This is the accepted version of the article, but might differ from the final archived version, which can be found at: https://doi.org/10.1016/j.cogdev.2022.101214

Children's Understanding of Relational Language for Quantity Comparisons

Michelle A. Hurst[1], Abrea Greene[1], Sarah H. Eason[2], Amy Claessens[3], & Susan C. Levine[1]

[1] University of Chicago Department of Psychology

[2] Purdue University Human Development and Family Studies

[3] University of Wisconsin-Madison Educational Policy Studies

Corresponding Author:
Michelle Hurst
5848 S University Ave Chicago IL 60637
hurstm@uchicago.edu

Author Note

This research was funded by a generous grant (#2018-0680) from the Heising-Simons Foundation to the Development and Research in Early Math Education (DREME) Network, of which S.C. Levine and A. Claessens are members. In addition, we would like to thank all our participants and participant partners, including the schools and organizations that participated and the Museum of Science and Industry (Chicago, IL). Author Contributions: MAH was involved in conceptualization, design, programming, data collection, personnel management, data analysis, and manuscript writing. AG: conceptualization, design, data collection, manuscript editing. SHE: conceptualization, design, personnel management, manuscript editing. AC: funding, conceptualization, design, manuscript editing. SCL: funding, conceptualization, design, manuscript editing. Declarations of interest for all authors: none

Abstract

Numerical order and magnitude are important aspects of early number knowledge. We investigated children's understanding of relational vocabulary for representing and communicating about order (*before/after*) and magnitude (*bigger/smaller*, *more/less*). In Experiment 1, 4- to 7-year-old children compared symbolic numbers, non-symbolic discrete quantities, and continuous amounts using relational words (N=151). In Experiment 2, 4- to 6-year-old children made yes/no judgements of ordinal and magnitude relations between symbolic numbers (N=60). Children showed lower performance with ordinal vocabulary compared to magnitude vocabulary (Experiment 1). Further, children were less likely to endorse the use of ordinal language for non-consecutive numbers than consecutive numbers, but showed no difference for magnitude language (e.g., 6 and 7 are *bigger* than 5, but only 6 comes *after* 5; Experiment 2). These results suggest a divergence in children's understanding of magnitude and ordinal vocabulary, suggesting a dissociation between these two concepts and/or the language used to communicate about them.

Keywords: relational language, mathematical language, math vocabulary, numerical order, numerical magnitude

Highlights

- Children show lower understanding of vocabulary for numerical order than magnitude
- Children do not systematically apply ordinal vocabulary to continuous areas
- Children apply ordinal language narrowly to refer to consecutive numbers

Introduction

Children's early number learning is multifaceted. Children must learn the meaning of numbers in terms of the quantity those numbers represent (i.e., cardinality; 3 = | | |) as well as their relative magnitudes (e.g., 4 > 3) and conventional order (e.g., 3, 4, 5). Furthermore, individual differences in children's ability to compare numerical magnitudes (i.e., "which number is bigger?") and judge numerical order (i.e., "are these three numbers in order?") are uniquely predictive of children's later math ability (Holloway & Ansari, 2009; Lyons et al., 2014; Lyons & Beilock, 2011; Starr et al., 2013). Given the importance of understanding these numerical relations, decades of research have investigated cardinal and magnitude representations of symbolic number, as well as how magnitude information is used to construct numerical order (e.g., Anderson & Cordes, 2013; Brannon, 2002; Dehaene, 2007; Moyer & Landauer, 1973; Wynn, 1992). However, less research has focused on how children's ordinal knowledge may be distinct from other numerical concepts, such as magnitude (Lyons & Beilock, 2013; Turconi et al., 2006; Tzelgov & Ganor-Stern, 2005). In the current set of studies, we investigate how young children compare quantities and amounts in terms of their magnitude and in terms of their ordinal position using specific vocabulary that describes each of these relations (e.g., *bigger, more, after*). In general, substantial work highlights the importance of math language and vocabulary for math learning (Hornburg et al., 2018; Powell & Nelson, 2017; Purpura et al., 2017; Purpura & Reid, 2016). Here, we focus on children's understanding of specific mathematical vocabulary that references different relational information (e.g., *bigger* references relative magnitude, *after* references relative order) to better understand how children map the meaning of these vocabulary words to quantity relations.

Ordinal Understanding

Quantities, such as sets of items, can be ordered in terms of their exact cardinal values, which inherently contain magnitude information. Symbolic numbers can similarly convey ordinal information in a way that maps onto precise magnitudes. However, symbolic numbers can also be used to convey ordinal information in a way that is divorced from the precise magnitudes they represent. For example, house numbers typically provide ordinal information without magnitude information; if a street includes the addresses of 1200, 1210, and 1220, they are likely to occur that order, but the numerical magnitudes are not meaningful in that the numerical differences of the addresses are unlikely to map onto the distances between the houses or the number of houses in between. In this ordinal sense, numbers are much like letters of the alphabet – M comes after H, but there are no magnitudes attached to M or H.

Children are able to recite the count list well before they have an understanding of the quantities representing those values (Huang et al., 2010; Sarnecka & Carey, 2008; Wynn, 1992) and part of this procedural knowledge is learning the “stable-order” principle, that the words must be recited in the same order (Gelman & Gallistel, 1978; Ip et al., 2018). This rote counting procedure may then provide children with an early, albeit impoverished, sense of the ordinal relations among symbolic numbers before those numbers are mapped to magnitudes. Indeed, it takes quite a while for children to connect the ordinal relations of number words to the magnitude relations of the underlying quantities they represent. For example, many children know the cardinal principle (i.e., that when counting, the last number word recited denotes the size of the set) but continue to struggle with knowing “ten” is more than “six” (i.e., the later-greater principle; Davidson et al., 2012; Le Corre, 2014) or that adding one item to a set results in the next number word in the count list (i.e., the successor function; Cheung et al., 2017; Sarnecka & Carey, 2008; Schneider et al., 2020). Furthermore, children’s ability to place quantities in an exact numerical order may be dependent on these other aspects of their number knowledge, such as the cardinal principle, the later-greater principle, and the successor function (Spaepen et al., 2018).

However, much of the research investigating ordinal knowledge has focused on children's knowledge of greater-than and less-than relations by measuring ordinal judgements of non-symbolic arrays (e.g., Brannon & Van de Walle, 2001; Bullock & Gelman, 1977; Cantlon et al., 2007; Cantlon & Brannon, 2006). Although these greater-than and less-than magnitude relations are deeply entangled with ordinal relations, whether children flexibly integrate their magnitude and ordinal knowledge into a single concept that is accessible for thinking about relative magnitude and conventionalized order separate from magnitude is an open question. This is important given that children's ability to decide which of two numbers is larger and whether or not three numbers are "in order" show separable and significant relations with advanced math ability (Lyons et al., 2014; Lyons & Beilock, 2011).

Connecting Magnitude and Order

Non-symbolic number is easily orderable in terms of relative magnitude and cardinality (e.g., from largest to smallest, with one-unit increments) and these constructs are likely deeply connected in the way people think about numerical information. Researchers have proposed that an ordered and approximate mental number line is an appropriate analogy for how numerical magnitudes are represented (Dehaene, 2011; Dehaene et al., 1998). Moreover, the spatial arrangement of numerical information impacts people's ability to make magnitude judgements. For example, judging which of two numbers is larger is easier when the numbers are in ascending order than descending order (Turconi et al., 2006) and responding to larger numbers is easier with the right hand than the left hand (termed the SNARC effect; Dehaene et al., 1993;

Gevers et al., 2006). Lastly, the connection between magnitude and order is evident early in infancy and in non-human primates (e.g., Brannon, 2002; Brannon & Terrace, 1998; Cantlon & Brannon, 2006; Picozzi et al., 2010; Suanda et al., 2008). For example, infants generalize numerical order from monotonically increasing discrete sets to monotonically increasing line lengths (de Hevia & Spelke, 2010) and children and non-human primates are better able to learn the order of stimuli that monotonically change in size (e.g., circles from smallest to largest) than when the same stimuli are ordered arbitrarily and nonmonotonically (e.g., Ohshiba, 1997; Terrace & McGonigle, 1994). Together, these findings suggest that numerical order as defined in terms of relative magnitude (e.g., Russell, 1903) may be an ontogenetically and phylogenetically early concept (see Anderson & Cordes, 2013 for a review).

However, ordinal and magnitude relations do show distinct patterns in the case of symbolic number. When making magnitude judgements about symbolic number, people demonstrate a robust distance or ratio effect, such that it is easier to compare two numbers that are further apart than two numbers that are closer together (e.g., Buckley & Gillman, 1974; Moyer & Landauer, 1973). In contrast, ordinal judgements about symbolic number show a reverse distance effect, such that values closer together (e.g., 4, 5, 6) are easier to judge than values further apart (e.g., 3, 5, 7; Lyons & Ansari, 2015; Turconi et al., 2006). Moreover, the relations between symbolic magnitude and ordinal knowledge to arithmetic skills change over the 1st to 6th grade time period: magnitude judgements start out as a stronger predictor, but by 6th grade ordinal knowledge is a stronger predictor of arithmetic success than magnitude knowledge (Lyons et al., 2014). Therefore, it may be that children's understanding of magnitude and order are dissociable when reasoning about symbolic numbers, which, as discussed earlier, represent cardinal magnitudes, but can also be used for ordinal representations in the absence of magnitude through the conventionalization of the count list.

Accessing Numerical Relations Through Math Language

Mathematical vocabulary is a critical component of early developing numeracy skills (e.g., Hornburg et al., 2018; Purpura et al., 2017; Purpura & Reid, 2016), and mathematical comparatives are one specific type of mathematical vocabulary (e.g., *more, bigger*). Although researchers have investigated the development of children's knowledge of specific comparatives, such as *more* (Gathercole, 1985; Odic et al., 2013; Palermo, 1973), *bigger* (e.g., Marschark, 1977), and the temporal comparatives *before/after* (e.g., Amidon & Carey, 1972; Clark, 1971; Johnson, 1975), whether children flexibly interpret these different vocabulary words as targeting the same underlying numerical concepts in the same way is an open question. Given the importance of relational language for children's relational reasoning in non-numerical domains (e.g., Christie & Gentner, 2014), understanding whether children understand mathematical vocabulary for different numerical relations in similar or different ways is a promising approach for better understanding children's developing numerical knowledge.

Prior work suggests that children have at least some understanding of these relational comparatives in unambiguous contexts by around 4 years old. For example, 3-year-olds perform above chance identifying *more* in both discrete and continuous contexts (Odic et al., 2013), 4-year-olds can act out sentences with the correct order of events described using *before* and *after* (although some children still make systematic errors; Clark, 1971), and 3- and 4-year-olds can readily point to the *biggest* and *littlest* in a display (although, still struggle with *next littlest;* Marschark, 1977). Thus, in the current study, we focused our age group on children who are likely to be at least somewhat familiar with the target relational vocabulary (i.e., 4-year-olds and older), but are continuing to develop more advanced numerical concepts, such as the “later-greater” principle and knowledge of the successor function (e.g., Davidson et al., 2012; Le Corre, 2014; Schneider et al., 2020), which might impact their ability to attend to ordinal or magnitude relations or their understanding of how these vocabulary words apply to ordinal and magnitude relations. This approach allows us to investigate children’s understanding of comparative vocabulary for targeted numerical relations in terms of their specific (and potentially systematic) interpretation errors, rather than general differences in children’s knowledge of these words.

Furthermore, it is worth noting that children’s early knowledge of words like *more* and *bigger* (as just described) is in reference to size, amount, and number. In contrast, children’s early knowledge of the words *before* and *after* may be specific to the context of temporal order (e.g., Sam eats breakfast *before* going to school). Thus, children may not directly learn ordinal words for quantities, but instead apply already known temporal words in the new and different context of quantitative order. What components of children’s early number knowledge allow them to make this mapping between vocabulary for temporal order and numerical order? One possibility is children’s early learning of the count list. Although the count list is initially learned in a rote, but stable, order (e.g., Huang et al., 2010; Sarnecka & Carey, 2008; Wynn, 1992), other work suggests that it is understanding the structure of counting that is necessary (although, not always sufficient) for a range of early number concepts (e.g., Cheung et al., 2017; Davidson et al., 2012; Schneider et al., 2019; Schneider et al., 2020). In either case, children’s experience with counting may be important for their ability to think about symbolic number in terms of order using temporal vocabulary (e.g., when counting, larger numbers are said *after* smaller numbers).

The Current Study

In the current study, we investigated children's ability to compare two quantities or amounts when prompted to do so with quantitative terms that refer to magnitude (i.e., *bigger, smaller, more, less*) or order (i.e., *before, after*). Our primary question is whether children map both magnitude and ordinal vocabulary onto the same underlying conceptualization of relative quantity (e.g., a "later is greater" concept that directly aligns relative magnitude and relative order), or whether their receptive knowledge of magnitude and ordinal vocabulary show dissociations. We address this primary question in two experiments. In Experiment 1, we systematically compare children's ability to compare quantities across (a) relational vocabulary that refer to either magnitude or ordinal relations and (b) different quantity types that vary in the type of ordinal information available. In Experiment 2, we narrow our focus to test a specific hypothesis about where children's understanding of magnitude and ordinal vocabulary might diverge: when judging the relations between consecutive and non-consecutive symbolic numbers, ordinal vocabulary might be less readily applied to values that are not consecutive, compared to magnitude vocabulary (e.g., Turconi et al., 2006).

If children's ordinal understanding of quantities is inherent in their understanding of the magnitude of quantities, and they can flexibly and readily access this understanding regardless of whether they are prompted to compare values using magnitude or ordinal relational vocabulary, then we would not expect children's behavior to systematically differ across either distinct relational vocabulary or quantity types. If, on the other hand, children show differences in their knowledge of ordinal and magnitude vocabulary across different quantities, their response patterns may reveal different magnitude and ordinal concepts and/or differences in how they map these terms (bigger/smaller, before/after) onto a unified magnitude/ordinal concept.

Finally, in Experiment 2, we also explore the role of counting knowledge in children's understanding of ordinal vocabulary for symbolic numbers. Given children's early learning of a rote and stable count list, followed by increasing understanding of how counting relates to magnitude and the structure of the count list (e.g., Davidson et al., 2012; Sarnecka & Carey, 2008), we included various measures of counting ability (e.g., what comes next, counting by 2s) to begin generating hypotheses about how children's ability to flexibly think about the order of number words in the count list might be related to their ability to map ordinal words to numerical order.

Experiment 1

In Experiment 1, we asked children to compare two symbolic numbers, two discrete quantities, or two continuous amounts based on the terms *bigger/smaller*, *more/less*, or *after/before.* Although all three quantity types convey magnitude information, they differ in the kind of ordinal information available. The "later-greater" principle of symbolic number provides a conventionalized order in addition to magnitude, such that numbers representing larger values come later in the count list. In contrast, the "later-greater" principle does not apply to continuous area-based magnitudes, as they do not have a discrete or conventionalized order (i.e., there are not immediate successors and magnitude-based monotonic order could be increasing or decreasing). In other words, although continuous area-based magnitudes are orderable by magnitude, and prior work suggests that even infants and non-human primates can reason about the magnitude-based order of continuous magnitudes (de Hevia & Spelke, 2010; Ohshiba, 1997; Terrace & McGonigle, 1994), this order is entirely based on relative magnitude and is not conventionalized. Lastly, discrete sets are similarly orderable by non-symbolic magnitude, again a feature that even infants attend to (de Hevia & Spelke, 2010). However, unlike continuous area, discrete sets can be enumerated and their cardinalities can be represented by symbolic numbers, making it possible that children will apply the conventionalized "later-greater" ordering of symbolic numbers to non-symbolic discrete sets. Thus, in Experiment 1 we ask two specific questions: (1) Within a particular quantitative context (e.g., numbers, dots, area) how do children's relational judgements vary across different relational vocabulary? and (2) How does children's understanding of specific relational vocabulary vary across different quantitative contexts? By examining these questions, we can begin to investigate whether children apply both ordinal and magnitude vocabulary similarly and flexibly or whether there are either general differences in the level of understanding and/or systematic differences in how these vocabulary words are applied (specific predictions are detailed in *Analysis Plan and Predictions*).

This experiment was pre-registered on AsPredicted.org (#18471; https://aspredicted.org/cy742.pdf) and the pre-registration, materials, data, and data-analysis code are available on the Open Science Framework (https://osf.io/e9y76/).

Method

Participants

One-hundred-and-fifty-one[1] 4 through 7-year-old children are included in the final sample. Children (ages reported as years;months) were randomly assigned to one of three conditions: Symbolic Number (n = 52; M_{age} = 5;9; Range: 4;0 to 7;10; 28 girls, 24 boys), Discrete Sets (n = 49; M_{age} = 5;8; Range: 4;1 to 7;10; 28 girls, 20 boys, 1 undisclosed), and Continuous Area (n = 50; M_{age} = 5;8; Range: 4;0 to 7;9; 25 girls, 25 boys). This sample size was

[1] We pre-registered a final sample of 150, with 50 children in each condition. Due to the dynamic nature of data collection at museums and other programs, as well as an experimenter error in condition assignment, we ended data collection with one additional child overall and a slight imbalance across conditions.

chosen a priori to provide 80% power (calculated using G*Power; Faul et al., 2007) to detect effect sizes as small as 0.4 for paired *t*-tests and 0.56 for independent *t*-tests.

Children were recruited through the greater Chicago, IL area and participated in our campus laboratory, at a local science museum, or at local schools and childcare programs from January 2019 to November 2019. Parents provided written consent and children provided verbal assent. All children received a small prize (e.g., sticker or small toy). Parents of children coming into the lab received $10 for travel compensation and schools or childcare programs received a $50 gift card for supplies. Additional demographic data were not systematically collected, limiting our ability to contextualize these findings. However, these data were collected for the experiment reported in Supplemental Materials, and we expect similar demographic characteristics across the two samples because we used similar recruitment methods. All procedures were approved by the University of Chicago Institutional Review Board (IRB17-1599 "Relational Math Reasoning").

Design

Children were presented two quantities and questions about those quantities in a forced-choice task. All children received three types of questions: Order ("which comes after [before]?"), Size-Based Magnitude ("which is bigger [smaller]?"), and Quantity-Based Magnitude ("which is more [less]?"). In a between-subject design, children were randomly assigned to one of three conditions that varied the stimuli being compared: written numerals paired with number words (Symbolic Number Condition), sets of small squares that varied in the number of items (Discrete Set Condition), or two single squares that varied in area (Continuous Area Condition).

Stimuli

All conditions included the same 36 trials and only differed in the stimulus presented to represent the quantity (see Figure 1). In the Symbolic Numeral condition, the two options were Arabic numerals centered on the left and right half of the screen and were paired with English number words spoken aloud by the experimenter. In the Discrete Sets condition, the two options were sets of small red squares placed randomly within a larger rectangle (13 cm wide by 10 cm tall, presented as a border for all trial types). We did not vary the item size of the Discrete Sets (each item was 1.7 cm square, resulting in an area of 2.89 cm^2), meaning that cumulative area co-varied with number. In the Continuous Area condition, the two options were each a single red square that had the same area as the cumulative area of the corresponding Discrete Sets trial (e.g., the stimulus corresponding to 9 items had an area of 9 x (2.89 cm^2)).

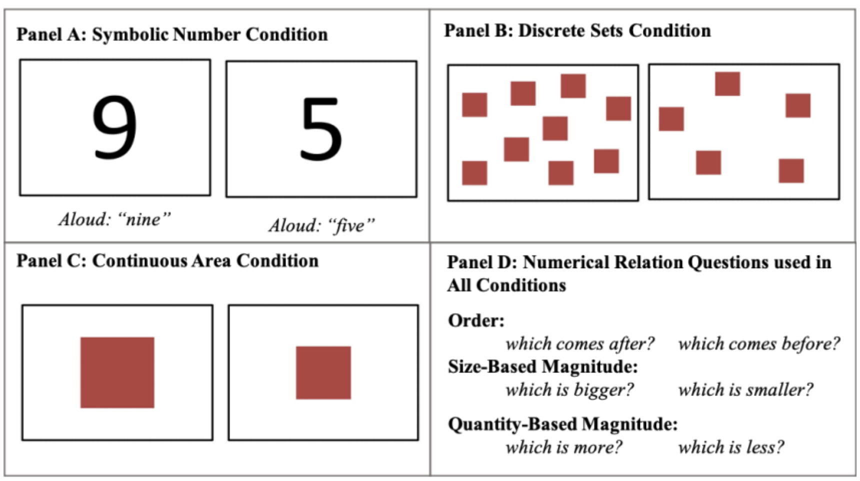

Figure 1: Example stimuli from Experiment 1. In the Symbolic Number Condition (Panel A), the experimenter pointed to each stimulus while saying the number word (e.g., *nine* and *five*). In the Discrete Set Condition (Panel B) and the Continuous Area Condition (Panel C), the experimenter pointed to each stimulus referring to it as "this". In all conditions, children were asked the same set of questions (Panel D) that probed both ordinal and magnitude relations.

The 36 trials were made up of 6 unique comparisons (2 vs. 5; 3 vs. 6; 3 vs. 8; 4 vs. 7; 4 vs. 8; 5 vs. 9) shown twice for each of the three quantitative questions (which comes after [before], which is bigger [smaller], which is more [less]), once for each positive direction (after, bigger, more) and once for each negative direction (before, smaller, less). To ensure that variability in performance could be attributed to interpreting the vocabulary word rather than difficulties discriminating the magnitudes, stimuli on each trial differed by at least a 1:2 ratio, which even infants can discriminate (see Cordes & Brannon, 2008 for a review).

Procedure

Trained research assistants administered the task using a 13-inch laptop computer running PsycoPy2 Version 1.85.4 and the task typically lasted no more than 10 minutes. On each trial, the screen displayed two visual stimuli corresponding to the condition and a question written in English along the top of the screen. The experimenter read the question aloud, then pointed to the left and right options while saying “this or this” (Continuous Area and Discrete Sets conditions) or the number words (e.g., “two or three”; Symbolic Number condition). Children responded by pointing to the left or right option and/or saying the number word aloud (Symbolic Number condition only). The experimenter recorded the child’s response with the left or right arrow key, which also immediately advanced to the next trial. The correct answer was on the right side for half of the trials of each question (i.e., if children had a side bias, they would be at 50%).

Analysis Plan and Predictions

The dependent variable is children’s proportion correct on each trial type, where correctness was defined as selecting the larger magnitude when asked about *bigger*, *more*, or *after*, and the smaller magnitude when asked about *smaller*, *less*, or *before*[2]. We had two central

[2] Given the spatial layout of the task, children might have interpreted the ordinal questions in terms of spatial order, with the left side = “before” and the right = “after”. When the ordinal trials are scored in this way, average performance is around 50% on all three quantity types, suggesting there was not a group level tendency to use this interpretation. However, some children may have interpreted it this way: 8 children in the Continuous Area condition and 2 children in the Discrete Sets condition responded in this way on all ordinal trials. Given that our interest was specifically in whether children would interpret these ordinal vocabulary words in terms of relative magnitude (i.e., *after = bigger/more; before = smaller/less*) and that

questions. Our analysis plan and primary hypotheses were pre-registered and deviations from the pre-registration are clearly described.

First, within a particular quantitative context (e.g., numbers, dots, area) how do children's relational judgements vary across different relational vocabulary? To address this question, we pre-registered performing *paired t*-tests within each quantity stimulus type (discrete sets, continuous area, symbolic number), comparing performance on the ordinal questions (*before/after*), the size-based magnitude questions (*bigger/smaller*), and the quantity-based magnitude questions (*more/less*). Although not preregistered (due to an oversight in the preregistration), we also report children's performance within each cell to chance, to better interpret the pattern of differences. We predicted that children would perform better on magnitude questions than ordinal questions when asked about continuous areas, as prior work suggests children of this age are able to reason about relative magnitude of amounts using magnitude vocabulary (e.g., Marschark, 1977; Odic et al., 2013), but continuous amounts do not have a conventionalized order (i.e., they do not have immediate successors and could readily be ordered by increasing or decreasing size). However, for symbolic numbers and discrete sets, which can be readily mapped to magnitude and have a conventional order via the count list, we did not make a directional prediction, but instead were interested in investigating which kind of relation children more readily map to these quantities. Lastly, for all quantity types, we were interested in whether children's performance would differ across the two terms for magnitude, *more/less* and *bigger/smaller*.

there was not a group-level tendency to interpret ordinal words spatially, we continued to score and interpret children's ordinal knowledge in terms of a magnitude-based interpretation, as originally planned. However, when excluding these children from the analyses, the pattern remains unchanged.

Second, how does children's understanding of specific relational vocabulary vary across different quantitative contexts? To address this question, we pre-registered performing *independent t*-tests across each of the three quantity type conditions, within each of the question types, which used different relational vocabulary. When asked about order (*before/after*), we predicted that children would perform better with Symbolic Numbers than Continuous Area, with performance on Discrete Sets being intermediate, because symbolic numbers are most closely tied to a conventional order whereas continuous areas do not typically have a conventional order. Further, if children readily map discrete sets (i.e., non-symbolic numerical values) to numerical order then we would expect children to be able to readily make ordinal judgements with these stimuli using ordinal vocabulary, similar to their ability to do so for symbolic number. On the other hand, if children's conception of order is specifically tied to symbolic number, and not (at this age) readily generalized to non-symbolic number (potentially akin to differences in knowing numerical successors for symbolic count list versus discrete sets; Davidson et al., 2012), then we would expect children to show relatively worse performance on ordinal words that refer to discrete sets, closer to what we predict for judgements of continuous area. We did not have specific directional predictions about how children's performance might vary across contexts when asked about relative magnitude because all three quantity types convey magnitude.

As a secondary exploratory question, we were also interested in variations in children's performance across the age range in our sample. This age range was chosen such that children should have basic knowledge of the relational vocabulary tested and basic number knowledge (e.g., counting, the cardinal principle; Sarnecka & Carey, 2008), but are still developing more advanced numerical concepts (e.g., successor function; Davidson et al., 2012; Schneider et al., 2020). This developing knowledge might impact how children interpret magnitude and ordinal vocabulary across different quantity contexts. To explore this hypothesis, we investigated children's ordinal judgements across the three quantitative contexts as a function of age, using correlations with age in months to investigate general trends (with n = 50, we have 80% power to detect correlations as small as .37) and using comparisons to chance (50%) with younger (4 and 5 years old) and older children (6 and 7 years old) separately to contextualize these trends (with n = 17-32 in each group, we have 80% power to detect medium to large effects of .51 to .72). Importantly, these analyses are exploratory (i.e., although we did plan on exploring age a priori, we did not pre-register a specific analysis plan or hypotheses) and because we are only powered to detect medium to large age effects, they are intended to support hypothesis generation about the potential developmental time course of children's knowledge.

All analyses were conducted in R 4.0.2 (R Core Team, 2020) with RStudio (R Studio Team, 2016), and the following packages *stringr_1.4.* (Wickham, 2019), *forcats_0.5.0* (Wickham, 2020), *effsize_0.8.0* (Torchiano, 2018), *gt_0.2.2* (Iannone et al., 2020), *psychReport_1.1.0* (Mackenzie, 2018), *rstatix_0.6.0* (Kassambara, 2020), *purrr_0.3.4* (Henry & Wickham, 2019), *readxl_1.3.1* (Wickham & Bryan, 2019), *ggplot2_3.3.2* (Wickham, 2016), *tidyr_1.1.1* (Wickham & Henry, 2018), *dplyr_1.0.1* (Wickham, Francois, et al., 2018), *readr_1.3.1* (Wickham, Hester, et al., 2018), and *cocor*_1.1.3 (Diedenhofen & Musch, 2015).

Results and Discussion

Based on our pre-registered plan, our primary analyses do not separate the different directions of the relation (i.e., *bigger* vs. *smaller*), but the descriptive data for each word and collapsed across each pair of words are reported in Table 1.

Table 1: Mean Proportion Correct (Standard Deviation) of Performance in Each Visual Condition for Each Relational Word

	Magnitude – Size			Magnitude – Quantity			Order		
	bigger	smaller	both	more	less	both	after	before	both
Symbolic Number	0.92 (0.17)	0.91 (0.17)	0.92 (0.14)	0.91 (0.16)	0.82 (0.30)	0.87 (0.20)	0.71 (0.34)	0.79 (0.26)	0.75 (0.26)
Discrete Sets	0.94 (0.16)	0.93 (0.14)	0.93 (0.14)	0.94 (0.14)	0.86 (0.25)	0.90 (0.16)	0.64 (0.33)	0.67 (0.31)	0.65 (0.26)
Continuous Area	0.97 (0.07)	0.93 (0.14)	0.95 (0.08)	0.86 (0.23)	0.89 (0.16)	0.87 (0.17)	0.54 (0.33)	0.54 (0.32)	0.54 (0.27)

Within a particular quantitative context, how do children's relational judgements vary across different relational terms?

First, we investigated how children's performance varied for different kinds of relational language on the different quantity types (see Figure 2, comparing across relational vocabulary within each condition, shown on the x-axis). Within each family of tests, we control for family wise error rates using Holm's sequential procedure to adjust the alpha level (Holm, 1979). Thus, an alpha level is reported separately for each test comparing across the relational language categories.

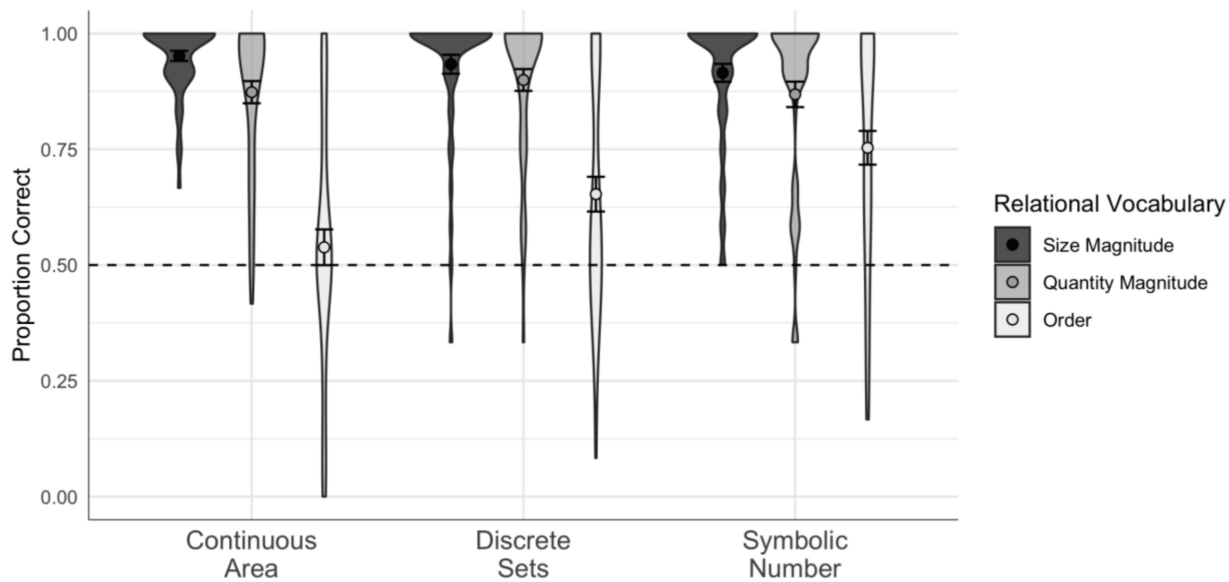

Figure 2: Proportion correct in Experiment 1 within each stimulus type (x-axis) when judging Size Based Magnitude (*bigger* and *smaller;* darkest grey, left), Quantity Based Magnitude (*more* and *less;* light gray, middle), and Order (*after* and *before;* white gray, right). Means and standard errors are plotted with the violin plots of the underlying distribution of child-level data.

Continuous Area. When making judgements about squares of different sizes, children scored highest when asked which was bigger/smaller (Size Magnitude), followed by more/less (Quantity Magnitude), and then before/after (Order), with all pairwise comparisons revealing significant differences: Size Magnitude vs. Order, $t(49) = 10.2$, $p < 0.001$, $d = 2.1$ ($\alpha_{1/3} = 0.017$); Quantity Magnitude vs. Order, $t(49) = 8.0$, $p < 0.001$, $d = 1.46$ ($\alpha_{2/3} = 0.025$); Size Magnitude vs. Quantity Magnitude, $t(49) = 3.6$, $p < 0.001$, $d = 0.55$ ($\alpha_{3/3} = 0.05$). Furthermore, children scored significantly above chance (.50) when asked about Size Magnitude, $t(49) = 41.3$, $p < 0.001$, and Quantity Magnitude, $t(49) = 15.6$, $p < 0.001$, but not when asked about Order, $t(49) = 0.99$, $p = 0.33$.

Discrete Sets. When making judgements about sets, children scored highest when asked which was bigger/smaller or more/less (Size and Quantity Magnitude, respectively), which were not significantly different from each other, Size Magnitude vs. Quantity Magnitude, $t(48) = 1.75$, $p = 0.09$, $d = 0.2$ ($\alpha_{3/3} = 0.05$). Children scored lowest when asked which comes before/after (Order), which was significantly different from both bigger/smaller, Size Magnitude vs. Order, $t(48) = 7.8$, $p < 0.001$, $d = 1.3$ ($\alpha_{1/3} = 0.017$), and more/less, Quantity Magnitude vs. Order, $t(48) = 6.7$, $p < 0.001$, $d = 1.1$ ($\alpha_{2/3} = 0.025$). Unlike the Continuous Area condition, children scored above chance (.50) on all three relations: Size Magnitude, $t(48) = 21.2$, $p < 0.001$, Quantity Magnitude, $t(48) = 17.0$, $p < 0.001$, and Order, $t(48) = 4.1$, $p < 0.001$.

Symbolic Number. When making judgements about symbolic numbers, children scored highest when asked which was bigger/smaller (Size Magnitude), followed by more/less (Quantity Magnitude), and then before/after (Order), with all pairwise comparisons showing significant differences: Size Magnitude vs. Order, $t(51) = 4.8$, $p < 0.001$, $d = 0.73$ ($\alpha_{1/3} = 0.017$); Quantity Magnitude vs. Order, $t(51) = 3.5$, $p = 0.001$, $d = 0.49$ ($\alpha_{2/3} = 0.025$); Size Magnitude vs. Quantity Magnitude, $t(51) = 2.5$, $p = 0.02$, $d = 0.25$ ($\alpha_{3/3} = 0.05$). As in the Discrete Sets condition, children scored significantly above chance on all three relations: Size Magnitude, $t(51) = 21.5$, $p < 0.001$, Quantity Magnitude, $t(51) = 13.4$, $p < 0.001$, and Order, $t(51) = 7.0$, $p < 0.001$.

How does children's understanding of specific relational terms vary across different quantitative contexts?

Next, we investigated how performance varied on each of the relational vocabulary types (e.g., bigger/smaller) across the three quantity types (area, discrete sets, and symbolic number; see Figure 2; comparing across conditions on the x-axis for each type of relational vocabulary).

When a test of equality of variances was significant, a Welch's *t*-test is reported (not assuming equal variances), however, the interpretation does not change when variances are assumed to be equal. As in the previous analyses, Holm's sequential procedure (Holm, 1979) was used to control family wise error rates within each family of tests comparing across condition within a relational language type.

Size Based Magnitude. Children's scores did not significantly differ across the three quantity conditions when asked which was *bigger*/*smaller*: Continuous Area vs. Symbolic Number, $t(80.3) = 1.65$, $p = 0.10$, $d = 0.32$ ($\alpha_{1/3} = 0.017$); Continuous Area vs. Discrete Sets, $t(73.4) = 0.8$, $p = 0.44$, $d = 0.16$ ($\alpha_{2/3}$ = NA, because test 1/3 was not significant); Discrete Sets vs. Symbolic Number, $t(99) = 0.66$, $p = 0.51$, $d = 0.13$ ($\alpha_{3/3}$ = NA).

Quantity Based Magnitude. As shown for size-based magnitude, children's scores did not significantly differ across the three quantity conditions when asked to decide which was *more*/*less*: Discrete Sets vs. Symbolic Number, $t(99) = 0.85$, $p = 0.40$, $d = 0.17$ ($\alpha_{1/3} = 0.017$); Continuous Area vs. Discrete Sets, $t(97) = 0.78$, $p = 0.44$, $d = 0.16$ ($\alpha_{2/3}$ = NA); Continuous Area vs. Symbolic Number, $t(100) = 0.13$, $p = 0.90$, $d = 0.03$ ($\alpha_{3/3}$ = NA).

Order. Unlike the two magnitude-based relations, children's judgments of the ordinal relations *before*/*after* varied across the three conditions. Children scored higher when comparing symbolic numerals than when comparing continuous areas, $t(100) = 4.1$, $p < 0.001$, $d = 0.80$, $\alpha_{1/3} = 0.017$. Judgements about discrete sets were intermediate, but not statistically different (when correcting for multiple comparisons) from either continuous area, $t(97) = 2.13$, $p = 0.04$, $d = 0.43$ ($\alpha_{2/3} = 0.025$), or symbolic number, $t(99) = 1.9$, $p = 0.06$, $d = 0.38$ ($\alpha_{3/3}$ = NA).

How does children's age relate to their understanding of ordinal language?

The largest differences between quantitative contexts were in children's judgements of order. When comparing symbolic numbers, children scored well above chance when judging magnitude and order, suggesting they understand both categories of relational terms in this context. However, the fact that children were not able to systematically judge the relative order of continuous areas suggests that children do not readily map the conventionalized numerical order of symbolic numbers (i.e., determined by the count list) to magnitude in a way that generalizes to continuous areas. Moreover, children's performance comparing discrete sets with order vocabulary was in between these other two conditions – suggesting that they may be beginning to map discrete magnitudes onto the conventionalized order of number symbols, but that their understanding of order vocabulary for these sets is less robust than for symbolic number.

Children's performance mapping order terms onto the order of both discrete sets, n = 49, $r = .56$, $p < .001$, and symbolic numbers, n = 52, $r = .31$, $p = .024$, showed a significant correlation with age, with children's performance improving across the age range (see Figure 3). However, children's performance with continuous area was not significantly correlated with age, n = 50, $r = .10$, $p = .473$. Moreover, younger children (4- and 5-year-olds) were only above chance in the symbolic number condition, n = 31, $M = 0.72$, $p < .001$, but not in the discrete sets condition, n = 32, $M = 0.57$, $p = .115$, or in the continuous area condition, n = 32, $M = 0.52$, $p = .642$). On the other hand, older children (6- and 7-year-olds) scored above chance with both symbolic numbers, n = 21, $M = 0.81$, $p < .001$, and discrete sets, n = 17, $M = 0.81$, $p < .001$, but were again around chance with continuous area, n = 18, $M = 0.57$, $p = .360$.

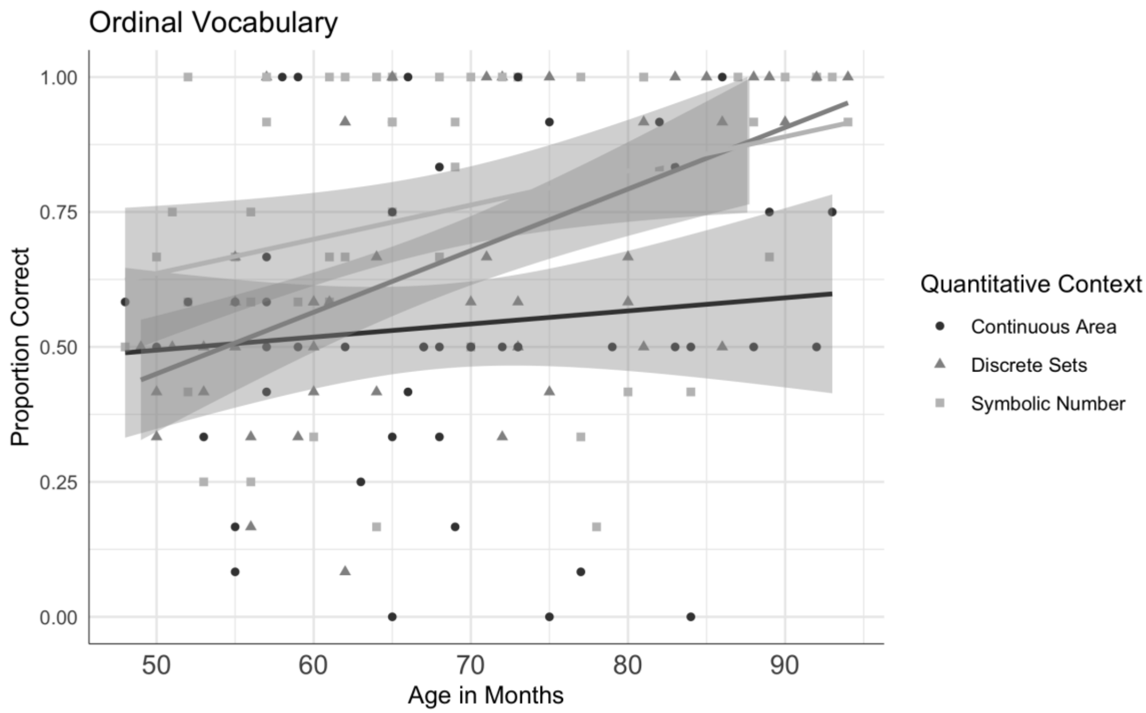

Figure 3: Scatterplot displaying the relation between proportion correct (x-axis) on the ordinal vocabulary words (before and after) and age in months (y-axis) separated by quantitative contexts (by color). Fit lines are based on linear models with standard error.

Summary of Experiment 1

In Experiment 1, we have two primary findings: children were better able to judge magnitude relations than ordinal relations across all three quantity contexts and only judgements of ordinality varied for different quantitative contexts. Specifically, children did not systematically apply the ordinal words *before* and *after* to amounts of continuous area. In contrast, children were able to systematically judge the relative order of symbolic numbers using ordinal vocabulary. Finally, children's ability to use ordinal words to judge the relative order of discrete sets was between these other two quantity contexts.

In addition, although overall children scored significantly above chance on their use of ordinal vocabulary, our exploratory age analyses suggest that children's understanding of ordinal vocabulary for discrete sets may emerge sometime between 4- and 7-year-old. Although speculative (given the narrow age range tested and limitations with sample size), we discuss some important open questions generated by this age-related pattern of performance in the General Discussion.

Experiment 2

Although children in Experiment 1 demonstrated worse performance with ordinal numerical relations than magnitude numerical relations, even for symbolic number, the overall performance differences in the paradigm used in Experiment 1 does not provide specific insight into where children's understanding of ordinal language goes wrong. In Experiment 2, we further address the question of how children map magnitude and ordinal vocabulary to symbolic number but make several modifications to the paradigm to allow for a better comparison of the two contexts and to test specific hypotheses about how children's understanding of ordinal language systematically diverges from their understanding of magnitude language.

In the prior experiment, the two values being compared were always about a 2-fold change apart (e.g., 4 vs. 8 or 3 vs. 6). This was deliberate to ensure that children would be able to visually discriminate the values (Cordes & Brannon, 2008; Xu et al., 2005), but it may have made the ordinal comparisons more difficult than the magnitude comparisons overall. As discussed in the Introduction, prior work on magnitude comparison typically reveals ratio or distance effects such that comparing values further apart is easier than values closer together (Moyer & Landauer, 1973). However, the opposite pattern has been shown for ordinal comparisons, such that values closer together are easier to compare than values further apart (Lyons & Ansari, 2015). Thus, it may be that testing far-apart values facilitated magnitude comparisons, while hindering ordinal comparisons. In Experiment 2, we include both consecutive (e.g., 5 vs. 6) and non-consecutive (e.g., 5 vs. 7) relations to directly test this hypothesis.

We made three additional modifications from the methods used in Experiment 1. First, we narrowed our age range slightly to focus on 4- through 6-year-olds, rather than 4- through 7-year-olds, because data from Experiment 1 suggested that children were approaching ceiling at the upper end of the age range. Second, we did not include *more/less* because these terms are not entirely symmetric across continuous and discrete entities. Although the exact usage of each word is often debated, *less* is often meant to refer to continuous quantities and substances, whereas *fewer* is often more correct for discrete quantities. Thus, we chose to focus on *bigger/smaller* and *after/before* to streamline the number of required trials and more cleanly contrast relational terms that are unambiguously about magnitude and order, respectively. Third, we changed the paradigm and phrasing of the question to allow for a better comparison across magnitude and ordinal vocabulary: in Experiment 2, we use a yes/no relational verification task rather than a comparison task. In the prior experiment, the ordinal phrasing of the comparison task ("which comes before?" or "which comes after?") may have been less familiar and more difficult to understand than the magnitude phrasing (e.g., "which is more?" or "which is bigger?"). Although only anecdotal from the authors' experiences as English speakers, the question format of "which" may feel awkward and unusual for the words before/after (the ordinal words first, second, etc. might be more appropriate). Furthermore, an initial pilot study that changed the phrasing of the question but continued to use a comparison paradigm indicated that the comparison paradigm may mask children's reasoning by forcing a choice between two suboptimal options. For example, when asked whether 3 or 7 comes after 5, children may believe that neither really comes after 5, but when forced to choose, decide that 7 is still a *better* option than 3 (details of this experiment can be found in Supplemental and on the OSF project page). Instead, the yes or no verification paradigm does not rely on comparing the relations of multiple numbers and instead allows children to express their beliefs about specific relations independently.

Finally, we also included measures of children's counting ability to provide an initial exploration of the potential mechanisms that might contribute to children's systematic interpretation of ordinal language. Although we did not have strong a priori hypotheses, prior work suggests that children's knowledge of the structure of counting is necessary for a range of early number concepts, but often not entirely sufficient (e.g., Cheung et al., 2017; Davidson et al., 2012; Schneider et al., 2019; Schneider et al., 2020). Thus, we were interested in exploring whether children's knowledge of the symbolic count list would be related to their understanding of the ordinal relations *before* and *after* for symbolic numbers. Furthermore, given that we are most interested in children's interpretation of ordinal relations beyond the next number, we included counting tasks that measured knowledge of the consecutive numbers in the count list, as well as children's ability to count flexibly, including skip counting, backward counting, and counting from a number other than "one".

Thus, we had three research questions: (1) Do children's judgements with ordinal and magnitude vocabulary differentially depend on whether the numbers being compared are consecutive? (2) How does age relate to children's knowledge of these relational vocabulary terms? (3) How does counting ability, and in particular flexible counting, relate to children's knowledge of the relational vocabulary? This experiment was pre-registered on OSF (https://osf.io/9pc2s) and all materials, data, and analysis code is available (https://osf.io/cv763).

Method

Participants

The sample consisted of 60 4- to 6-year-old children (*M* = 5;5, age range from 4;0 to 6;9; 24 girls, 22 boys, 14 undisclosed gender). An additional three children participated but are excluded because they did not complete the primary task of interest. This sample size was chosen a priori to provide 80% power for testing our central hypothesis (i.e., detecting an interaction between the two repeated measures of at least d = 0.6), as well as appropriate pairwise tests (based on power analyses and simulations reported in Brysbaert, 2019).

Children were recruited from our lab database, Children Helping Science (childrenhelpingscience.com), and other advertisements and forums on social media. Data collection occurred from August 2020 to November 2020, and due to the COVID-19 pandemic, all children were tested over video chat using Zoom software. Parents completed an optional demographic survey (78% completed at least some of the demographic survey, of which 77% were mothers, 15% fathers, and 8% were nonbinary or did not disclose gender). Based on this subsample, parents reported that 11% of children identified as Asian, 9% as Black or African American, 89% as White, and 26% as Hispanic or Latino/a/x (parents could select multiple options). All children used English as a primary or secondary language at home, with 95% reporting it as their primary language. Parents were predominantly educated with at least a bachelor’s degree or equivalent (78%) and reported relatively high household incomes, with 60% reporting an annual family income of $100K or more, 26% between $50 to $99K, and 14% between $15 and $49K. All procedures were approved by the same institutional IRB, parents provided informed consent, and families were offered a $5 gift card via their email.

Procedure

Children completed a Relational Verification task, probing Ordinal and Magnitude relations, and a battery of counting tasks. All tasks were administered by a trained research assistant over Zoom video chat software and took approximately 15-20 minutes to complete.

Relational Verification Task. The Relational Verification task was programmed in PsycoPy3 version 3.2.4 (but was also administered using PsychoPy2020 v2020.1.2 for some children). The task was presented via the screen sharing feature of Zoom. Prior to beginning the task, children were shown a "Ready?" screen that had a different image in each corner and children were asked what they saw. This screen was used as a warmup task (given that some children have lower comfort with video chat) and to help troubleshoot any issues with visibility of the screen on the child's device.

The task consisted of two counterbalanced blocks: Ordinal and Magnitude (n = 30 per order). Each block consisted of 18 trials that presented children with a yes/no verification question about the given relation for that block, such as "*Does 5 come after 4?".* These 18 trials included three trial types, with six trials each: directionally correct consecutive trials, directionally correct non-consecutive trials, and directionally incorrect trials. The directionally correct trials are the critical trials for testing our hypothesis, as they ask about numbers in the correct relational direction as the probe but were either consecutive or non-consecutive with respect to the comparison number. On the consecutive trials, the number was right next to the target number, in the relational direction being probed (e.g., "*does 6 come after 5?*" / "*does 4 come before 5?*"). On the non-consecutive trials, the number was two away from the target number, in the correct relational direction being probed (e.g., "*does 7 come after 5?*" / "*does 3 come before 5?*"). In contrast, the six directionally incorrect trials asked about numbers that were incorrect regardless of whether or not children used a narrow, consecutive interpretation of *before/after*, because the numbers were relationally opposite of the correct answer (e.g., "*does 2 come after 4?*" / "*does 5 come before 4?*"). These trials were included so that "yes" was not always the correct answer. Within each of the three trial types, there were two trials for each target number (4, 5, and 6) and half of the questions asked about the positive direction (*after* on the ordinal block and *bigger* on the magnitude block), while the other half asked about the negative direction (*before* on the ordinal block and *smaller* on the magnitude block).

On each trial (see Figure 4 for an example), children were shown the written question on the top of the screen (e.g., "Does this number come after 5?"), with the target number displayed large, centered, and below the question (e.g., "3"). The question was read aloud by the experimenter with the target number replacing the phrase "this number" (e.g., "Does three come after five?"). The child responded verbally, and the experimenter recorded the child's response by pressing the "y" key for yes or the "n" key for no. Next, there was a 500ms intertrial pause with only a single dot in the center of the screen and then the next trial began. The experimenter repeated the questions as needed and provided general encouragement (e.g., "You're doing great!", "Let's keep going!"), but did not provide feedback about correctness.

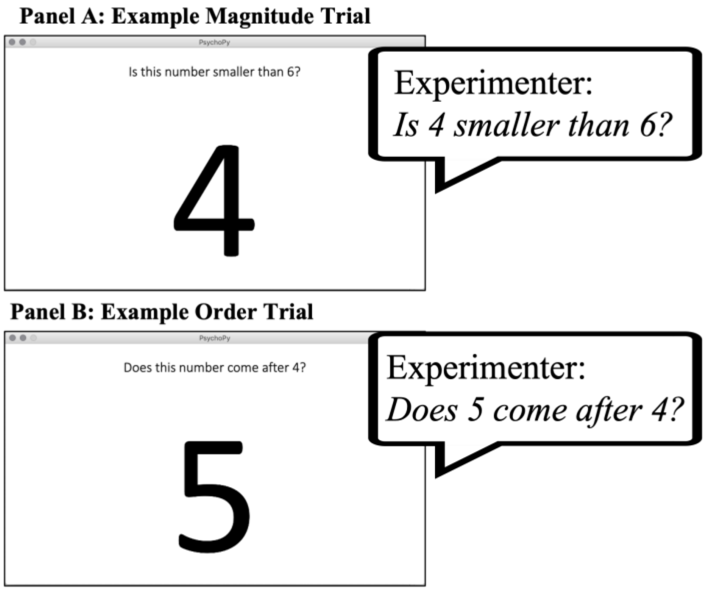

Figure 4: Example trials from Experiment 2 Magnitude Block (Panel A, upper) and Ordinal block (Panel B, lower). The screen was presented via screen sharing over video chat and the target question was asked verbally by the experimenter.

After the verification trials, children were asked “how many numbers come after five?” in the ordinal block and “how many numbers are bigger than five?” in the magnitude block. The experimenter typed the child’s response as the child was speaking, omitting some aspects of child speech (e.g., ums) but capturing the key components of their response. These free responses were coded from the experimenter’s live-typed response, except for 5/120 responses that were re-typed from the video recordings because the experimenter indicated they mistyped or could not capture the child’s response live.

Counting Tasks. The battery consisted of four types of trials, always presented in this order: What’s Next, Flexible Forward, Flexible Backward, and Skip counting. For all trials, the experimenter administered the question verbally and recorded the child’s verbal responses via a Qualtrics survey. As pre-registered, children were given two counting scores: What’s Next score based on the proportion of trials correct on the What's Next subtask and a Flexible Counting score based on the proportion of trials correct on the remaining subtypes (Flexible Forward, Flexible Backward, and Skip counting trials), calculated as the sum of the total number of problems correct out of the total number of problems administered.

The What's Next task (Schneider et al., 2019) contained two practice trials followed by eight test trials. On each trial, children were given a target number and asked what number comes next (e.g., "if I say 23, what comes next?") and the experimenter typed the child's response verbatim. Children were given accurate and corrective feedback on the practice trials (targets: 1 and 5) but not on test trials (targets, asked in this order: 23, 29, 37, 40, 59, 62, 70, 86).

The Flexible Forward trials (Raudenbush et al., 2020) contained one practice trial followed by six test trials. On the practice trial, the experimenter demonstrated how to count starting at three by counting from three to eight, and then asked the child to count starting at three. On each test trial, children were asked to start counting from the given number and that the experimenter would tell them when to stop (test trials, administered in this order: 5 to 7, 18 to 21, 6 to 12, 4 to 9, 21 to 24, 35 to 39).

The Flexible Backward trials (Raudenbush et al., 2020) contained a demonstration trial and six test trials. First, the experimenter demonstrated by counting from ten to one. On each test trial, children were asked to start counting backwards from the given number and that the experimenter would tell them when to stop (test trials, administered in this order: 5 to 1, 8 to 4, 26 to 22, 13 to 9, 7 to 3, 18 to 14). Additionally, to ensure children remembered to count backwards, the experimenter started the sequence for them by providing the first two numbers (e.g., on the 5 to 1 trial: "Can you count backwards from five? Like this, five, four…").

The Skip counting trials contained a demonstration trial and three test trials. First, the experimenter demonstrated counting by 3s from three to fifteen (i.e., 3, 6, 9, 12, 15). On each test trial, children were asked to count by a given number and the experimenter told them when to stop (test trials, administered in this order: by 10s from 10 to 50, by 5s from 5 to 25, by 2s from 2 to 10). As with backward counting, children were provided with the first two numbers in the sequence (e.g., "Can you count by 10s? Like this, ten, twenty…").

On the Flexible Forward, Flexible Backward, and Skip counting trials, the experimenter typed the child's response verbatim and live scored them as correct using a check box for analysis. Children were only scored correct if they started from the correct number (or the next number after the experimenter's lead) and counted all numbers correctly in the target sequence. Throughout, the experimenter did not provide specific feedback on test trials but sometimes provided general encouragement to keep children engaged.

Analysis Plan and Predictions

Our primary pre-registered analysis is a repeated measures ANOVA with Relation Block (2: Ordinal words vs. Magnitude words) and Trial Type (2: consecutive number vs. non-consecutive number) as within-subject comparisons on children's proportion correct on the critical directional trials, which ask about numbers on the correct side of the target but only varied in whether they were consecutive or not. To further interpret the results of this analysis, we also report comparisons to chance (although due to an oversight, the comparisons to chance were not included in the pre-registration). These analyses allowed us to test the question of whether children's judgements show distinct patterns of dependence on the numerical distance of symbolic numbers for ordinal and magnitude judgements. We predicted that they would, resulting in an interaction between Relation and Trial Type, with the Ordinal Block showing better performance (i.e., more “yes” responses) on the consecutive trials than the non-consecutive trials (in line with reverse distance effects for ordinal judgements; Lyons & Ansari, 2015; Turconi et al., 2006), and with the Magnitude Block showing a non-significant difference (given that a 2-fold change is typically easily discriminable in infancy; Cordes & Brannon, 2008) or a difference in the opposite direction (in line with typical distance effects for magnitude judgements; Moyer & Landauer, 1973). This analysis does not include the directionally incorrect trials because our hypothesis about children’s narrow interpretation of ordinal language does not make predictions about trials that are directionally inconsistent and are therefore incorrect for reasons that are irrelevant to consecutiveness. However, we do provide a brief descriptive analysis of these directionally incorrect trials to give a sense of children’s performance on unambiguous trials.

As another approach to addressing this question, we also coded children’s responses on the free-response questions of “how many numbers come after five?” and “how many numbers are bigger than five?”. We pre-registered coding children’s responses as suggesting either only a single number or more than one number satisfies the relation. Children were categorized as indicating a single number if they explicitly said “one” or if they said “six”, as they may have been answering the question of *which number* rather than *how many.* Children were categorized as indicating more than one number if they provided a single number larger than ten (e.g., a hundred), an approximate amount (e.g., a lot), or a list of numbers (e.g., 6, 7, 8, 9, 10). Smaller numbers (less than or equal to ten) other than one and six (e.g., “eight”), ambiguous responses (e.g., “the higher numbers”), and non-responses (including “I don’t know”) were not included in these categories. Smaller numbers other than one and six were not included because it is unclear whether children were answering *how many* or instead providing an example of a number that fit the relation. Given the variability of children's responses within these broader categories and the difficulty of interpreting some of their responses, a description of more detailed coding within each category and a breakdown of the data based on these narrower subcategories is provided in the Supplemental Materials. Two coders coded all responses and agreed on 95% (6/120 disagreement). Disagreements were discussed and recoded.

As secondary analyses, we also planned a priori to investigate relations between children's performance with the relational vocabulary and (a) their age in months and (b) their counting ability. In terms of children's age, we did not pre-register a specific analysis plan or predictions. However, to provide some initial exploration of age-related patterns we analyzed correlations between children's age in months and their performance on the ordinal and magnitude blocks, for both consecutive and non-consecutive trials. To explore the relations with counting, we pre-registered using correlations between performance on the ordinal and magnitude blocks and the two counting measures: What's Next Task (Cronbach's alpha = 0.91) and Flexible Counting composite (proportion of trials correct on the remaining three counting tasks calculated as the total number correct out of the 15 total trials administered; Cronbach's alpha = 0.92). For each correlation (involving either age or counting), we have sensitivity (with n = 60 and 80% power) to detect correlations as small as .35, and as such may be underpowered to detect small correlations.

Results and Discussion

First, we analyzed performance on the directionally inconsistent trials where the correct answer was "*no*" (e.g., does 5 come after 7?). Overall, children did fairly well on these trials on both the ordinal block, $M = 0.76$, $SD = 0.28$, and the magnitude block, $M = 0.74$, $SD = 0.33$, performing above chance on both, ps < 0.001, and not significantly differently across the two blocks, $p = 0.73$. Children's relatively high performance on these trials suggests they did have at least some understanding of the vocabulary words being tested.

Do children's judgements with ordinal and magnitude vocabulary show distinct patterns of dependence on the consecutiveness of the numbers being compared?

Positive Verifications on Directionally Consistent Trials. To test our primary hypothesis, we analyzed the proportion of trials on which children responded "*yes*" when the number was consistent with the probed relational direction (i.e., it was bigger than the target number when asking about after/bigger or was smaller than the target number when asking about before/smaller) using a 2 (Trial Type: Consecutive, Non-Consecutive) x 2 (Relation: Ordinal, Magnitude) repeated measures ANOVA (see Figure 5).

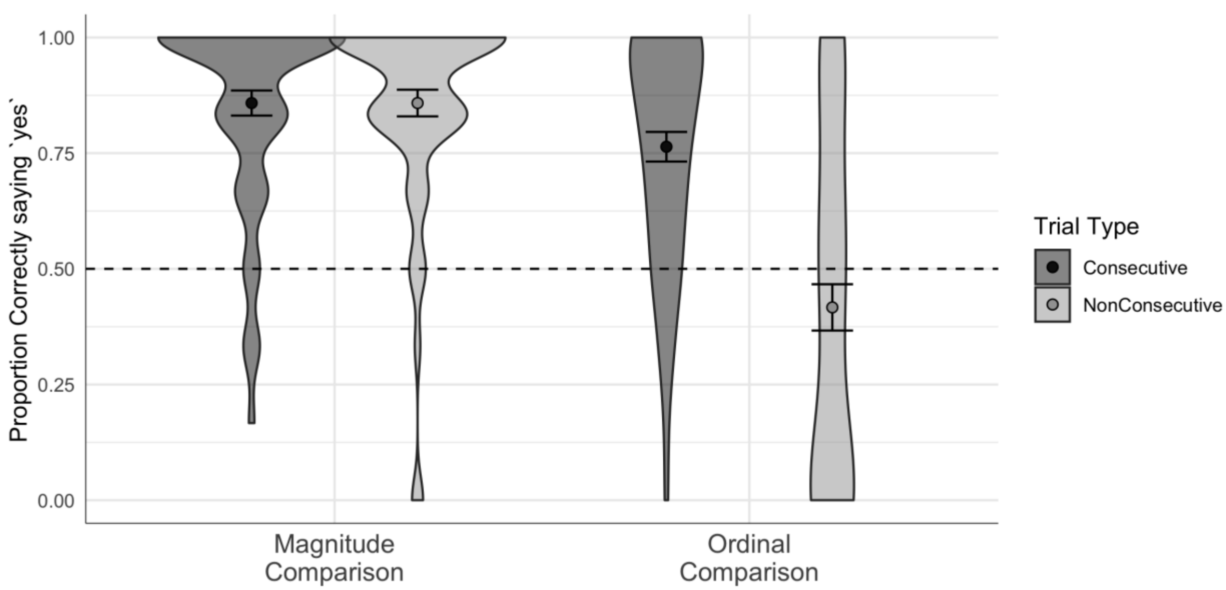

Figure 5: Children's performance in Experiment 2 on the Magnitude and Ordinal blocks (separated on the x-axis), for the consecutive (dark gray) and non-consecutive (light grey) trials, when those trials are directionally consistent (e.g., "is four after three?"). Points are mean performance, error bars are standard error of the mean, and kernel density violin plots show the distribution of the underlying data. Dotted line is at the chance level of 0.5.

The analysis revealed a main effect of Trial Type, $F(1, 59) = 36.7$, $p < 0.001$, $\eta^2_{partial} = 0.38$, a main effect of Relation, $F(1, 59) = 54.3$, $p < 0.001$, $\eta^2_{partial} = 0.48$, and an interaction between Trial Type and Relation, $F(1, 59) = 41.0$, $p < 0.001$, $\eta^2_{partial} = 0.41$. Given the interaction, we compared performance on the consecutive and non-consecutive trials within the ordinal and magnitude relations separately. When asked about order, children said the consecutive numbers came before/after the target number, $M = 0.76$, $SD = 0.25$, significantly more frequently than the non-consecutive numbers, $M = 0.42$, $SD = 0.39$, $t(59) = 7.1$, $p < 0.001$, $d = 0.92$. However, when asked about the magnitude of the numbers, children performed almost identically regardless of whether the numbers were consecutive, $M = 0.86$, $SD = 0.21$, or non-consecutive, $M = 0.86$, $SD = 0.22$, $t(59) < 0.001$, $p = 1$, $d < 0.001$.

Moreover, overall performance was significantly above chance when judging magnitude of both consecutive and non-consecutive numbers, $p < 0.001$. However, on the ordinal trials, performance was above chance for consecutive numbers, $p < 0.001$, and not significantly different from chance on non-consecutive numbers, $p = 0.10$.

Thus, in line with our hypothesis, children's performance reveals a narrow interpretation of before and after to be *immediately* before or after, despite knowing that both consecutive and non-consecutive numbers share the same magnitude relations. Next, we tested whether this same pattern is evident in children's explicit verbal responses about the same relations.

Open Ended Responses. The number of children whose responses fall into each of our categories are presented in Table 2. About a third of the children's responses were categorized as no response or ambiguous, and these children are not considered further in our analyses. When considering the two categories of interest (Single Number vs. Multiple Numbers), fewer than a quarter of the children indicated that only a single number was bigger than five (10%) or came after five (22%). On the other hand, around half the children responded in a way consistent with a belief that multiple numbers could be after five (47%) or bigger than five (52%). A chi-squared contingency table test on the two categories of interest between the ordinal (after) and magnitude (bigger) relations was not significant, $\chi^2 = 1.76$, $p = 0.18$. Thus, unlike their behavioral responses, children's verbal responses suggest that they did not have an overwhelming tendency to say that only a single number came after or was bigger than five and their pattern of responses did not significantly differ for ordinal or magnitude relations.

Table 2: The Distribution of Children's Open-Ended Responses

	"How many numbers come *after* five?"		"How many numbers are *bigger* than five?"	
Response Category	*n*	proportion	n	proportion
Single Number	13	0.22	6	0.10
Multiple Numbers	28	0.47	31	0.52
No Response or Ambiguous Response	19	0.32	23	0.38

Together, these findings suggest that many children provide a verbal response indicating that many numbers come after 5, but still respond on ordinal verification judgements with the vocabulary before/after as if the number 6, but not the number 7, comes after 5. In order words, children's behavior reflects a systematically narrow interpretation of the ordinal vocabulary words before/after, regardless of whether they articulated this narrow interpretation.

How does age relate to children's knowledge of relational vocabulary?

As secondary analyses, we were also interested in how children's narrow interpretation of order may vary across development. Children were evenly sampled across the three annual age bins within the larger sample (20 each of 4-, 5-, and 6-years-olds) and we analyzed continuous age in months.

As is evident from Figure 6, the difference in children's scores on consecutive and non-consecutive directionally consistent trials significantly increased across the age range tested: correlation between age and Consecutive – Non-Consecutive Score, $r = 0.44$, $p < 0.001$. Specifically, children's scores significantly increased on consecutive trials, $r = 0.33$, $p = 0.01$, but there was a negative, although not significant, relation between age and scores on the non-consecutive trials, $r = -0.22$, $p = 0.09$.

This pattern was not found for magnitude scores, which were relatively high across the age group tested and did not significantly correlate with age: difference score $r = -0.01$, $p = 0.97$; consecutive trials, $r = 0.13$, $p = 0.32$; non-consecutive trials, $r = 0.13$, $p = 0.33$.

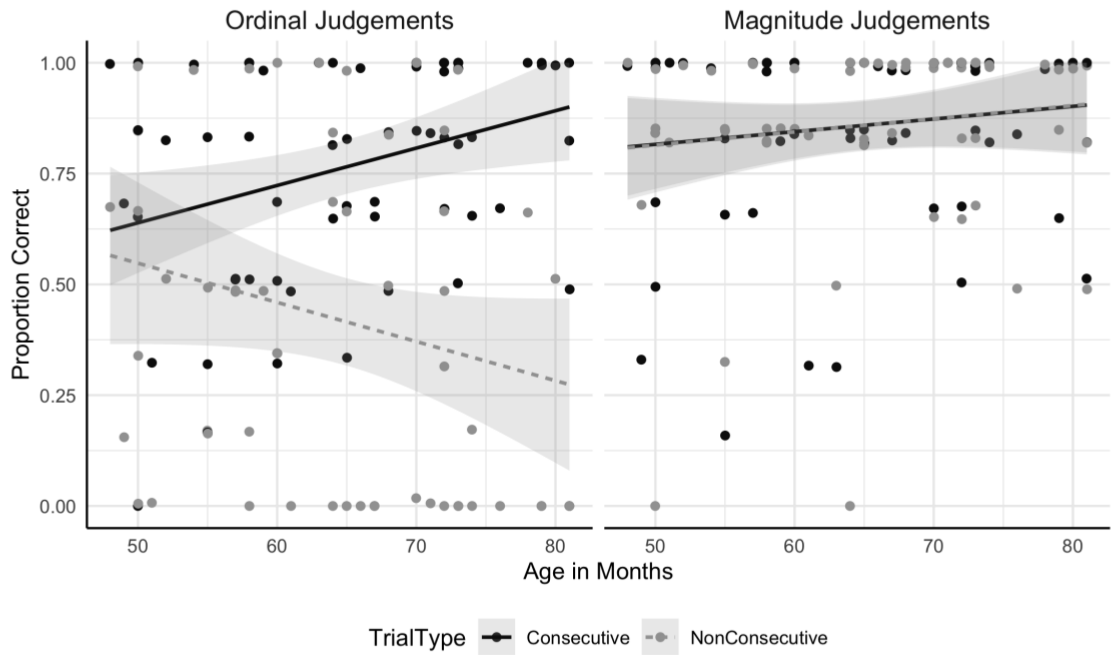

Figure 6: Correlation between children's age in months (x-axis) and their scores on the ordinal (left panel) and magnitude (right panel) judgement tasks, separated by consecutive (grey, dashed) and non-consecutive (black, solid) trials. Only directionally consistent trials are included. Some jitter has been added to the points to view overlapping data more easily. Note that the two lines on the plot of the magnitude judgements are very overlapping, making it difficult to see them separately.

How does counting ability relate to children's knowledge of relational language?

One child was excluded from these analyses because they did not complete the counting task, leaving a sample of 59 children. To investigate the relation between counting ability and children's relational knowledge, we analyzed bivariate correlations between children's scores on the ordinal block (on the three trial types separately: Directionally Correct & Consecutive, Directionally Correct & Non-consecutive, and Directionally Incorrect) and both the What's Next measure and the Flexible Counting measure (i.e., the composite of flexible forward, backward, and skip counting[3]). To investigate whether these relations were unique to ordinal relations, we used the same analyses with scores from the magnitude trials. Table 3 displays these correlations and the descriptive statistics of the counting scores.

Table 3: Correlation between the counting tasks and relational judgements

			Ordinal Comparisons			Magnitude Comparisons		
			Directionally Correct		Directionally Incorrect	Directionally Correct		Directionally Incorrect
	Mean (SD)		C	Non-C	Combined C & Non-C	C	Non-C	Combined C & Non-C
What's Next Task	0.59 (0.38)	*r*	.38**	-.11	.54***	.29*	.09	.51***
		p	.003	.39	< 0.001	.03	.48	< .001
Flexible Counting	0.58 (0.32)	*r*	.33**	-.16	.62***	.31*	.15	.54***
		p	.01	.22	< 0.001	.02	.26	< 0.001

Note: C = consecutive trials, Non-C = Non-Consecutive trials; Trials assessing the incorrect polar direction are collapsed across consecutiveness because there are only three trials of each type
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

First, both the What's Next task and the Flexible Counting composite scores positively correlated with children's ability to judge directionally correct consecutive numbers in terms of ordinal vocabulary (i.e., that 5 does come after 4 or 3 does come before 4), but neither significantly correlated with children's performance on directionally correct non-consecutive trials. These correlations of the counting tasks with the ordinal consecutive and non-consecutive correct trials significantly differed from each other: What's Next: z = 3.61, $p < 0.001$, Flexible Counting: z = 3.60, $p < 0.001$. It is also worth noting that although not significant, the

[3] When each component of the composite Flexible Counting measure is analyzed separately, we find a similar pattern as is reported for the overall composite. This suggests that combining the measures in this way does not obscure distinct patterns across trial types.

correlations with children's ability to judge non-consecutive trials in terms of ordinal vocabulary are numerically negative. Finally, there were also large significant correlations with children's scores on the directionally inconsistent trials for ordinal vocabulary, where children needed to say "*no*" (e.g., does 4 come after 5?).

A similar pattern was found for judgements of magnitude, with significant correlations between counting and consecutive magnitude trials, but not non-consecutive magnitude trials. However, in contrast to the ordinal judgements, all the correlations are numerically positive, and the size of the correlation did not significantly differ between consecutive and non-consecutive trials: What's Next, $z = 1.62$, $p = 0.10$, or Flexible Counting, $z = 1.33$, $p = 0.18$. There were also large significant correlations with children's scores on the directionally inconsistent trials.

Thus, it appears that children's counting knowledge positively predicts their ability to correctly reject directionally inconsistent relations and positively verify directionally consistent relations for *consecutive* values, for both ordinal and magnitude relations. Moreover, there are slight, albeit non-significant, negative correlations between counting and ordinal judgements on the non-consecutive trials.

Summary of Experiment 2

In Experiment 2, our primary goal was to investigate whether children's judgements with ordinal and magnitude vocabulary differentially depended on whether the numbers being compared are consecutive. Overall, we find that they do – children's ordinal judgements varied for consecutive vs. non-consecutive numbers, while their magnitude judgements did not. Specifically, children did not endorse the use of the words *before* or *after* for non-consecutive numbers but did endorse the use of the magnitude words *bigger* and *smaller* regardless of consecutiveness.

General Discussion

Across two experiments, we investigated children's understanding of vocabulary words for mathematical relations involving magnitude comparison and order. We find that when comparing quantities and amounts, children demonstrate lower understanding of the ordinal vocabulary words *before* and *after* compared to magnitude vocabulary words *bigger*, *smaller*, *more*, and *less* across all three quantitative contexts: symbolic number, discrete non-symbolic number, and continuous area (Experiment 1). Further, in the context of symbolic numbers, children's errors across ordinal and magnitude vocabulary show systematically distinct patterns, with children applying ordinal vocabulary more narrowly than magnitude vocabulary (Experiment 2). Together, these findings suggest some divergence between children's understanding of relational vocabulary words for ordinal and magnitude comparisons.

These findings identify a salient difference but leave open the question of what causes this divergence. We discuss three possible explanations, which are neither exhaustive nor mutually exclusive, but can set the foundation for generating several important hypotheses and future directions for better understanding the development of children's ordinal and magnitude knowledge: (1) children's underlying conception of magnitude and (conventionalized) order are at least somewhat independent, (2) children learn to interpret relational language referring to ordinal and magnitude relations in distinct ways, and (3) children have an understanding of language referring to magnitude and ordinal relations, but have difficulty extracting relevant magnitude and/or ordinal information when needed because aspects of their number knowledge are weighted differently when interpreting vocabulary words for magnitude versus ordinal judgements. We discuss each of these possible explanations in turn.

First, it may be that children's underlying number knowledge does not include well integrated concepts of relative magnitude and order, and instead these two relations are at least somewhat independent. This explanation is consistent with other work comparing people's understanding of magnitude and ordinal relations in the absence of specific relational vocabulary, which similarly reveal that consecutive numbers are privileged over non-consecutive numbers for ordinal processing (Turconi et al., 2006) and demonstrate divergences in terms of how children's performance judging each relation predicts their later math ability (Lyons et al., 2014). Notably, however, this divergence between magnitude and order may specifically be in terms of *conventionalized* numerical order. Given the early emergence of ordering magnitudes (e.g., Anderson & Cordes, 2013) and that typical theories of numerical magnitude represent number on a well ordered continuum (Dehaene, 2011; Dehaene et al., 1998), we are not arguing that the representation of magnitude values does not have an ordered component. Rather, conventional order in the context of symbolic numbers and a learned count list may lead to divergence of ordinal and magnitude concepts in the context of symbolic number, which in turn account for the different patterns seen for the ordinal and magnitude relational verification task. Future work could investigate how children's conceptions of conventionalized numerical order and magnitude-based ordinal relations are related both to each other and to other aspects of early number knowledge, including magnitude comparison and counting.

Second, given that our paradigm was specifically focused on children's understanding of comparative vocabulary, a natural explanation of the results is in terms of children's understanding of these words. Although Experiment 1 revealed lower understanding of ordinal vocabulary compared to magnitude vocabulary in general, Experiment 2 suggests a systematic difference in how children interpret ordinal vocabulary compared to magnitude vocabulary. It may be that children's narrow interpretation of ordinal vocabulary reflects an immature understanding of these words, relative to magnitude words. Notably, given that children are not learning the words *before* and *after* for the first time in a numerical context, but instead are applying these already-known temporal words to numerical information, an important next question is whether this immature understanding of ordinal vocabulary is unique to symbolic number or if it is evident in temporal contexts as well. If it is the former, and this pattern of interpreting ordinal vocabulary is unique to numerical contexts, then what is it about numerical contexts that leads to these systematic patterns? One possibility is that the conventional order of numerical information through the count list leads to a narrow interpretation of ordinal vocabulary. For example, we find that children's counting knowledge was related to their judgements of consecutive numbers, but not their judgements of non-consecutive numbers, which might indicate that counting, and even flexible counting, highlights the consecutiveness of numbers. However, similar relations to counting were found for magnitude vocabulary as well, leaving open the question of why children do not generalize this same narrow interpretation to both ordinal and magnitude vocabulary. Future work could investigate children's interpretation of ordinal vocabulary across a range of contexts (e.g., temporal events, the alphabet, monotonically ordered continuous area) to better understand how children apply the same ordinal vocabulary across different contexts that have different degrees of conventionalized order versus magnitude-based order, as well as what aspects of children's experience (e.g., parents use of ordinal and magnitude vocabulary, counting knowledge or experience) might be impacting their learning of these words.

Finally, a third possibility is that children rely on different strategies for extracting magnitude and ordinal information when needed, leading to performance differences across magnitude and ordinal contexts. Because magnitude and order may be intrinsically connected in the context of symbolic number, children's understanding of "more" when comparing two symbolic numbers is typically interpreted as ordinal knowledge, or understanding of the "later-greater" principle (e.g., Davidson et al., 2012; Le Corre, 2014). It may be that children's ability to judge relative magnitude, and in this case interpret magnitude vocabulary, draws upon this "later-greater" concept allowing them to seamlessly integrate both pieces of information: "later" (i.e., understanding what numbers are later in the count list) and "greater" (i.e., understanding the relative magnitudes of those numbers). In contrast, children may rely on other strategies when asked about ordinal relations, such as strategies that more directly rely on conventional order, including counting or successor knowledge (i.e., knowing the next number) and these different strategies result in different patterns of performance. Although the age-related patterns found in the current study should be interpreted with caution, they do generate some possible hypotheses that could be tested in future work. Research on the successor function suggests that it is typically mastered at around 5.5 to 6-years-old (e.g., Davidson et al., 2012; Schneider et al., 2020), around the same age that the children in our sample robustly applied ordinal vocabulary to discrete non-symbolic sets (Experiment 1) and showed a stronger narrow interpretation of ordinal vocabulary in symbolic contexts (Experiment 2). Future work could investigate whether the acquisition of the successor function is related to children's narrow interpretation of ordinal vocabulary and/or to their strategy use on other ordinal and magnitude tasks, and if so, what the causal direction might be and how it could be leveraged to support learning.

Limitations

The current study is of course not without limitations. First, most children in our samples came from high-income households with at least one college educated parent, which limits both our ability to investigate the effect of these factors (e.g., socioeconomic status) on children's understanding of relational vocabulary and our ability to generalize these findings to children with different experiences, particularly with different exposure to early number concepts. Second, although Experiment 1 was conducted in person using traditional developmental psychology approaches (e.g., museums, lab testing), Experiment 2 was conducted entirely online through video chat during the COVID-19 pandemic. As the field continues to explore online data collection, it is important to keep in mind the benefits, pitfalls, and best practices of online data collection in general (e.g., Lourenco & Tasimi, 2020; Sheskin et al., 2020), as well as its role during a pandemic, which may offer unique challenges for both logistics and generalizability. A third limitation is our sample size. Although preregistered and sufficient to detect approximately medium effects on our primary analyses of interest, the sample sizes limited our ability to detect smaller effects and explore other individual differences across subgroups. Most notably, substantial number learning typically occurs during the relatively narrow age range used in the current experiments (4 to 6 or 7 years old), making age-related patterns in our tasks particularly interesting. However, our sample size limited our ability to investigate developmental patterns, leaving it to future work to test our proposed age-related hypotheses more fully. Finally, and relatedly, the current study was entirely correlational, again limiting our ability to make any claims about the causal nature of children's understanding of relational language. Importantly, however, the current study lays the foundation for investigating causal relations by generating several hypotheses that could be tested in future work.

Conclusion

In summary, we investigated 4- to 7-year-old children's knowledge of ordinal and magnitude vocabulary for comparing quantities, both symbolically and non-symbolically. We find evidence of both general and systematic differences between children's performance judging quantities in terms of ordinal language versus magnitude language, suggesting a dissociation between children's understanding of these two concepts and/or the language used to communicate about them. These findings suggest promising new directions for investigating the potentially privileged status of consecutive numbers in children's early numerical reasoning, the role of counting knowledge in this reasoning, and what role language might play in children's learning of numerical relations, including how magnitude and ordinal relations are related to each other.

References

Amidon, A., & Carey, P. (1972). Why Five-Year-Olds Cannot Understand Before and After. *Journal of Verbal Learning and Verbal Behavior*, *11*, 417–423.

Anderson, U. S., & Cordes, S. (2013). 1<2 and 2<3: Non-Linguistic Appreciations of Numerical Order. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00005

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, *83*(3), 223–240.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the Numerosities 1 to 9 by Monkeys. *Science*, *282*(5389), 746–749. https://doi.org/10.1126/science.282.5389.746

Brannon, E. M., & Van de Walle, G. A. (2001). The Development of Ordinal Numerical Competence in Young Children. *Cognitive Psychology*, *43*(1), 53–81. https://doi.org/10.1006/cogp.2001.0756

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/joc.72

Buckley, P. B., & Gillman, C. B. (1974). COMPARISONS OF DIGITS AND DOT PATTERNSl. *Journal of Experimental Psychology*, *103*(6), 1131–1136.

Bullock, M., & Gelman, R. (1977). Numerical Reasoning in Young Children: The Ordering Principle. *Child Development*, *48*(2), 427–434.

Cantlon, J. F., & Brannon, E. M. (2006). Shared System for Ordering Small and Large Numbers in Monkeys and Humans. *Psychological Science*, *17*(5), 401–406. https://doi.org/10.1111/j.1467-9280.2006.01719.x

Cantlon, J. F., Fink, R., Safford, K., & Brannon, E. M. (2007). Heterogeneity impairs numerical matching but not numerical ordering in preschool children. *Developmental Science*, *10*(4), 431–440. https://doi.org/10.1111/j.1467-7687.2007.00597.x

Cheung, P., Rubenson, M., & Barner, D. (2017). To infinity and beyond: Children generalize the successor function to all possible numbers years after learning to count. *Cognitive Psychology*, *92*, 22–36. https://doi.org/10.1016/j.cogpsych.2016.11.002

Christie, S., & Gentner, D. (2014). Language Helps Children Succeed on a Classic Analogy Task. *Cognitive Science*, *38*(2), 383–397. https://doi.org/10.1111/cogs.12099

Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 266–275. https://doi.org/10.1016/S0022-5371(71)80054-3

Cordes, S., & Brannon, E. M. (2008). Quantitative competencies in infancy: Quantitative competencies in infancy. *Developmental Science*, *11*(6), 803–808. https://doi.org/10.1111/j.1467-7687.2008.00770.x

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*(1), 162–173. https://doi.org/10.1016/j.cognition.2011.12.013

de Hevia, M. D., & Spelke, E. S. (2010). Number-Space Mapping in Human Infants. *Psychological Science*, *21*(5), 653–660. https://doi.org/10.1177/0956797610366091

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. *Attention & Performance XXII. Sensori-Motor Foundations of Higher Cognition, Ed. P. Haggard & Y. Rossetti*, *52774*.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361. https://doi.org/10.1016/S0166-2236(98)01263-6

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Gathercole, V. C. (1985). More and more and more about more. *Journal of Experimental Child Psychology*, *40*(1), 73–104. https://doi.org/10.1016/0022-0965(85)90066-9

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Harvard University Press.

Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 32–44. https://doi.org/10.1037/0096-1523.32.1.32

Henry, L., & Wickham, H. (2019). *purrr: Functional Programming Tools* (0.3.2) [Computer software]. https://CRAN.R-project.org/package=purrr

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hornburg, C. B., Schmitt, S. A., & Purpura, D. J. (2018). Relations between preschoolers' mathematical language understanding and specific numeracy skills. *Journal of Experimental Child Psychology*, *176*, 84–100. https://doi.org/10.1016/j.jecp.2018.07.005

Huang, Y. T., Spelke, E., & Snedeker, J. (2010). When Is Four Far More Than Three? *Psychological Science*, *21*(4), 600–606.

Iannone, R., Cheng, J., & Schloerke, B. (2020). *gt: Easily Create Presentation-Ready Display Tables* (0.2.2) [R]. https://CRAN.R-project.org/package=gt

Ip, M. H. K., Imuta, K., & Slaughter, V. (2018). Which Button Will I Press? Preference for Correctly Ordered Counting Sequences in 18-Month-Olds. *Developmental Psychology*, *54*(7), 1199–1207.

Johnson, H. L. (1975). The meaning of before and after for preschool children. *Journal of Experimental Child Psychology*, *19*(1), 88–99. https://doi.org/10.1016/0022-0965(75)90151-4

Kassambara, A. (2020). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (0.4.0) [Computer software]. https://CRAN.R-project.org/package=rstatix

Le Corre, M. (2014). Children acquire the later-greater principle after the cardinal principle. *British Journal of Developmental Psychology*, *32*(2), 163–177. https://doi.org/10.1111/bjdp.12029

Lourenco, S. F., & Tasimi, A. (2020). No Participant Left Behind: Conducting Science During COVID-19. *Trends in Cognitive Sciences*, *24*(8), 583–584. https://doi.org/10.1016/j.tics.2020.05.003

Lyons, I. M., & Ansari, D. (2015). Numerical Order Processing in Children: From Reversing the Distance-Effect to Predicting Arithmetic: Development of Ordinality. *Mind, Brain, and Education*, *9*(4), 207–221. https://doi.org/10.1111/mbe.12094

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–261. https://doi.org/10.1016/j.cognition.2011.07.009

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. https://doi.org/10.1111/desc.12152

Mackenzie, I. G. (2018). *psychReport: Reproducible Reports in Psychology.* (R package version 0.4) [Computer software]. https://CRAN.R-project.org/package=psychReport

Marschark, M. (1977). Lexical Marking and the Acquisition of Relational Size Concepts. *Child Development*, *48*(3), 4.

Moyer, R. S., & Landauer, T. K. (1973). Determinants of reaction time for digit inequality judgments. *Bulletin of the Psychonomic Society*, *1*(3).

Ohshiba, N. (1997). Memorization of Serial Items by Japanese Monkeys, a Chimpanzee, and Humans. *Japanese Psychological Research*, *39*(3), 236–252. https://doi.org/10.1111/1468-5884.00057

Palermo, D. S. (1973). More about less: A study of language comprehension. *Journal of Verbal Learning and Verbal Behavior*, *12*(2), 211–221. https://doi.org/10.1016/S0022-5371(73)80011-8

Picozzi, M., de Hevia, M. D., Girelli, L., & Macchi Cassia, V. (2010). Seven-month-olds detect ordinal numerical relationships within temporal sequences. *Journal of Experimental Child Psychology*, *107*(3), 359–367. https://doi.org/10.1016/j.jecp.2010.05.005

Powell, S. R., & Nelson, G. (2017). An Investigation of the Mathematics-Vocabulary Knowledge of First-Grade Students. *The Elementary School Journal*, *117*(4), 664–686. https://doi.org/10.1086/691604

Purpura, D. J., Napoli, A. R., Wehrspann, E. A., & Gold, Z. S. (2017). Causal Connections Between Mathematical Language and Mathematical Knowledge: A Dialogic Reading Intervention. *Journal of Research on Educational Effectiveness*, *10*(1), 116–137. https://doi.org/10.1080/19345747.2016.1204639

Purpura, D. J., & Reid, E. E. (2016). Mathematics and language: Individual and group differences in mathematical language skills in young children. *Early Childhood Research Quarterly*, *36*, 259–268. https://doi.org/10.1016/j.ecresq.2015.12.020

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Studio Team. (2016). *RStudio: Integrated Development for R*. RStudio Inc. http://www.rstudio.com/

Raudenbush, S. W., Hernandez, M., Goldin-Meadow, S., Carrazza, C., Foley, A., Leslie, D., Sorkin, J. E., & Levine, S. C. (2020). Longitudinally adaptive assessment and instruction increase numerical skills of preschool children. *Proceedings of the National Academy of Sciences*, *117*(45), 27945–27953. https://doi.org/10.1073/pnas.2002883117

Russell, B. (1903). *The Principles of Mathematics*. Allen & Unwin. https://philpapers.org/rec/RUSTPO-77

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662–674. https://doi.org/10.1016/j.cognition.2008.05.007

Schneider, R. M., Sullivan, J., Guo, K., & Barner, D. (2020). *What counts? Sources of knowledge in children's acquisition of the successor function* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/vu47r

Schneider, R. M., Sullivan, J., Marušič, F., Žaucer, R., Biswas, P., Mišmaš, P., Plesničar, V., & Barner, D. (2019). Do Children Use Language Structure to Discover the Recursive Rules of Counting? *Cognitive Psychology*. https://doi.org/10.31234/osf.io/cf9nm

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online Developmental Science to Foster Innovation, Access, and Impact. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2020.06.004

Spaepen, E., Gunderson, E. A., Gibson, D., Goldin-Meadow, S., & Levine, S. C. (2018). Meaning before order: Cardinal principle knowledge predicts improvement in understanding the successor principle and exact ordering. *Cognition*, *180*, 59–81. https://doi.org/10.1016/j.cognition.2018.06.012

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences*, *110*(45), 18116–18120. https://doi.org/10.1073/pnas.1302751110

Suanda, S. H., Tompson, W., & Brannon, E. M. (2008). Changes in the Ability to Detect Ordinal Numerical Relationships Between 9 and 11 Months of Age. *Infancy*, *13*(4), 308–337. https://doi.org/10.1080/15250000802188800

Terrace, H. S., & McGonigle, B. (1994). Memory and Representation of Serial Order by Children, Monkeys, and Pigeons. *Current Directions in Psychological Science*, *3*(6), 180–185. https://doi.org/10.1111/1467-8721.ep10770703

Torchiano, M. (2018). *effsize: Efficient Effect Size Computation.* https://CRAN.R-project.org/package=effsize

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

Tzelgov, J., & Ganor-Stern, D. (2005). Automaticity in Processing Ordinal Information. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition*. Psychology Press.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Wickham, H. (2019). *stringr: Simple, Consistent Wrappers for Common String Operations* (1.4.0) [R package]. https://CRAN.R-project.org/package=stringr

Wickham, H. (2020). *forcats: Tools for Working with Categorical Variables (Factors)* (0.5.0) [R package]. https://CRAN.R-project.org/package=forcats

Wickham, H., & Bryan, J. (2019). *readxl: Read Excel Files* (1.3.1) [Computer software]. https://CRAN.R-project.org/package=readxl

Wickham, H., Francois, R., Henry, L., & Muller, K. (2018). *dplyr: A Grammar of Data Manipulation* (0.7.7) [Computer software]. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2018). *tidyr: Easily Tidy Data with "spread()" and "gather()" Functions* (0.8.1) [Computer software]. https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & Francois, R. (2018). *readr: Read Rectangular Text Data* (1.3.1) [Computer software]. https://CRAN.R-project.org/package=readr

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

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8*(1), 88–101. https://doi.org/10.1111/j.1467-7687.2005.00395.x