Children's interpretation of "before" and "after" for consecutive and non-consecutive numbers and events

Michelle A. Hurst[1], Paige Dadika[1], Madeleine Oswald[2], & Susan C. Levine[2]

[1] Rutgers University, New Brunswick

[2] University of Chicago

This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Cognition and Development in Sept 2025, available at: https://doi.org/10.1080/15248372.2025.2560861

Author note

Following the credit taxonomy (https://credit.niso.org/), we outline the roles and responsibilities of each author. In addition, we want to thank the research assistants involved in data collection (i.e., Investigation): Emily Kim, Emily Pardo, and Leslie Villaverde.

The authors made the following contributions. Michelle A. Hurst: Conceptualization, Data analysis, Investigation, Project administration, Supervision, Writing - Original Draft Preparation, Review, and Editing; Paige Dadika: Data curation, Formal analysis, Visualization, Writing - Original Draft Preparation, Review, and Editing; Madeleine Oswald: Conceptualization, Formal analysis, Investigation, Supervision, Writing - Original Draft Preparation, Review, and Editing; Susan C. Levine: Conceptualization, Funding acquisition, Resources, Supervision, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Michelle A. Hurst, 152 Frelinghuysen Rd, Piscataway, New Jersey. E-mail: michelle.hurst@rutgers.edu

Abstract

When learning the integers, children must learn that number words are in a stable order and that this order is tied to magnitude. In previous work, US English speaking 5- and 6-year-old children's interpretation of ordinal vocabulary (after, before) to compare numbers in the count list was restricted to consecutive numbers (e.g., 6 is after 5, but 7 is not), despite readily applying magnitude vocabulary (i.e., both 6 and 7 are bigger than 5). In the current study, we investigate whether this narrow interpretation of before/after is a more general feature of ordinal comparisons by comparing it to children's interpretation of before/after in temporal event sequences (Experiment 1, N = 120) and to adults' interpretation of before/after for both numerical and temporal sequences (Experiment 2, N = 120). First, we replicate the original finding: US 5- and 6-year-old children are more likely to endorse using before/after for consecutive numbers than non-consecutive numbers. Second, we do not find the same pattern in children's judgements of event sequences, suggesting that it is caused by a feature of the count list that is not shared by event sequences. Third, we find that US adults predominantly endorse before/after for both consecutive and non-consecutive comparisons, though there is some variation consistent with children's responding. These findings suggest that the narrow interpretation of before and after in US children's numerical judgements is not due to their interpretation of ordinality more generally, and instead may be tied to a specific feature of numerical sequences.

Keywords: count list, number, ordinality, relational language

Children's interpretation of "before" and "after" for consecutive and non-consecutive numbers and events

Introduction

Relative order is a fundamental relation used to compare numerical quantities, temporal events, and other arbitrary and non-arbitrary sequences. In the case of numerical quantities, evidence suggests that ordinal relations are distinct from other relevant relations, such as magnitude. For example, children's ordinal and magnitude judgments each uniquely predict later math achievement (Holloway & Ansari, 2009; Lyons & Beilock, 2011; Lyons et al., 2014; Starr et al., 2013). Additionally, although knowledge of numerical order is often measured in the absence of specific relational vocabulary (e.g., are "3 4 5" in order?), we can – and often do – apply ordinal vocabulary terms to numerical information. For example, we can describe numerical sequences using temporal words like before, after, and later (e.g., the "later greater" principle), and sequential ordering words like next (e.g., "what number comes next?"). Importantly, a recent study demonstrated a distinction in children's reasoning about numerical magnitude and numerical order in terms of how they interpreted relational vocabulary (Hurst et al., 2022). Specifically, English speaking 4- to 6-year-old children in the US were unlikely to endorse sentences involving the terms before and after to reference non-consecutive numerical values (e.g., saying no to "does 7 come after 5?"), despite correctly endorsing the use of magnitude language to compare non-consecutive values (e.g., saying yes to "is 7 bigger than 5?"). Thus, these children seemed to hold a narrow consecutive-only (or primarily) interpretation of the words before and after when comparing symbolic numerical values. Given the relations between both children's ordinal knowledge of numbers (e.g., Lyons & Beilock, 2011; Lyons et al., 2014) and of mathematical vocabulary (e.g., Hornburg et al., 2018; Purpura et al., 2017; Purpura & Reid, 2016) with mathematical achievement, it is important to better understand both the causes and possible consequences of this narrow interpretation. In the current study, we begin addressing this question by investigating the generality of this narrow consecutive-only interpretation.

One possibility is that the consecutive-only interpretation of before and after is a more general property of children's interpretation of these words. There is a substantial literature on children's acquisition of ordinal vocabulary, including the words before and after, which has revealed several biases and errors in children's use and interpretation of these terms. For example, children may rely more on the order of mention in a sentence to guide their behavior, rather than correctly interpreting ordinal words (e.g., Clark, 1971; Johnson, 1975), focus only on the main clause of a sentence (e.g., Amidon & Carey, 1972), or default to a "next in line" strategy, in which they select the next event in time as the default (e.g., Coker, 1978). Although this prior work has not revealed a narrow interpretation of before and after as described here, that may be a function of the typical paradigms and procedures used. Many studies only include two events that must be ordered or score all possible preceding events (when prompted for before) or subsequent events (when prompted for after) as being correct, making it difficult to disentangle whether children have any bias for interpreting consecutive events differently from non-consecutive events.

An alternative possibility is that the consecutive-only interpretation of before and after previously reported in numerical sequences is due to something unique about numerical sequences. That is, it may be that children interpret the meaning of before and after differently in numerical sequences compared to temporal event sequences, where the terms before and after are likely to be initially encountered and to be discussed more frequently (e.g., “Before you go to sleep, you need to brush your teeth.”). Unlike temporal events, numerical sequences are tied to the count list and numerical magnitude, which then has several consequences that may lead to differences in children’s interpretation of ordinal vocabulary across numerical and temporal sequences. Children learn to recite the count list from memory, well before learning the meanings of the number words and their connections to magnitude (e.g., Carey, 2004; Fuson, 1988; Wynn, 1990, 1992). After learning the stable verbal count list, children learn that the order of numbers is tied to magnitude through the successor function: the next number is always one more than the previous. Importantly, this provides an external constraint on numerical order, tied to the magnitudes of the underlying numerical values (Davidson et al., 2012; Spaepen et al., 2018). Of course, numbers can be used as arbitrary symbols in other orders (e.g., house numbers, phone numbers), but the count list is tied to a specific numerical order governed by immediate successors. In contrast, events can occur in many different orders, some that are related to causal world structure (e.g., opening a container before reaching into it) and some that are not (e.g., touching your head and then touching your knees). Further, although people do tend to segment events into meaningful event chunks (Zacks & Swallow, 2007), this discreteness is not an inherent property of the event sequence itself, which has a more continuous temporal structure. Indeed, mothers appear to modify their actions in ways that help infants detect the underlying structure of dynamic events (Brand et al., 2002).

Previous work has shown that young children do not endorse non-consecutive numbers as being “in order” (Gilmore & Batchelor, 2021; Hutchison et al., 2022). When asked whether number sequences such as “3 4 5” or “2 4 6” are in order, 4- to 9-year-olds are above chance on consecutive numbers but at chance [6- to 9-year-olds; Gilmore and Batchelor (2021)] or below chance [4- and 5-year-olds; Hutchison et al. (2022)] with non-consecutive numbers. This narrow interpretation of ordinality may stem from children's knowledge of the verbal count sequence (Gilmore & Batchelor, 2021), and it may be that this narrow interpretation of ordinality in turn explains 4- to 6-year-old children's narrow interpretation of before and after when making ordinal judgements about number (Hurst et al., 2022). If this is the case, then children who have a consecutive-only interpretation of before and after for number should not have this interpretation for temporal event sequences, as these sequences do not have the same connection to counting.

In the current study, we address two research questions, with the goal of understanding which contexts give rise to children's consecutive-only interpretation of before and after. First, is the previously reported narrow interpretation of before and after also evident in temporal event sequences? If so, this would suggest that it may be a feature of the experimental paradigm and/or children's interpretation of these temporal words more generally, and not caused by features of numerical order specifically. In contrast, if the narrow interpretation is found in numerical sequences and not temporal event sequences, this would suggest there is a feature of numerical sequences driving this interpretation, such as children's understanding of numerical order or reliance on a sequential verbal count list. In Experiment 1, we had 5- and 6-year-old children, primarily in the US, watch videos of a person counting aloud or completing a series of actions. After the videos, we asked children questions about the order of the numbers (e.g., "was 6 after 5?") or the events (e.g., "was [screenshot of pouring juice from video] after [screenshot of inserting straw from video]?"). We expected to replicate prior work with numbers, where 5- and 6-year-old children relatively rarely apply before/after to non-consecutive numbers (Hurst et al., 2022). The critical question is how these children respond to analogous questions involving events from the series of actions. That is, would children endorse the use of before/after to describe the relations between non-consecutive events as well as between consecutive events? We chose to test 5- and 6-year-old children because prior work found that a consecutive-only interpretation of before/after is evident using a similar paradigm by around 5-year-olds (Hurst et al., 2022).

Second, do we see any evidence of this consecutive-only interpretation of before and after with numerical sequences in adults? To interpret children's behavior as being an error or an immature response pattern, we need a baseline to compare it to. Although our adult intuitions (i.e., the authors) are that we would be unlikely to respond like children in terms of our choices, the bias may still be evident in more subtle ways, such as reaction time. In Experiment 2, we use the same stimuli and paradigm as in Experiment 1 but measure both accuracy and reaction time with a sample of adults in the US to provide a baseline for interpreting children's behavior.

Experiment 1

Methods

Participants

Our primary analyses include data from 117 typically developing 5- and 6-year-old children (52 female, 49 male, 16 undisclosed; M_{age} = 71.4 months). An additional three children (total N = 120) are included in analyses of the first block only (see Supplemental), because they did not complete the second block. Finally, an additional 11 children participated but were excluded from the sample entirely because of refusal to answer questions (n=4), parental interference (n=3), technical difficulties (n=1), experimenter error (n=1), or because the child was outside the target age range (n=2).

We pre-registered a sample size of 120 completing at least the first block of trials, as we expected some attrition between the two blocks. We intended to rely primarily on the between subject analysis of the first task only, comparing within subject performance only if there is sufficient complete data. However, almost all children completed both tasks (only 3/120 did not), allowing us more power to do the within subject analyses. Thus, the within-subject sample of 117 is our primary sample reported in the manuscript and the analysis of the full sample of 120 is reported in Supplemental.

Children were recruited from an existing database of interested families, Children Helping Science prior to the 2023 merge with Lookit (Sheskin et al., 2020), and social media. Children were not recruited from a particular geographic region, as they could participate online, though most of the children were living in the US. Families received a $5 gift card for participating. Parent reported child race was 55% White, 13% Asian, 5% Black or African American, 11% one or more races, 2% other, and 14% undisclosed, with 15% of participants also identifying as Hispanic or Latine. Parent’s reported highest level of education was 44% graduate degree, 6% some graduate training, 29% 4-year undergraduate degree, 3% 2-year undergraduate degree, 3% some college, 2% high-school/GED, and 13% undisclosed.

Materials and Procedure

Participants met with an experimenter on video-chat using Zoom and were directed to an online slideshow to open in their own browser (using www.slides.com) and controlled by the experimenter (if technical difficulties prevented families from opening the slideshow, they viewed it via screen sharing instead). All participants completed two tasks, each organized into two blocks. The order of the tasks and blocks within each task were counterbalanced across participants. Trials within a block were in a set order.

Each context consisted of two blocks, randomly counterbalanced across participants. In the temporal context, participants watched a silent video of a female actor performing a series of five actions (videos are available on our OSF project page, linked below). In Block A the series of actions were as follows: the actor placed a glass on a table, opened a juice container, poured juice into the glass, placed a straw in the glass, and sipped from the straw. In Block B the series of actions were: the actor waved at the camera, took a banana out of a bowl, peeled the banana, took a bite of the banana, and drank from a glass of water. In the numerical context, participants watched a video of a female actor reciting a sequence of five consecutive numbers (videos are available on OSF). In Block A the number sequence was 2-3-4-5-6, in Block B it was 8-9-10-11-12.

After each video, children were shown 12 order judgement trials. On the temporal task, each trial displayed two stills of actions from the preceding video positioned vertically with the text "Is this [before/after] this?" (Figure 1). The experimenter read the text aloud and used a cursor to indicate which picture "this" referred to on the slide show. On the numerical task, each trial used the same structure as the temporal task, but using symbolic numerals instead of stills. The experimenter read the text aloud, providing the number word for the numerals (e.g., "is five after three?"). The questions were presented vertically because it allowed us to remove some of the left-right spatial biases associated with numerical order (Dehaene et al., 2005). Of course, biases associated with up and down (e.g., Gevers et al., 2006) and order of mention (e.g., Clark, 1971; Johnson, 1975) remain, as we always asked about the top option first relative to the bottom option. Children's verbal responses (yes/no) were recorded by the experimenter.

Within each block of 12 trials, half of the trials used the word "before" and half used "after". Half of the trials displayed consecutive actions/numbers (e.g. opening juice container and pouring juice) and half displayed non-consecutive actions/numbers (e.g. opening juice container and placing a straw in the glass). Eight trials were relationally consistent; that is, asking about a target that is in the correct relational direction (e.g. "Is 3 before 5?"). Four trials were relationally inconsistent (e.g. "Is 5 before 3?"), and therefore incorrect regardless of children's narrow interpretation of the ordinal word. We refer to these relationally inconsistent trials as "opposite" trials.

Figure 1. Example judgement trials for Number (left) and Temporal (right) Tasks.

Transparency and Openess

All data analyses were performed in R (Version 4.4.2; R Core Team, 2023) and the R-packages *afex* (Version 1.4.1; Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2024), *broom* (Version 1.0.7; Robinson, Hayes, & Couch, 2023), *diptest* (Version 0.77.1; Maechler, 2024), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *flextable* (Version 0.9.7; Gohel & Skintzos, 2024), *forcats* (Version 1.0.0; Wickham, 2023a), *ggdist* (Version 3.3.2; Kay, 2024), *ggplot2* (Version 3.5.1; Wickham, 2016), *ggpubr* (Version 0.6.0; Kassambara, 2023a), *knitr* (Version 1.48; Xie, 2015), *lme4* (Version 1.1.35.5; Bates, Mächler, Bolker, &

Walker, 2015), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.1; Bates, Maechler, & Jagan, 2023), *papaja* (Version 0.1.3; Aust & Barth, 2023), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2023), *rstatix* (Version 0.7.2; Kassambara, 2023b), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2023), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and *tinylabels* (Version 0.2.4; Barth, 2023).

We pre-registered the sample size, design, and analysis on OSF (https://osf.io/de24t/?view_only=a0d3866817ea4be1bdc237e50ec844b5). However, our primary analysis deviates from the pre-registered analysis. We pre-registered using a between-subject analysis of the first block of trials because we anticipated attrition between blocks, preventing a large enough sample for a within subject analysis. However, because we had less attrition than expected (only 3/120 did not complete both blocks), we opted to report the within subject analysis in the manuscript because this analysis provides more power and allows us to investigate order effects (even with the slightly smaller sample size, 117 vs. 120). For transparency, we report the pre-registered between subject analyses with the full sample in Supplemental, though the pattern of results is identical. All deidentified data, materials, and analysis scripts are provided on the OSF: https://osf.io/u4m5t/?view_only=1008bf234bf9426eb0497b83633a93ee.

Results

Preliminary Analyses

Means (including 95% confidence intervals) and comparisons to chance in each trial type and context separately are presented in Table 1. First, to ensure participants understood the nature of the task, we analyzed their performance on the opposite trials, where the answer was unambiguously “no”. As seen in Figure 2A and Table 1, average performance was significantly above chance for opposite trials in both the Number and Temporal Contexts, showing that, in general, children seemed to understand the task.

There were 9 children who consistently responded “yes” on all opposite trials within a task (3 on number trials, 4 on temporal trials, and 2 on both number and temporal trials), suggesting that they did not understand the task. For our primary analysis, we maintain these children in the analyzed sample, as we did not pre-register excluding them. However, as a robustness check, the analysis excluding these children resulted in an identical pattern and is reported in Supplemental.

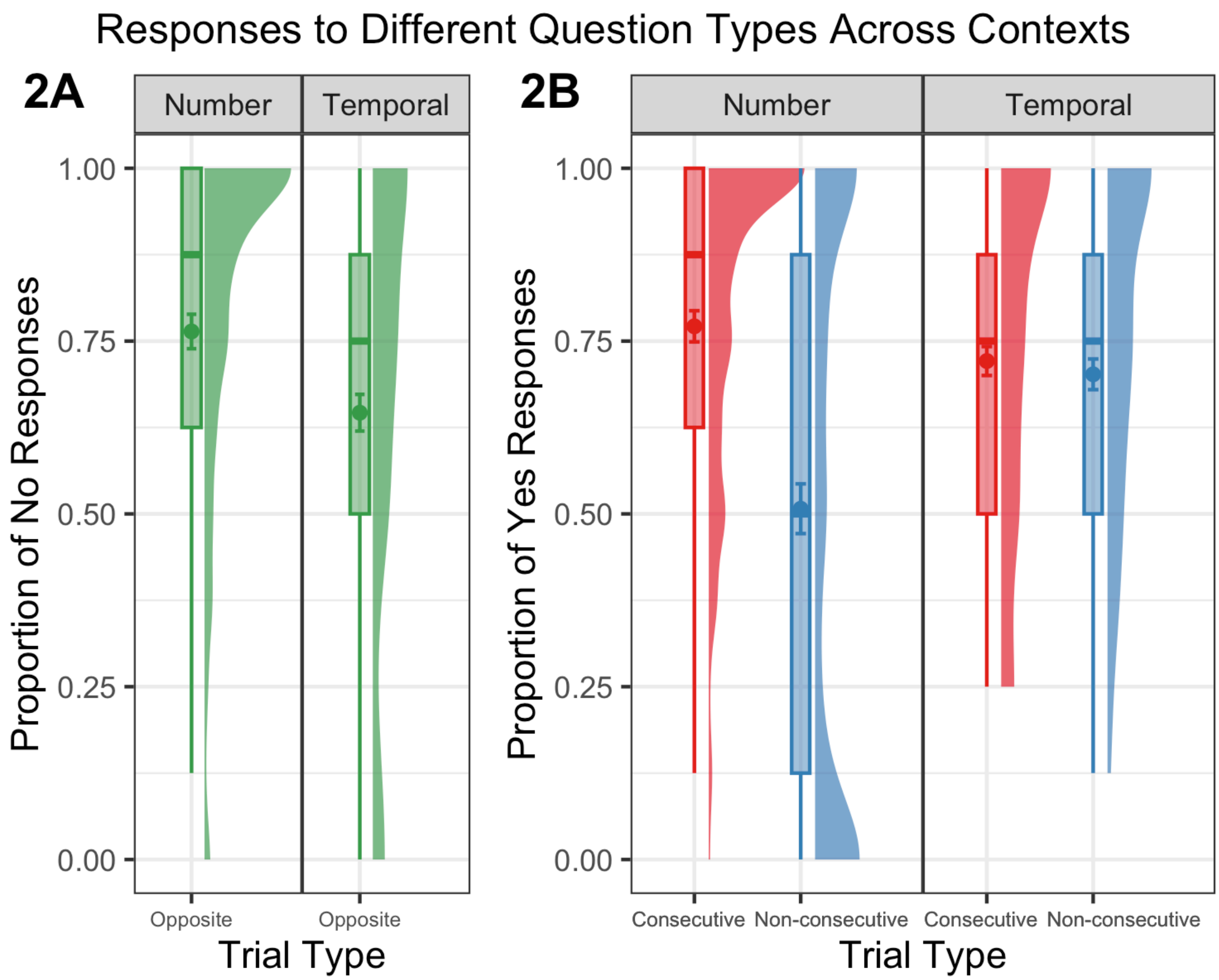

Figure 2. Children's performance across trial types and contexts in Experiment 1. Panel A: Proportion of "No" responses to opposite trials (e.g., "Is 5 before 3?"). Panel B: Proportion of "Yes" responses to consecutive (left, red) and non-consecutive (right, blue) trials (e.g., "Is 4 before 5" and "Is 3 before 5", respectively) in the number and temporal contexts as separate facets.

Table 1: Descriptive data and comparisons to chance for each trial type and context separately

Context	Trial Type	M	95% CI	t(116)	p
Number	Opposite	0.76	[0.71, 0.81]	10.59	< .001
	Consecutive	0.77	[0.73, 0.82]	12.05	< .001
	Nonconsecutive	0.51	[0.44, 0.58]	0.21	0.836
Temporal Events	Opposite	0.65	[0.59, 0.70]	5.52	< .001
	Consecutive	0.72	[0.68, 0.76]	10.56	< .001
	Nonconsecutive	0.70	[0.66, 0.75]	9.10	< .001

Second, initial analyses including context order as a between-subject factor in the 2 (order: number first, temporal first) x 2 (trial type: consecutive, non-consecutive) x 2 (context: number, temporal) ANOVA did not reveal a significant main effect of order $F(1{,}115) = 1.47$, $p = .228$, $\hat{\eta}_p^2 = .013$, or interactions with trial type, $F(1{,}115) = 2.98$, $p = .087$, $\hat{\eta}_p^2 = .025$, or context, $F(1{,}115) = 0.46$, $p = .499$, $\hat{\eta}_p^2 = .004$, or a three-way interaction, $F(1{,}115) = 1.43$, $p = .235$, $\hat{\eta}_p^2 = .012$. Given these results, and that the between-subject analyses on the first context only parallel these results (see Supplemental), we do not include context order as a factor in our analyses, except where specified in the Exploratory Analysis section.

Primary Analysis of Relationally Consistent Trials

For our primary analysis, we focused on only the relationally consistent trials – that is, the trials where the number or event was on the correct relational side but was either consecutive or non-consecutive. Using a 2 x 2 ANOVA with Context (Temporal versus Number) and Trial Type (Consecutive versus Non-consecutive) as within-subject factors, we find significant main effects of Context $F(1{,}116) = 10.65$, $p = .001$, $\hat{\eta}_p^2 = .084$ and Trial Type $F(1{,}116) = 45.54$, $p <$

$.001$, $\hat{\eta}_p^2 = .282$, as well as a significant interaction between Context and Trial Type $F(1{,}116) = 60.97$, $p < .001$, $\hat{\eta}_p^2 = .345$ (Figure 2B). To further investigate the interaction, we performed two t-tests comparing consecutive vs. non-consecutive trials in each context separately. In the number context, there was a significant difference between trial types, with children more likely to respond "yes" to consecutive judgements than non-consecutive judgements, $M_D = 0.26$, 95% CI $[0.20{,}0.33]$, $t(116) = 8.17$, $p < .001$, $d = 0.81$. In contrast, there was not a significant difference between trial types in the temporal context, $M_D = 0.02$, 95% CI $[-0.02{,}0.06]$, $t(116) = 1.06$, $p = .291$, $d = 0.08$. Furthermore, children responded significantly above chance (Table 1) on both trial types in the temporal context, but only on consecutive trials in the number context, and instead responded to non-consecutive trials at a level that was not significantly different from chance. Together, these findings suggest that children's narrow interpretation of before and after applies to sequences of natural numbers but not to temporal sequences of events.

Exploratory Analysis: Bimodal Response Distribution

Although the average proportion of "yes" responses to judgements comparing non-consecutive numbers was at chance, inspection of Figure 2B reveals a bimodal distribution. Using Hartigan's dip statistic (Freeman & Dale, 2013), we find significant evidence against the data being unimodal: $D = 0.11$, $p < .001$. Thus, rather than being concentrated around 0.5, as it would be if children were guessing between the two response options (i.e., yes or no), the data is skewed towards the extremes, with some children consistently responding "yes" and others consistently responding "no". To investigate the cause of this bi-modal distribution, we used exploratory analyses to ask whether children's behavior differed as a function of age or the order in which they received the two blocks (i.e., whether the numerical block was after or before the temporal block). Notably, because the study was within-subject, children who are consistently responding either “yes” or “no” are doing so on both the smaller number range (2 to 6) and the larger number range (8 to 12), suggesting that numerical range does not explain the bimodal distribution.

First, we divided children into two groups: all-yes responders on non-consecutive number trials (i.e., those who do not have a narrow interpretation, n = 25) and all-no responders on the non-consecutive number trials (i.e., those who do have a narrow interpretation, n = 27). There was not a significant difference in age between the all-yes responders, M_{age} = 71.2 months, and the all-no responders, M_{age} = 72.2 months, $\Delta M = 1.02$, 95% CI $[-2.81, 4.86]$, $t(49.93) = 0.54$, $p = .595$, d = 0.15.

However, there was a hint that the order of the two contexts impacted children’s responses (Figure 3). There were proportionally more all-yes responders among children who saw the temporal context first, 18 all-yes responders vs. 13 all-no responders (i.e., 58% all-yes), compared to children who saw the number context first,7 all-yes responders vs.14 all-no responders (i.e., 33% all-yes), though the difference is not significant, χ^2(1, N=52) =3.07, p = 0.08. Additionally, there is a small but not significant difference in overall performance as a function of order: the average proportion of “yes” responses on non-consecutive number trials for children who had the number context first, M = 0.45, is lower, although not significantly, than children who had the number context second, M = 0.57, $\Delta M = -0.12$, 95% CI $[-0.26, 0.02]$, $t(114.53) = -1.64$, $p = .103$, d = -0.30. Although neither analysis is statistically significant, the pattern of the means and the distribution of responses in Figure 3 suggests that there may be some malleability – at least for some children – in judgements of non-consecutive numerical order. If that is the case, future research can investigate the contexts that support children’s judgements of numerical order as well as whether there are educational implications of any such differences.

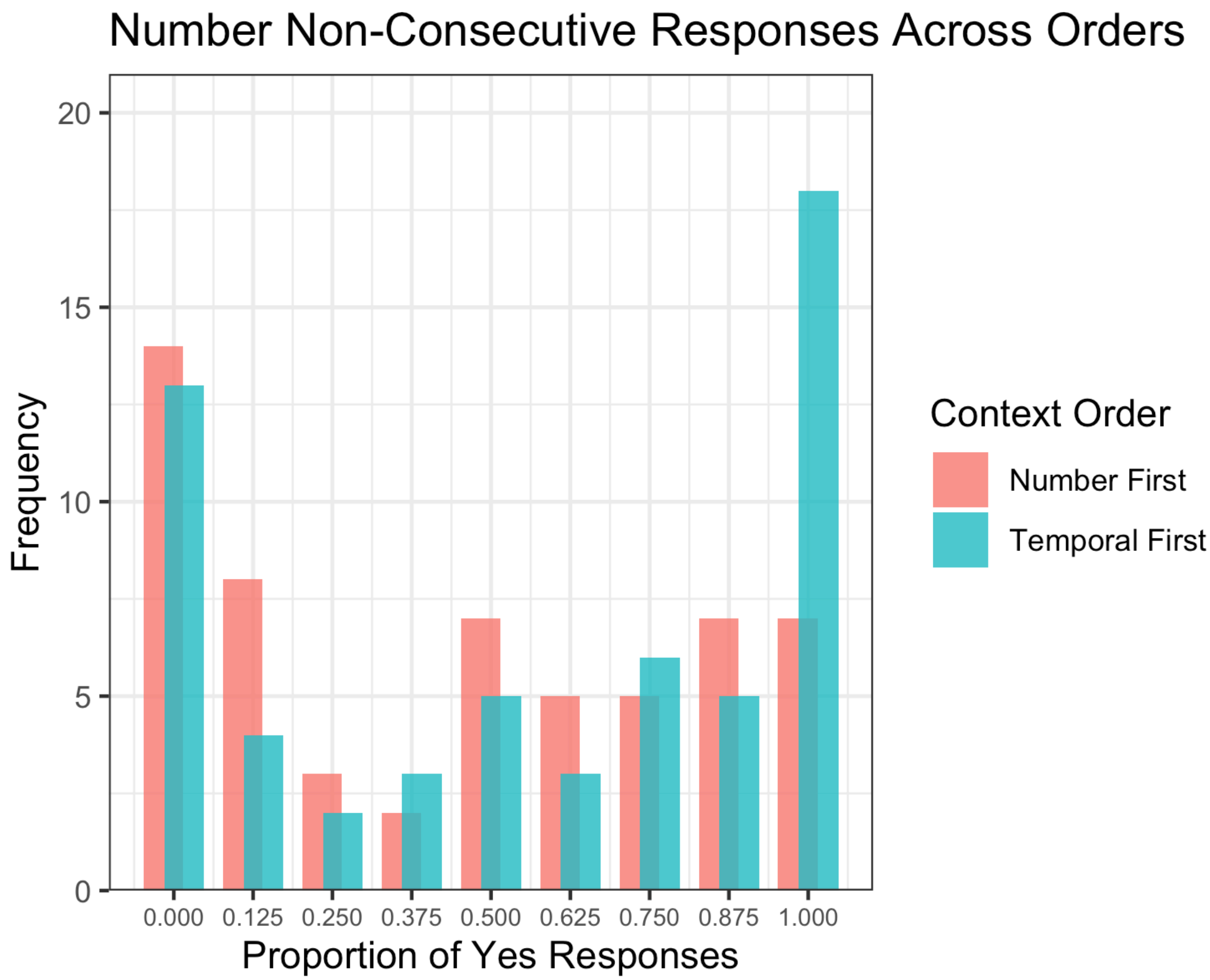

Figure 3. A histogram showing the number of children with different proportions of yes responses across both context orders, only for the non-consecutive trials within the number context.

Exploratory Analysis: Before vs. After

Although the prior study using a very similar task did not report findings separately for the words before vs. after (Hurst et al., 2022), in non-numerical contexts children acquire the word before before acquiring the word after (Clark, 1971). Thus, as additional exploratory analyses, we compared performance within each context and trial type when asked about “before” relations versus when asked about “after” relations.

We found that children were significantly more accurate (where accuracy is defined as “no” for opposite trials and “yes” for both consecutive and non-consecutive trials) with *before* compared to *after* on consecutive number trials and on both consecutive and non-consecutive temporal trials (Table 2). However, the pattern found in our primary analyses is consistent even when each word is analyzed separately, with a significant difference between consecutive and non-consecutive trials in the number context, and not significantly different in the temporal context, for both before and after (Table 3)

Table 2: Comparing performance across relational terms before and after

Context	Trial Type	M (after)	M (before)	t(116)	p	Cohen's d [95%CI]
Number	Opposite	0.76	0.76	-0.08	0.936	-0.01 [-0.21, 0.16]
	Consecutive	0.76	0.79	-0.87	0.388	-0.08 [-0.25, 0.1]
	Nonconsecutive	0.47	0.54	-2.72	0.008	-0.25 [-0.43, -0.08]
Temporal Events	Opposite	0.63	0.66	-0.89	0.378	-0.08 [-0.25, 0.12]
	Consecutive	0.64	0.80	-5.17	< .001	-0.48 [-0.7, -0.31]
	Nonconsecutive	0.65	0.76	-3.58	< .001	-0.33 [-0.53, -0.16]

Table 3: Comparing performance across consecutive vs. non-consecutive trials, separated for before and after.

Context	Ordinal Word	t(116)	p	Cohen's d [95% CI]
Number	after	7.84	<.001	0.72 [0.54, 0.92]
	before	6.44	<.001	0.6 [0.43, 0.77]
Temporal Events	after	-0.24	0.814	-0.02 [-0.2, 0.15]
	before	1.71	0.09	0.16 [-0.02, 0.35]

Exploratory Analyses: Causal Connections in Temporal Sequences

One possible difference between numerical sequences and temporal events is that the order of events in time can have causal or logical constraints. For example, the actor in our video needed to pick up the banana before peeling it and could not have peeled it before picking it up. In our stimuli, both videos included strong causal or logical constraints on the order of some events (e.g., opening a juice container before pouring juice in Video A; the banana example above in Video B), but also events that had weaker connections (e.g., getting the glass before opening the juice container; drinking water after eating the banana). Overall, however, the logical connections between events seem to be stronger and more consistent in Video A (the actor placed a glass on a table –> opened a juice container –> poured juice into the glass –> placed a straw in the glass –> sipped from the straw) than in Video B, which included more disconnected events at the beginning and end (the actor **waved at the camera** –> took a banana out of a bowl –> peeled the banana –> took a bite of the banana –> **drank from a glass of water**). To explore whether this impacted performance, we compared children's behavior on all three trial types across the two videos and within Video B across trials that included only logically ordered events (i.e., the middle three events involving the banana) or included the less causally ordered events (i.e., waving and drinking water, which could in theory occur at any time relative to the banana sequence). However, children's behavior did not vary in any comparison (see Tables 4 and 5).

Table 4: Comparing average proportion of trials correct across the two temporal videos

Trial Type	Video A	Video B	t(116)	p	Cohen's d [95% CI]
Opposite	0.65	0.64	-0.21	0.831	-0.02 [-0.22, 0.18]
Consecutive	0.74	0.70	-1.43	0.156	-0.13 [-0.31, 0.05]
Nonconsecutive	0.69	0.72	1.18	0.239	0.11 [-0.07, 0.3]

Table 5: Comparing average proportion of trials correct within Video B comparing only trials that involve the banana (i.e., include only causally or logically ordered events) vs. trials that include one of the more ambiguous events

Trial Type	Trials including an ambiguous event	Trials including only causally or logically ordered events	t(116)	p	Cohen's d [95% CI]
Opposite	0.64	0.65	-0.31	0.757	-0.03 [-0.22, 0.16]
Nonconsecutive	0.70	0.74	-1.34	0.183	-0.12 [-0.31, 0.07]

Discussion

In Experiment 1, we replicate the previously reported result: English speaking 5- to 6-year-old children in the US are more likely to interpret before/after as referring narrowly to consecutive numbers than to non-consecutive numbers. Importantly, however, we find that these same children do not apply before/after as narrowly when comparing events in a temporal sequence, suggesting that this pattern is not due to general constraints of the paradigm or their general interpretation of the words before/after across contexts. Instead, there may be something about number in particular that is driving this pattern. We further found that behavioral patterns on the numerical task was bimodal: some children consistently responded with a constrained interpretation and some consistently responded with a broad interpretation. Exploratory analyses investigating some potential explanations for these patterns did not provide an obvious answer: age was not a significant predictor of children's narrow-interpretation and children's narrow-interpretation was consistent for the smaller and larger numerical range and for both the terms *before* and *after*. There was a small (but not significant) pattern such that children who responded to the temporal videos first (where they do not have a narrow interpretation) were slightly less likely to display a narrow interpretation of before/after on numerical trials, compared to children who responded to the numerical videos first. This suggests that some children may have modified their numerical responses because they were primed to interpret before/after broadly, as they do for temporal event sequences. Even so, however, this pattern was not significant overall, and children's consecutive-only interpretation was apparent in both task orders – suggesting that the consecutive-only interpretation is robust.

We also analyzed the data in two additional ways that warrant discussion. First, although our primary analysis held for both the words *before* and *after* separately, it is worth noting that when there was a difference between the two words, it reflected a benefit for the word *before*, over the word *after*. This may be due to our paradigm, where *before* questions were consistent with the order-of-mention. That is, when asked about *before* children can use whether the two mentioned numbers or events are in the correct order; for example, "is 3 before 5" is akin to evaluating "3, 5", whereas "is 3 after 5" requires reversing the to-be-compared values to evaluate "5, 3", even though order-of-mention was "3, 5". The order-of-mention has been shown to impact children's interpretation of *before*/*after* in other paradigms as well (e.g., Clark, 1971; Johnson, 1975).

Second, one possible hypothesis for why numerical sequences and temporal sequences result in different interpretations of before/after is due to the causal and/or logically structured nature of temporal events (e.g., you need to peel a banana before you can eat it). However, exploratory analyses across our two trial blocks (which may have differed in how clear the logical connections were) and across trials with weaker vs. stronger causal/logical connections (e.g., waving then eating a banana vs. peeling a banana then eating a banana) revealed that children's behavior was very similar in these different contexts. However, our experiment was not designed to test this hypothesis, and so future work is needed to more carefully test whether causal or logical relations between events may impact children's interpretation of their ordinal relations.

Despite the bimodal responding and the robustness of this pattern in 5- and 6-year-old children in the US, it is unclear whether the narrow interpretation of before and after in the context of number should be interpreted as being *incorrect*, as we do not have an adult baseline for comparison. In Experiment 2, we use the same paradigm with a sample of US adults, to provide a baseline for further interpreting children's behavior.

Experiment 2

Methods

Participants

We collected data from 120 adults on Prolific to match the planned sample size from Experiment 1. Based on our pre-registered exclusion criteria, we excluded six participants: not completing the entire experiment in one attempt (n = 2), incorrectly answering more than 25% of the attention check questions (n = 3), or self-reporting they were not paying attention (n = 1).

This resulted in a sample of 114 adults: M_{age} = 39 years, range = 19 – 77 (one missing); 57 males, 52 females, 5 non-binary, and 1 undisclosed. Participant reported race was 74% White, 12% Asian, 4% Black or African American, 5% one or more races, less than 1% other option not listed here, and less than 1% undisclosed. Additionally, 7% of participants identified as Hispanic or Latine. Participants' highest level of education was 12% graduate degree, 3% some graduate training, 35% 4-year undergraduate degree, 11% 2-year undergraduate degree, 24% some college, 14% high school/GED, < 1% did not complete high-school, and < 1% undisclosed.

Materials and Procedures

Participants were recruited via Prolific and paid $1.80 (equivalent to $12/hour). Stimuli were presented online via cognition.run. Participants were explicitly told "This experiment was designed for children and so some of the questions might seem simple or silly. Just answer as best you can without overthinking it". The design and stimuli were identical to Experiment 1, except: there was no experimenter verbally providing instructions, there were additional instruction screens to replace what the experimenter would typically say in synchronous studies with children, participants recorded their answer by selecting a "yes" or a "no" button on the screen with their mouse, and reaction time (in addition to response choice) was recorded.

Following the primary task, participants were asked whether they were paying attention, with the options: "Yes: I was trying to pay attention most of the time.", "Somewhat: I was trying to pay attention some of the time.", and "No: I was not paying attention for most of the time." They were also asked what strategy they used to make their judgements, using separate open-ended text boxes for each context (i.e., for number and for temporal events).

Responses to the open-ended strategy questions were coded into predetermined categories by two coders. Due to the possible overlap of strategy categories, coders were allowed to assign a response as belonging to more than one strategy. For the temporal context the categories were: an expected progression of events, a labeling of events, a reliance on memory of the video, an ordering of the events, a commonsense strategy, and/or no response. For the number context the categories were: knowledge of counting and number order, magnitude of numbers, a reliance on their memory of the video, the order of presentation within the video, a commonsense strategy, and/or no response. Examples of these coding categories are provided in Table 7. The two coders agreed on 89% and 91% of the strategy codes for the temporal and number contexts, respectively. The remaining discrepancies were decided by open dialogue between the two coders with the help of a moderator.

Transparency and Openness.

We preregistered the sample size, design, and analysis on OSF (https://osf.io/ybcsd/?view_only=544d9a49949642229d6dffda1f9a3bd6). We deviated from our planned analysis because of extreme skew in the data. We had originally preregistered performing a 2 x 2 ANOVA between Context (2: Temporal vs. Numerical) and Trial Type (2: consecutive vs. non-consecutive) as within-subject comparisons. However, adults' responses were extremely internally consistent, resulting in highly skewed data with insufficient variability to analyze using parametric statistics. Because of this, we use non-parametric tests on binned data, targeting the same research questions. As in Experiment 1, our pre-registration, materials, and all data are available on OSF: https://osf.io/u4m5t/?view_only=1008bf234bf9426eb0497b83633a93ee

Results

Preliminary Analyses

We used performance on the opposite trials as exclusion criteria – as they are objectively false and should be trivially easy for adults. Adults that failed to correctly judge at least 75% of these questions were excluded from our analyses (n = 3). Overall, adults performed very well with both the number, M = 0.98, SD = 0.05 and the temporal, M = 0.95, SD = 0.08 contexts. Additionally, on these opposite trials adults were slower to respond in the temporal context than the numerical context, $M_D = -1{,}789.67$, 95% CI $[-1{,}964.95, -1{,}614.38]$, $t(113) = -20.23$, $p < .001$, d = -1.89 (Table 6). This benefit for number is not surprising given that the numbers were in a standard increasing order, whereas the temporal events were videos they had never seen before, requiring additional memory resources. For the primary analyses we focus only on consecutive and non-consecutive trials and do not include data from opposite trials.

Table 6: Descriptive statistics for Experiment 2

Context	Trial Type	M (ms)	SD (ms)	N All Yes Responders	N At least One No Response
Number	Consecutive	2,316	1,427.389	97	17
	Nonconsecutive	2,318	1,331.942	85	29
	Opposite	2,442	1,464.454	-	-
Temporal Events	Consecutive	4,353	2,141.654	54	60
	Nonconsecutive	3,952	2,121.662	65	49
	Opposite	4,232	2,018.268	-	-

Behavioral Responses

For each trial type and context separately, we compared the number of adults who responded “yes” 100% of the time versus the number of adults who responded with at least one “no” (Table 6). Using McNemar’s Chi Square test, we find that the distribution of participants in each of these two groups significantly differed across consecutive and non-consecutive trials in the number context, χ^2(1, N=114) = 4.24, *p* = 0.04, but not in the temporal context, χ^2(1, N=114) = 2.81, *p* = 0.09. Thus, although adults’ behavior is unlike children’s behavior – in that most adults consistently said “yes” to both consecutive and non-consecutive number comparisons – the occasional deviations from this behavior mirrored children’s consecutive-only interpretation of before and after in numerical contexts. Notably, although not significant, the temporal context is in the opposite direction of the number context, with numerically more adults responding with at least one no (vs. all yes) on the consecutive trials and the opposite on non-consecutive trials.

Reaction Time

We also analyzed average reaction time, using a 2 x 2 ANOVA between Context and Trial Type (Table 6). There were significant effects of Context, $F(1,113) = 454.91$, $p < .001$, $\hat{\eta}_p^2 = .801$, with adults responding faster on judgements about numbers than events, and Trial Type, $F(1,113) = 15.81$, $p < .001$, $\hat{\eta}_p^2 = .123$, as well as a significant interaction, $F(1,113) = 14.92$, $p < .001$, $\hat{\eta}_p^2 = .117$. We further explored this interaction using paired t-tests comparing consecutive versus non-consecutive response times on both the number and temporal block. There was not a significant difference between response times for consecutive and non-consecutive trials in the number context, $M_D = -1.26$, 95% CI $[-106.89,104.36]$, $t(113) = -0.02$, $p = .981$, *d* = <.01, but there was a significant difference in the temporal context $M_D = 405.77$, 95% CI $[229.85,581.68]$, $t(113) = 4.57$, $p < .001$, *d* = 0.43, with participants answering more quickly on non-consecutive trials, which is consistent with their slightly higher scores on these question types as compared to consecutive trials.

As preregistered, we also analyzed average reaction time only from trials where the participant responded "yes." We found similar results, including significant effects of Context, $F(1,112) = 441.80$, $p < .001$, $\hat{\eta}_p^2 = .798$, with adults performing faster on the number context, Trial Type, $F(1,112) = 5.22$, $p = .024$, $\hat{\eta}_p^2 = .045$, and a significant interaction between the two $F(1,112) = 7.16$, $p = .009$, $\hat{\eta}_p^2 = .060$. Once again, we find no significant difference in reaction time across trial type for the number context $M_D = -18.71$, 95% CI $[-127.09, 89.67]$, $t(112) = -0.34$, $p = .733$, d = -0.03, but there was a significant difference for the temporal context $M_D = 270.51$, 95% CI $[83.37, 457.65]$, $t(112) = 2.86$, $p = .005$, d = 0.27, with faster responding on non-consecutive trials vs. consecutive trials.

Strategies

Data from 112 participants were included in the strategy analyses, because 2 participants were excluded for reporting on the wrong context (i.e. referencing numbers when asked about strategy use within the temporal context). Examples of each category and the number of responses grouped into each category are presented in Table 7. For temporal events, the most common strategy reported was relying on their memory of the video, followed by the order of events, and expected progression of events. For number sequences, the most common strategy reported was the order of the video, followed by memory of the video, and the "common sense" strategy. In both contexts, participants relied most on strategies that referenced the specific order of occurrence of events in the video or referenced a less specific video memory. This was true even in the number context, where the judgement trials were essentially questions about conventional numerical order.

Table 7: Coding of adults' reported strategies in Experiment 2

	Example Responses		Proportion of Responses	
Strategy	Temporal Events	Number	Temporal Events	Number
Order of video events	“The sequence of events in the video."	“I was listening to the order the numbers were listed in. I heard 5 after 3.”	0.45	0.59
Memory of video	“I remember what happened.”	“I remember what was said in the video”	0.56	0.25
Expected progression of events	“You cannot pour juice until after you have opened it”	NA	0.33	NA
Common sense	“I know what happened.”	“I know math”	0.03	0.15
Labeling of the video's events	“I just saw she poured the juice before she drank it.”	NA	0.21	NA
Counting	NA	"I counted to 4 and it came after 3."	NA	0.12
Magnitude	NA	“I knew 4 came after 3 because 4 is bigger than 3.”	NA	0.09
No response			0.06	0.08

Note: Some strategies are included in a singular context because they are not applicable to the other (e.g. a counting strategy cannot be used in a temporal context). This is denoted by NA. Additionally, because participants could report multiple strategies, the values do not add up to 100%.

Discussion

In Experiment 2, we find that adults’ overall behavior is inconsistent with a consecutive-only interpretation, and thus is different than children in Experiment 1. However, in the very few cases when behavior deviates from the predominant “yes” responding, the pattern of deviation parallels the narrow interpretation of before/after for number shown by children. Additionally, in our reaction time analyses, we see evidence of a traditional distance effect for temporal event sequences, such that comparing events further apart is faster (i.e., less difficult) than comparing events closer together. However, for numerical sequences we do not see differences in reaction time across consecutive and non-consecutive trials – meaning, we do not see significant evidence of either a traditional distance effect or the reverse distance effect sometimes reported for ordinal comparisons (e.g., Lyons & Ansari, 2015). One possibility is that the presence of both a distance effect and reverse distance effect, either across participants or across trials, is resulting in a non-significant difference on average. Overall, this suggests that there might be some child-like bias in adult's interpretation that guides behavior only on a small number of trials (e.g., a failure of inhibitory control). Notably, there was one adult who responded strongly in terms of the consecutive-only interpretation of before/after, responding "no" on all numerical non-consecutive comparisons. When asked to report their strategy, the inherent inconsistencies in the use of "before" become apparent: "Even though the number 3 comes before number 5 I chose to answer by what was in the video. So only 4 would be the correct answer. Not everything that comes before 5 for example, 1,2,3,4." This participant's explanation also suggests there might be task-specific features supporting this interpretation, in ways that aren't necessarily present in everyday interpretations of before and after. These task-specific factors may also differ from the factors that contribute to children's narrow interpretation of before/after for numbers. Future work could contribute to our understanding of what these features might be and/or whether they arise in everyday settings that impact behavior.

General Discussion

In the current study, our goal was to begin to understand the cause of US children's narrow interpretation of before/after as only applying to consecutive numbers, but not non-consecutive numbers, by comparing this behavior to (1) children's interpretation of before/after in temporal event sequences and (2) adults' interpretation of before/after in both numerical sequences and temporal event sequences. Overall, we find that children in our sample did not demonstrate a consecutive-only interpretation of before/after when comparing temporal events, while replicating the previous findings with numerical sequences. Further, adults' behavior was also not consistent with a consecutive-only interpretation, with most adults responding "yes" on all relationally consistent trials. Together, these results suggest that English speaking 5- and 6-year-old US children's narrow interpretation of before/after is being caused by a feature of numerical sequences that is not present for temporal events and may be less salient or apparent to adults.

What causes this constrained interpretation of before/after specifically for number?

These findings provide initial insight into the question of why children respond narrowly to before/after in the context of numerical sequences by telling us some of the things that are not causes: it is not due to their understanding of before/after more broadly, details of the specific paradigm shared across our two versions of the task, or contextual features that are also available to and salient for adults. Furthermore, the prior work using a similar paradigm suggests that it is not driven by lack of number knowledge in general, as children do well on judgements of magnitude (e.g., bigger/smaller) despite responding narrowly in terms of order (i.e., before/after; Hurst et al. (2022)). However, the current study alone cannot tell us what about numerical sequences is driving children's interpretation.

Although prior work on numerical ordering more generally suggests that the consecutive bias for number may be driven by the learned count sequence (e.g., Gilmore & Batchelor, 2021), future research is needed to tease apart what features of the numerical count list cause this effect, such as the habitual learning of a stable and arbitrary sequence and/or the presence of magnitudes or successors. For example, if it is due to the habitual learning of a stable sequence of arbitrary symbols (something that is very different than events, which often have a causal or logical order) then we would expect the same patterns to hold for similarly learned stable sequences, such as the alphabet or days of the week. These sequences have immediate successors, but do not have direct ties to magnitude. In contrast, if the connection to magnitude is necessary, then even those learned habitual sequences may differ from numerical sequences. Other contrasts, such as numerical sequences that are not in counting order (e.g., phone numbers), action sequences that are arbitrary, learned, and referenced using verbal labels (e.g., the children's song "head, shoulders, knees, and toes", which is a series of actions that might be habitual for some children, but with no causal relations), size-based sequences (e.g., an ordered set of circles increasing in size), as well as sequences that have some combination of these features could provide insight into which feature(s) are necessary for children to hold a narrow understanding of order and ordinal vocabulary.

Additionally, future work is necessary to investigate how children demonstrate ordinal knowledge across tasks, such as counting tasks (e.g., count as high as you can), number line tasks (e.g., place 25 on a line from 0 to 100), order verification tasks (e.g., are 2 3 4 in order), and ordinal language tasks (e.g., does 5 come before 7). Each of these tasks requires aspects of number knowledge beyond ordinality, while also measuring some aspect of ordinality.

Finally, research on children's experience with the words before and after in different contexts could provide insight into whether their everyday exposure to these words may parallel this interpretation (i.e., is it the case that before/after are less often used for non-consecutive numbers [vs. consecutive numbers], but that this is not the case for events).

How does this pattern change across development?

Although we did not find age to be a significant predictor of behavior within our sample of 5- to 6-year-olds, we did find robust differences between the child sample and the adult sample. When and how this interpretation develops and when it disappears, however, are open questions. Prior work on children's judgements of numerical order more generally reveals a similar bimodal distribution as found in the current study (e.g., Gilmore & Batchelor, 2021; Hutchison et al., 2022). One possibility is that children's conceptual shift in understanding numerical order as not requiring consecutive values [i.e., accepting both 2 4 6 and 2 3 4 as being in order; Hutchison et al. (2022)] also underlies the shift in children's interpretation of the ordinal vocabulary terms before and after. If this is the case, then we might expect the developmental pattern to follow the developmental pattern found for these kinds of ordinal judgements. Specifically, prior studies have suggested that 4- and 5-year-olds may hold a stronger narrow understanding of ordinality, with below-chance responding about non-consecutive numbers (Hutchison et al., 2022), with 6- to 9-year-olds at chance on average (Gilmore & Batchelor, 2021). These findings suggest that the shift in the narrow interpretation of before/after for number may happen sometime in this time window. Future work is needed to test this hypothesis, by investigating the potential connections between conceptual understanding of order and interpretation of specific ordinal vocabulary across development.

Limitations

There are several limitations in the current study that must be considered. First, our sample of children are predominantly from highly educated parents, with almost three-quarters of the families having a parent with at least an undergraduate degree and almost half having a graduate degree. These children may be more exposed to relational language and numerical language than is typical, changing their interpretation of these terms and/or revealing a different developmental pattern than children with less educated parents and/or less exposure to relational and numerical language. Second, we did not investigate any other variables that might have impacted performance differentially across tasks, such as working memory or inhibitory control. For example, it may be that inhibitory control is important for inhibiting the dominant response of saying “no” to non-consecutive numbers, allowing children to respond “yes”. Additionally, the temporal sequences may have relied more on short-term memory, as the specific sequence of events in the video was unfamiliar, whereas the numerical sequences may have relied more on long term memory and number knowledge, as the videos were consistent with the count sequence. On the other hand, the temporal event sequences included causally or logically related sequences (e.g., peeling a banana before eating it), which may aid performance, though our exploratory analyses did not find differences in behavior across the two videos or trials with or without a causally or logically linked comparison. Finally, the numerical and temporal paradigms differed in the preciseness of the question posed. For the numerical task, we embedded the number word into the sentence (e.g., is 3 before 5?), whereas for the temporal task, we opted not to label the events, instead referring to images generically (e.g., is this before this?). It may be that the extra processing required to turn the question with deictic terms into a specific question about the image prompts changed performance relative to the specific question involving the numerical values. We opted not to label the events (e.g., “is peeling the banana before eating the banana?”) because we did not want to heavily impose a discrete structure onto the continuous temporal events; however, future work will be needed to investigate whether this difference in question specificity might be a contributing factor to differences in children’s performance in numerical and temporal contexts.

Conclusion

In conclusion, across two experiments we find that English speaking children in the US interpret the words before and after narrowly to refer to consecutive, but not (or less so) non-consecutive numbers. Further, these same children do not show this pattern when comparing temporal event sequences and US adults do not show this pattern for either numerical or temporal event sequences, albeit displaying a trace of children's narrow construal of before/after in the context of number. There are many open questions about children's interpretation of before/after, including: Are there other sequences for which US children also show a consecutive bias? What is the developmental trajectory of this bias? Are there educational implications for learning numerical concepts that accompany this bias? The current work provides a jumping off point for better understanding US children's relational judgements of sequences and the words used to describe them.

References

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Verbal Learning and Verbal Behavior*, *11*(4), 417–423. https://doi.org/10.1016/S0022-5371(72)80022-7

Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R Markdown*. Retrieved from https://github.com/crsh/papaja

Barth, M. (2023). *tinylabels: Lightweight variable labels*. Retrieved from https://cran.r-project.org/package=tinylabels

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Maechler, M., & Jagan, M. (2023). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from https://CRAN.R-project.org/package=Matrix

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*(1), 7283. https://doi.org/10.1111/1467-7687.00211

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*(1), 59–68. Retrieved from https://www.proquest.com/docview/210571119/abstract/79FDD1A49D264084PQ/1

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

Coker, P. L. (1978). Syntactic and semantic factors in the acquisition of. *Journal of Child Language*, *5*(2), 261–277. https://doi.org/10.1017/S0305000900007467

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*(1), 162173. https://doi.org/10.1016/j.cognition.2011.12.013

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

Fuson, K. C. (1988). *Early Relationships Among Sequence Number Words, Counting Correspondence, and Cardinality* (K. C. Fuson, Ed.). New York, NY: Springer. https://doi.org/10.1007/978-1-4612-3754-9_10

Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., & Fias, W. (2006). Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychologica*, *122*(3), 221–233. https://doi.org/10.1016/j.actpsy.2005.11.004

Gilmore, C., & Batchelor, S. (2021). Verbal count sequence knowledge underpins numeral order processing in children. *Acta Psychologica*, *216*, 103294. https://doi.org/10.1016/j.actpsy.2021.103294

Gohel, D., & Skintzos, P. (2024). *Flextable: Functions for tabular reporting*. Retrieved from https://CRAN.R-project.org/package=flextable

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. Retrieved from https://www.jstatsoft.org/v40/i03/

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

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

Hurst, M. A., Greene, A., Eason, S. H., Claessens, A., & Levine, S. C. (2022). Children's understanding of relational language for quantity comparisons. *Cognitive Development*, *63*, 101214. https://doi.org/10.1016/j.cogdev.2022.101214

Hutchison, J. E., Ansari, D., Zheng, S., De Jesus, S., & Lyons, I. M. (2022). Extending ideas of numerical order beyond the count-list from kindergarten to first grade. *Cognition*, *223*, 105019. https://doi.org/10.1016/j.cognition.2022.105019

Johnson, H. L. (1975). The meaning of before and after for preschool children. *Journal of Experimental Child Psychology*, *19*(1), 8899. https://doi.org/10.1016/0022-0965(75)90151-4

Kassambara, A. (2023a). *Ggpubr: 'ggplot2' based publication ready plots*. Retrieved from https://CRAN.R-project.org/package=ggpubr

Kassambara, A. (2023b). *Rstatix: Pipe-friendly framework for basic statistical tests*. Retrieved from https://CRAN.R-project.org/package=rstatix

Kay, M. (2024). ggdist: Visualizations of distributions and uncertainty in the grammar of graphics. *IEEE Transactions on Visualization and Computer Graphics*, *30*(1), 414–424. https://doi.org/10.1109/TVCG.2023.3327195

Lyons, I. M., & Ansari, D. (2015). *Chapter three - foundations of children's numerical and mathematical skills: The roles of symbolic and nonsymbolic representations of numerical magnitude* (J. B. Benson, Ed.). JAI. https://doi.org/10.1016/bs.acdb.2014.11.003

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

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 16. *Developmental Science*, *17*(5), 714–726. https://doi.org/10.1111/desc.12152

Maechler, M. (2024). *Diptest: Hartigan's dip test statistic for unimodality - corrected*. Retrieved from https://CRAN.R-project.org/package=diptest

Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. Retrieved from https://CRAN.R-project.org/package=tibble

Nuerk, H.-C., Wood, G., & Willmes, K. (2005). The Universal SNARC Effect: The Association between Number Magnitude and Space is Amodal. *Experimental Psychology*, *52*(3), 187–194. https://doi.org/10.1027/1618-3169.52.3.187

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. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Robinson, D., Hayes, A., & Couch, S. (2023). *Broom: Convert statistical objects into tidy tibbles*. Retrieved from https://CRAN.R-project.org/package=broom

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., … Schulz, L. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences*, *24*(9), 675–678. https://doi.org/10.1016/j.tics.2020.06.004

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2024). *Afex: Analysis of factorial experiments*. Retrieved from https://CRAN.R-project.org/package=afex

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

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Wickham, H., Hester, J., & Bryan, J. (2023). *Readr: Read rectangular text data*. Retrieved from https://CRAN.R-project.org/package=readr

Wickham, H., Vaughan, D., & Girlich, M. (2023). *Tidyr: Tidy messy data*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155193. https://doi.org/10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220251. https://doi.org/10.1016/0010-0285(92)90008-P

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.org/knitr/

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*(2), 8084. https://doi.org/10.1111/j.1467-8721.2007.00480.x