Cognitive Psychology 78 (2015) 57–77

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Learning the language of time: Children's acquisition of duration words

Katharine A. Tillman [*], David Barner

University of California, San Diego, United States

ARTICLE INFO

Article history:
Accepted 6 March 2015

Keywords:
Word learning
Time
Abstract words
Time words
Time concept development
Duration estimation

ABSTRACT

Children use time words like *minute* and *hour* early in development, but take years to acquire their precise meanings. Here we investigate whether children assign meaning to these early usages, and if so, how. To do this, we test their interpretation of seven time words: *second*, *minute*, *hour*, *day*, *week*, *month*, and *year*. We find that preschoolers infer the orderings of time words (e.g., hour > minute), but have little to no knowledge of the absolute durations they encode. Knowledge of absolute duration is learned much later in development – many years after children first start using time words in speech – and in many children does not emerge until they have acquired formal definitions for the words. We conclude that associating words with the perception of duration does not come naturally to children, and that early intuitive meanings of time words are instead rooted in relative orderings, which children may infer from their use in speech.

1. Introduction

Time is an ephemeral yet central dimension of human experience. The nature of time – and how it is mentally represented – has been a source of fascination for centuries, beginning with early philosophical inquiries (e.g., Augustine, 398/1992; James, 1890; Kant, 1781/2009; McTaggart, 1908), and extending to modern debates in cognitive and developmental psychology (e.g., Bender & Beller, 2014; Boroditsky, 2011; Casasanto, Fotakopoulou, & Boroditsky, 2010; Piaget, 1969; Whorf, 1956).

---

[*] Corresponding author at: UCSD, Department of Psychology, 9500 Gilman Drive #0109, La Jolla, CA 92037, United States. *E-mail address:* katillman@ucsd.edu (K.A. Tillman).

http://dx.doi.org/10.1016/j.cogpsych.2015.03.001

Many of these debates concern the role that natural language plays in mental representations of time. Linguistically, reference to time is pervasive both in speech and in written text (Brysbaert & New, 2009; Kucera & Francis, 1967). However, learning to encode time in language is difficult. Although both temporal syntax (e.g., tense marking) and time-related lexical items (e.g., *before*, *after*, *today*, *minute*) emerge early in children's language production (Ames, 1946; Brown, 1973; Dale & Fenson, 1996; De Villiers & De Villiers, 1978; Grant & Suddendorf, 2011; Harner, 1982; Weist & Buczomska, 1987), the meanings of time words are learned slowly in development, resulting in a long period of frequently incorrect usage and incomplete comprehension (Ames, 1946; Clark, 1971; Grant & Suddendorf, 2011; Harner, 1982; Shatz, Tare, Nguyen, & Young, 2010; Weist, Wysocka, & Lyytinen, 1991). Here, we investigate the lag between production and comprehension of duration words, such as *minute* and *hour*. We ask how children interpret these abstract words before they receive formal instruction, and thus whether they construct interim meanings for time words early in development. Characterizing the initial meanings children assign to time words may contribute both to our understanding of how time is represented in the child's mind and to our understanding of how abstract words are learned more generally.

Time words pose a difficult problem for children during language acquisition. First, although time and its passing are fundamental to experience, duration words like *second*, *minute*, and *hour* carve out relatively arbitrary units that cannot be directly seen or heard. Their boundaries are rarely demarcated explicitly in conversation—and, in fact, can only be precisely indicated via the use of measuring devices like clocks. Second, and relatedly, the units that define such time words are couched in a system of numerical knowledge that children take many years to master. For example, to acquire an adult-like understanding of the word *hour* (i.e., that it contains 60 minutes), children must also learn about minutes (which comprise 60 seconds), and in turn about seconds. In each case, mastery of the number words in question is difficult and protracted, and is typically not achieved until 5 or 6 years of age, if not later (for discussion of the stages of number word development, see Carey, 2004, 2009; Davidson, Eng, & Barner, 2012; Fuson, 1988; Fuson & Hall, 1983; Gelman & Gallistel, 1978; Le Corre & Carey, 2007; Schaeffer, Eggleston, & Scott, 1974; Wynn, 1990, 1992). Third, because duration words depend on numeracy, children are not explicitly taught formal definitions of these time words until very late in development, generally when they enter grade school. In standard US K-12 curricula, an introduction to clocks and time measurement begins in Grade 1, when children are 6 or 7 years of age, and instruction on time telling continues until Grade 3 (Common Core State Standards Initiative, 2010). Finally, children's experience with time words and how they are used in speech is relatively haphazard, given the variety of uses for words like *minute* and *second*. For example, a word like *minute* is frequently used informally in expressions like "just a minute" and "wait a minute," which only rarely reflect precise or accurate durations (Tare, Shatz, & Gilbertson, 2008).

Perhaps because of these challenges, children do not acquire an adult-like understanding of duration words until quite late in development. Given this, two questions arise: When children use these words early in development, what do they mean, if anything? And, if these words have meaning for children, how are these meanings learned? Previous studies suggest that very early on, children recognize that time words are relevant to questions about time and that they form a class of lexical alternatives. For example, when asked questions regarding temporal extent, children typically respond with duration words like *minute* and *hour* despite failing to use them accurately (Shatz et al., 2010). Beyond this, however, remarkably little is known about the acquisition of time words, and what happens between the point when children begin using these words, and when they acquire their formal definitions many years later. One early study reported that at age 6, more than 50% of children had still not learned the precise meaning of the word *hour* (i.e., that an hour is 60 minutes; Ames, 1946), and no studies have documented how children acquire other duration words, like *second* and *minute*, though it is known that their ability to read this information from clocks remains imperfect until age 9 or later (e.g., Friedman & Laycock, 1989).

In the present study, we investigate children's early interpretation of duration words like *second*, *minute*, and *hour*, and ask whether children assign early preliminary meanings to these words before they acquire formal definitions in grade school, and if so, how these meanings are learned. Specifically, we explored two types of meanings that children might assign to duration words early in development: (1) meanings rooted in the perception of approximate duration, and (2) meanings defined by the rank ordering of time words, independent of their actual durations.

Well before acquiring the formal meanings of time words, children – like adults and many non-human animals – exhibit a robust capacity to perceive and discriminate the approximate duration of brief temporal events (e.g., Brannon, Suanda, & Libertus, 2007; Droit-Volet, Tourret, & Wearden, 2004; Droit-Volet & Wearden, 2001; Lewkowicz, 2003; McCormack, Brown, Maylor, Darby, & Green, 1999; Pouthas, Droit, & Jacquet, 1993; VanMarle & Wynn, 2006), much like they can represent approximate numerosity of sets (Izard, Sann, Spelke, & Streri 2009; Lipton & Spelke, 2003; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005). Early work on rats and pigeons, including studies by B.F. Skinner (e.g., Ferster & Skinner, 1957), demonstrated an ability to estimate duration to guide feeding and reinforcement behaviors, governed by Weber's law (e.g., Catania, 1970; Gibbon, 1977; Stubbs, 1968). Human infants also display remarkable sensitivity to differences between brief durations: babies as young as 1.5 months of age adapt to temporal patterns in stimulus presentation and display anticipatory behaviors timed to those stimuli (e.g., Clifton, 1974; Colombo & Richman, 2002; Demany, McKenzie, & Vurpillot, 1977; Fitzgerald & Brackbill, 1968, 1976; Lewkowicz, 2003; Rivière, Darcheville, & Clément, 2000; Trehub & Thorpe, 1989). At 4–6 months of age infants can be trained to discriminate longer vs. shorter sounds (Provasi, Rattat, & Droit-Volet, 2011; VanMarle & Wynn, 2006), a capacity which improves gradually over development. For example, performance on temporal bisection tasks steadily improves between 3 years and 10 years of age (Droit-Volet, Clément, & Wearden, 2001; Droit-Volet & Wearden, 2001; Droit-Volet et al., 2004).

Despite lacking precise meanings for time words, children's early ability to perceive and discriminate duration might allow them to associate time words with approximate duration information – what we call the *Duration Mapping* hypothesis. How might such a link between perception and language be acquired? One possibility is that children make a direct mapping between experienced duration and time words. This might occur spontaneously, as children watch events unfold over time while hearing concurrent adult speech about the durations of those events (e.g., "She's been watching TV for an hour already!"). However, this process might be complicated by the fact that the start and endpoints of events are rarely explicitly marked, either perceptually or in speech. A second possibility is that the association between duration and time words is mediated by children's prior knowledge of the durations of familiar events (Friedman, 1990). For instance, if a parent said to her child, "Dinner is in an hour, so we can only watch one show," the child might constrain her hypothesis about what an hour is by associating it with the known approximate duration of watching one TV show. Thus associative mappings between duration words and temporal magnitudes could be achieved in a two-step process in which events are associated with temporal magnitudes via duration perception, and duration words are associated with events via linguistic input.

Although it is possible that children begin acquisition by associating time words with approximate durations, it is also possible that their first hypotheses about the meanings of these words do not involve nonverbal representations of duration. In addition to picking out absolute durations, time words can also be distinguished according to their rankings within a common lexical class. Hours are longer than minutes, which are longer than seconds. Children might learn this information, without having any knowledge about the precise durations denoted by such words, if they learned their initial meanings on the basis of their contrastive use in language. For example, a child who heard an utterance like, "The whole show lasts an hour, but there are only a few minutes left" might infer that an hour is longer than a minute, without having any sense of the precise duration denoted by either word.

In order to make such inferences, a child would first need to know that words like *minute* and *hour* denote temporal extent, and thus contrast along this dimension (Au & Markman, 1987; Clark, 1990). Consistent with this, children appear to know that duration words belong to a common lexical category by at least the age of 4, long before they learn their individual meanings (Shatz et al., 2010; Tare et al., 2008). For example, when asked "how long" or "how much time" an event takes, many preschoolers are able to respond using domain-appropriate expressions (i.e., using duration words paired with quantity words) despite being inaccurate in their responses (Ames, 1946; Shatz et al., 2010). Because preschoolers know which words in their lexicon refer to duration in general, this knowledge – combined with conversational cues – might allow them to infer that duration words denote contrasting durations, and that their lexical rank ordering reflects a corresponding rank ordering of relative duration – e.g., that a year is longer than a month, which is longer than a week, and so on. We call this the *Lexical Ordering* hypothesis.

Though prior work has shown that preschoolers possess incomplete comprehension of duration words, these studies cannot differentiate between the Duration Mapping and Lexical Ordering hypotheses. For example, in one recent study, children's early comprehension of duration words was assessed by probing how well they could match familiar activities (e.g., "a boy eating breakfast") to adult-estimated approximate durations of these activities (e.g., "ten minutes"). This study showed that 5-year-olds performed just above chance, demonstrating a rudimentary understanding of duration expressions and how they relate to familiar activities. Even 6- to- 7-year-old children, some of whom presumably had received instruction on time word definitions, were only 67% accurate on this binary-choice task (Shatz et al., 2010). These results clearly demonstrate that children struggle to relate conventional expressions of time to event durations. However, the findings are difficult to interpret because each prompt combined duration words, number words, and particular events (e.g., "learning to surf"). Children could succeed (or fail) at the task based on their level of understanding in any of these domains. Thus it is unclear if knowing the rank ordering of duration words would be sufficient for success, or if knowledge of their absolute durations is needed. In fact, this could vary based on a child's knowledge of the events and number words in question. Furthermore, though prior studies have shown that children as young as 4 years old can rank-order events in terms of duration by placing markers onto a spatial representation of time (Friedman, 1990), these studies did not test conventional duration words like *minute* and *hour*. Thus it is unclear whether children can use a spatial scale to represent the rank ordering and durations of conventional time words.

In the present study, we therefore focused specifically on how children represent the rank order and absolute durations encoded by words like *second*, *minute*, and *hour*. In Experiment 1, we tested whether children know the temporal orderings of duration words. Children answered forced choice questions in which duration words were the only cue that could support judgments (e.g., "Farmer Brown slept for a minute. Captain Blue slept for an hour. Who slept more?"). Success on this task indicates knowledge of the rank ordering of duration words, but would also be consistent with knowing their absolute durations. Experiment 2 provided a first test of whether children are sensitive to the difference in absolute duration between terms, by testing how well children can combine duration words and number words (e.g., "Farmer Brown slept for three minutes. Captain Blue slept for two hours. Who slept more?"). While success in Experiment 2 is consistent with knowledge of approximate durations, failure could also be attributed to syntactic difficulties combining number words and duration words, or to difficulty with number words more generally. In Experiment 3, we used a number-line task to test the Duration Mapping hypothesis more directly, asking how well children can estimate the durations of familiar events and of conventional time words, on the logic that accurate estimation of duration words could not be achieved without associations between these words and approximate durations.

On the Lexical Ordering account, we predicted that children should be able to correctly compare *hour* vs. *minute* before having representations of the approximate duration represented by each word, and that mappings to duration may not arise until children have learned the precise definitions of time words. On the Duration Mapping account, which posits that each duration word is linked to an approximate magnitude, we predicted that before learning their precise definitions, children should not only know that an hour is longer than a minute, but also approximately *how much* longer it is (i.e., $hour \approx minute \times 60$).

2. Experiment 1

The goal of Experiment 1 was to test children's knowledge of the rank ordering of time words – e.g., when they learn that an hour is longer than a minute. To do this, we tested 3- to 6-year-old children using a forced choice task in which they were asked to decide which of two verbally described events was longer – e.g., "jumping for a minute" vs. "jumping for an hour." Unlike in previous studies of time word comprehension (Shatz et al., 2010), children could not use their knowledge of the typical durations of events to guide their choices, nor could they rely on their knowledge of number words (because none were included in the prompts). To succeed at this task, children must know the rank ordering of the duration words, within their common lexical class.

2.1. Methods

2.1.1. Participants

Subjects were 92 children recruited from the San Diego area, including 25 3-year-olds (mean age = 3;6), 26 4-year-olds (mean age 4;6), 20 5-year-olds (mean age 5;5), and 21 6-year-olds (mean age 6;5). An additional 4 children participated but were excluded from analysis due to failure to complete the task. Children were either recruited by phone and tested in our UCSD laboratory, or tested in local daycares, preschools, and museums. Parents gave informed consent for their children to participate in the study, and children indicated their willingness to play a game with the experimenter before testing began. Parents who brought their children into the laboratory were compensated for their travel expenses, and children received a small gift in thanks for their participation.

2.1.2. Procedure

The child was first introduced to two action figures, Farmer Brown and Captain Blue, who were placed on a table in front of the child. On each trial, the experimenter read a short vignette of the form, “Farmer Brown [jumped] for [a minute]. Captain Blue [jumped] for [an hour].” This was followed by a two-alternative forced choice, “Who [jumped] more, [Farmer Brown or Captain Blue]?” If the child was reluctant to give a verbal response, they were encouraged to point to the character who did the action more. The experimenter then proceeded to the next trial.

Children completed two blocks of trials, each containing thirteen duration word comparisons. Seven time words were tested: *second*, *minute*, *hour*, *day*, *week*, *month*, and *year*. The specific comparisons tested were: week vs. month, day vs. week, month vs. year, hour vs. day, day vs. month, week vs. year, minute vs. hour, second vs. minute, hour vs. week, day vs. year, minute vs. day, second vs. hour, and second vs. day. Each comparison appeared once in each block, and each duration word described an activity that could in principle be done for any length of time. These activities were described using the past tense form of six high-frequency action verbs: *jump*, *sleep*, *cry*, *play*, *dance*, and *talk*. Critically, on a given trial, the same verb was paired with both duration words. Within each block, trials were presented in quasi-random order. Verbs were pseudo-randomly assigned to duration comparisons, with the stipulation that the same verb was never used in two consecutive trials. Trials were counterbalanced with respect to whether the larger duration word came first, which character performed the longer action, and which character was mentioned first in the question prompt. Half the participants received one item-order, and the other half received the reverse order. For analysis, the child’s response on each trial was coded as correct (1) or incorrect (0). These numbers were then converted into proportions correct for group-wise comparisons.

2.1.3. Analyses

Linear mixed-effects analyses were performed in R (R Core Team, 2013) using the *lme4* (Bates, Maechler, Bolker, & Walker, 2013) extension package. Unless otherwise noted, *p*-values were obtained by likelihood ratio tests of the full model in question against the model without the effect in question.

2.2. Results and discussion

Our first analysis examined whether children knew the rank ordering of duration words. To assess children’s overall understanding of the rank ordering of duration words (e.g., their knowledge of which words indicate longer times than others), we first calculated each child’s proportion of correct responses, across all trials and duration word comparisons. An ANOVA revealed a significant effect of age group on overall accuracy ($F(3,84) = 23.1$, $p < 0.001$). The youngest group of participants, the 3-year-olds, did not perform better than 50% accuracy, consistent with random guessing, M (SEM) = 0.48 (0.02), $t(24) = 1.2$, n.s. The 4-, 5-, and 6-year-old groups all performed significantly better than chance (t’s > 2; p’s < 0.05) and significantly better than each younger group: 4-year-olds, *M* (SEM) = 0.57 (0.02); 5-year-olds, *M* (SEM) = 0.67 (0.04); 6-year-olds, *M* (SEM) = 0.82 (0.03); *t*'s > 2; *p*'s < 0.05. Thus, beginning at 4 years of age, children exhibited partial knowledge of the rank ordering of duration words. However, somewhat surprisingly, even our oldest age group (6-year-olds) frequently failed to judge, for example, that a minute was longer than a second, *M* (SEM) = 0.68 (0.09); that a day was longer than an hour, *M* (SEM) = 0.65 (0.09); or that a year was longer than a month, *M* (SEM) = 0.60 (.08).

While this result shows that children have some knowledge of the rank ordering of duration words, it is also consistent with knowledge of their absolute durations. If children have knowledge of the absolute durations of time words (even if rudimentary), we might expect them to perform better on comparisons in which their magnitudes differ more in scale. To explore this, we next asked whether children's performance on a given trial was affected by the difference in absolute duration represented by the two terms being contrasted – e.g., if *hour vs. second* was easier than *minute vs. second*. To test this possibility, we conducted a mixed effects logistic regression testing whether the ratio between the terms in each comparison (e.g., *minute vs. second* = 60, *hour vs. second* = 360) predicted the accuracy (correct vs. incorrect) of children's responses. We entered age and duration-ratio as fixed effects in the model, and as random effects we included intercepts for subjects, as well as by-subject random slopes for the effect of duration-ratio. We observed an effect of duration-ratio that came in the form of an interaction with age, $\chi^2(2) = 3.1$, $p = .04$. To investigate this interaction, we analyzed the data from each age group separately. We found no effect of duration ratio on accuracy in the 3- or 5-year-olds groups, χ^2's < 0.7, all *p*'s > 0.4, a *negative* effect in the 4-year-olds, $\beta = -0.0001$, $z = -2.9$; $\chi^2 = 7.6$, $p < 0.01$, and a small positive effect of duration ratio in the oldest age group, the 6-year-olds, $\beta = 0.09$, $z = 2.6$; $\chi^2(1) = 5.8$, $p < 0.01$. In both cases, the model estimates (β) for this factor were tiny. Thus, while increasing difference in duration between two time words did not predict better performance among children under age 6 – i.e., they were not better at comparing *second vs. day* than they were *second vs. minute* – our results suggest that knowledge of the absolute durations of these words may begin to emerge at age 6.

If children know the rank of each term in their list of known duration words, but not their approximate magnitudes, their performance should be better predicted by the difference in ranking between the two terms. For example, 'minute' and 'second' differ by a rank of 1, while 'day' and 'second' differ by a rank of 3. In an analysis similar to the one described above, we entered this rank difference factor into our model in place of the duration-ratio. Here, we again found a significant interaction of this factor with age $\chi^2(2) = 11.2$, $p < 0.001$. In follow-up analyses, we found that while this factor also did not predict performance in the 3- and 4-year-old groups (all $\chi^2 < 2.4$, all $p > 0.05$), it had a marginally-significant positive effect on performance in the 5-year-old group ($\beta = 0.34$, $z = 2.0$; $\chi^2 = 3.4$, $p = 0.07$) and a highly-significant effect in the 6-year-old group ($\beta = 0.89$, $z = 3.6$; $\chi^2 = 10.8$, $p = 0.001$). This pattern of effects suggests that by age 5 children's knowledge of the rank ordering of duration words guides their ability to contrast them. However, it should be noted that none of the comparisons we tested differed more than a rank of 3, and further experiments with a wider range of rank differences would be needed to best assess the role of this factor.

Our final analysis examined whether individual words exhibited different learning trajectories. It is possible that, although children begin acquisition by ranking time words according to their relative magnitudes, they nevertheless have some absolute magnitude knowledge for some words, but not for others. To test for differences between duration words (item effects), we collapsed the data across all comparisons involving each word, yielding an accuracy score for each child for each word. Means and standard errors for each age group are depicted in Fig. 1. A mixed-effects ANOVA with age group as the between subjects factor and duration word as the within subjects factor revealed a main effect of age ($F(3,547) = 24.3$, $p < 0.001$) and a marginal effect of duration word ($F(6,547) = 2.1$, $p = 0.05$), with no significant interaction. However, when we analyzed the data from each age group separately, an effect of duration word was found only in the 3-year-old group, who, despite showing chance performance overall, nevertheless performed better on comparisons involving the word *week* (see Fig. 1). Given the overall failure of this group, this finding is difficult to interpret and is perhaps not a reliable effect. More importantly, there was no effect of duration word on accuracy within the 4-, 5-, or 6-year-olds groups (all *F*'s < 1.6, all *p*'s > 0.05). This finding suggests that as accuracy improves over these years, it may improve across the board, with equal improvement on each tested word (Fig. 1). However,

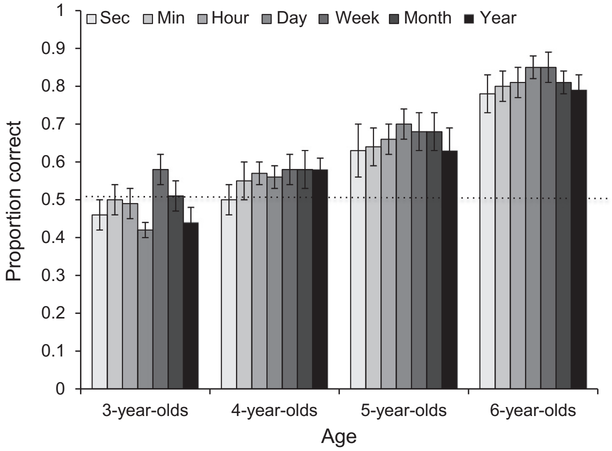

Fig. 1. Proportion of correct responses on all trials in Experiment 1 containing the indicated time word, for each age group. Error bars represent SEM. Dashed line indicates chance performance.

because each trial in this experiment tested children's knowledge of two duration words, differences between individual words may have been masked. We will return to this issue in Experiment 3.

3. Experiment 2

In Experiment 1, we found that by around age 4 many children know the rank order of some duration words, but that this learning process is gradual, and not yet complete in 6-year-olds. Also, surprisingly, we found no systematic differences in children's understanding of individual time words, and no evidence that children under 6 were more accurate when the time words that they compared differed more in duration.

These preliminary data suggest that children acquire the rank ordering of time words between the ages of 4 and 7, but that they may not be sensitive to their absolute durations until quite late in development. In Experiment 2 we explored this possibility further, by introducing number words into judgments like those used in Experiment 1. In addition to simply comparing *hour* to *minute*, the task required children to compare, e.g., *3 minutes* to *2 hours*. We reasoned that if children have an approximate understanding of the duration of an hour, then they should know that 2 hours is much longer than 3 minutes, so long as they also understand the meanings of 2 and 3. Although children in the US generally comprehend such numbers and can accurately compare them by the age of 3 or 4, we nevertheless verified that children were able to compare cases in which the larger number was paired with the longer duration word – e.g., *3 hours* vs. *2 minutes*.

3.1. Methods

3.1.1. Participants

We tested 93 children, including 25 4-year-olds, 22 5-year-olds, 25 6-year-olds, and 21 7-year-olds. Children were recruited from the same population as those in Experiment 1. An additional 8 children participated but were excluded from analysis due to failure to complete the study (7) and experimenter error (1). None had previously participated in Experiment 1.

3.1.2. Procedure

The procedure for Experiment 2 was identical to that of Experiment 1, except that the time words were modified by number words. For example, "Farmer Brown [jumped] for [two] [minutes]. Captain Blue [jumped] for [three] [hours]. Who jumped more?" Each child completed a total of 30 trials.

Table 1
Trial types used in Experiment 2.

Number comparison	Number size	Example
No numbers	None	A minute vs. an hour
Same	Small	2 minutes vs. 2 hours
	Large	6 minutes vs. 6 hours
Congruent	Small	2 minutes vs. 3 hours
	Large	6 minutes vs. 9 hours
Incongruent	Small	3 minutes vs. 2 hours
	Large	9 minutes vs. 6 hours

Trials in Experiment 2 included the same six verbs from Experiment 1. However, only five time–word comparisons were used in Experiment 2: minute vs. hour, week vs. year, day vs. year, day vs. week, and second vs. hour. Findings in Experiment 1 indicated that children's performance on these comparisons was similar. For each time–word pair, 7 different types of number–word comparisons were made (Table 1). For each pair, one trial included no numbers (identical to Experiment 1), 3 included "small" numbers (2 and/or 3), and 3 three included "large" numbers (6 and/or 9). Each comparison was designated Same, Congruent, or Incongruent as defined in Table 1. Trials were presented in quasi-random order. Half the participants received one item-order while the other half received the reverse order.

3.1.3. Analyses

Linear and logistic mixed-model analyses were conducted as in Experiment 1.

3.2. Results and discussion

To determine which experimental variables influenced children's performance in Experiment 2, we first conducted a mixed-effects logistic regression to predict the accuracy of each child's response (correct vs. incorrect) using the participant's age, the duration word comparison type (e.g., *hour/s vs. minute/s*), the number size (none, small, or large), and the number comparison type (same, congruent, or incongruent) as fixed effects. As random effects, we had intercepts for subject, as well as by-subject random slopes for word comparison, number size, and number comparison type. The only significant main effects were those of age group, $\chi^2(3) = 67.9$, $p < .001$, and number comparison type, $\chi^2(2) = 16.5$, $p < 0.001$. Because there were no significant effects of word comparison or number size, χ^2's < 5.6, p's > 0.2, the data were collapsed across these factors in all subsequent analyses.

A linear mixed effects analysis of the collapsed dataset (with age group and number comparison type as fixed effects, random intercepts for subject, and by-subject random slopes for the effect of comparison type) revealed that the effect of comparison type on accuracy was driven by an interaction with age, $\chi^2(6) = 29.9$, $p < 0.001$. Mean accuracy for each age group on the Same (including no-number), Congruent, and Incongruent comparisons is depicted in Fig. 2. As shown, the 4-, 5-, and 6-year-old groups all performed significantly worse on the critical Incongruent trials than on the Congruent trials (all t's > 2.9, all p's < 0.005), and the 4- and 5-year-old groups also performed significantly worse on Incongruent trials than on Same number trials (t's > 2.2, p's < 0.05). 6-year-olds did not perform significantly worse on Incongruent trials than Same trials, though performance still remained far from adult-like at 74% correct. In contrast, the 7-year-old children's performance was consistently near ceiling and unaffected by comparison type, $F(2,20) = 0.25$, $p = 0.8$, n.s., indicating a much more robust understanding of the durations being contrasted.

To understand these data it is useful to consider a concrete example. The results of Experiment 2 show that many 5-year-olds knew both that 3 is greater than 2 (76% correct overall on Congruent trials) and that an hour is longer than a minute (75% correct overall on Same trials), but performed significantly worse when judging that 2 hours is more than 3 minutes (61% correct overall on Incongruent trials). Given that the ratio between an hour and a minute greatly exceeds 3:2, this failure

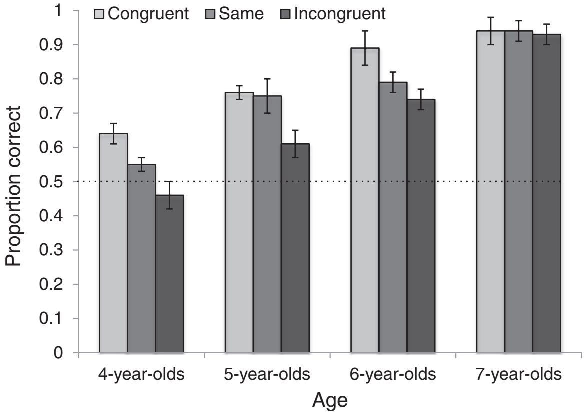

Fig. 2. Effect of congruency between time word comparisons and number word comparisons in Experiment 2.

indicates that children often cannot combine their knowledge about duration words with their knowledge about number words, suggesting that children's early meanings for duration words may not include information about approximate duration.

4. Experiment 3

The results of Experiments 1 and 2 suggest that while some children learn the rank ordering of duration words by age 4, most children do not associate these words with even approximate durations until as late as 6 or even 7 years of age. Many 6-year-olds in Experiment 2, for example, did not know that 2 hours was longer than 3 minutes.

A limitation of the forced-choice method used in the first two experiments is that each trial probed the participant's knowledge of two different duration expressions. A problem with this approach is that data for individual words are non-independent, making it more difficult to determine whether the words have absolute meanings for children that do not depend on comparison to one another. To further probe children's knowledge of the absolute durations encoded by individual time words, Experiment 3 tested children using a task that required them to place individual duration expressions onto a time line representing increasing duration. We asked children and adults to estimate the durations of familiar events (e.g., "watching a movie"), duration words (e.g., *hour*), and duration words paired with numbers (e.g., *3 hours*), using this spatial representation of time. This allowed us to disentangle children's ability to reason about relative duration (e.g., of events) from their knowledge of linguistic labels for durations (i.e., words like *minute*). Also, because an adult-like understanding of time words requires reasoning about large precise numerosities (e.g., quantities of 60), we tested each child with a standard number line task, in which they were required to place numbers on a line representing values from 0 to 100.

Finally, we also tested how children's understanding of duration words is related to their knowledge of formal definitions. We hypothesized that children may first represent the approximate durations of time words when they are formally taught the meanings of each word – e.g., that an hour equals 60 minutes, and a minute equals 60 seconds. We tested this hypothesis by asking children follow-up questions regarding the definitions of *second*, *minute*, *hour*, and *day* – e.g., "How many seconds are in a minute?".

4.1. Method

4.1.1. Participants

We tested 64 children, including 22 5-year-olds (*M* = 5;6), 21 6-year-olds (*M* = 6;6), and 21 7-year-olds (*M* = 7;6). We also tested 36 adults (*M* = 20;7). One additional child participated but was excluded from analysis due to failure to complete the task. Children were from the same population as those in Experiments 1 and 2. Twelve children had previously participated in either Experiment 1 or 2, on another day. Adults were members of the UCSD Psychology Department subject pool and received course credit for participation.

4.1.2. Procedure

Participants were given a sheet of $8.5' \times 11'$ paper with four horizontal, 17-cm lines printed in a vertical column down the center of the page. Each line had small, filled dots on both endpoints and no other markings (i.e., the midpoint of the line was not marked). The participant's task was to estimate the magnitude or duration of various stimuli by drawing vertical marks bisecting each line using colored pencils.

The first line was a standard number line. The experimenter provided the following instructions to the participant: "This is a number line. Each number has its own place on the line. You're going to show me where certain numbers go on the number line. Look, 0 goes here [to demonstrate, the experimenter drew a vertical line bisecting the left endpoint] and 100 goes here [the experimenter marked the right endpoint]." Then, for each of four number stimuli (see Table 2), the experimenter said, "The [first] number is [4]. Can you show me where [4] goes? Can you draw a line with the [blue] pencil?" A different colored pencil was used for each target, so the data could be easily interpreted later.

Next, children were presented three time lines, in order, which tested (1) event durations (e.g., "eating lunch"), (2) time word durations (e.g., "an hour"), and (3) numerically modified time words (e.g., "9 minutes"). Target items for Experiment 3 are shown in Table 2. Before the first duration estimation task, the line was explained to the participants as follows: "Now, this line is different. It shows how much time things take to do. It goes from a very short amount of time to a very long amount of time. Each amount of time has its own place on the line, and the further you go over here [experimenter gestured along the line from left to right], the more time something takes. You're going to show me how long certain things take to do on the line. Something very short, like blinking your eyes, goes here [experimenter marked left endpoint]. Something very long, like the time from waking up in the morning to going to bed at night, goes here [experimenter marked right endpoint]." For each item (see Table 2), the child was instructed to think about how long the activity takes to do and to mark the line accordingly. Participants were reminded that each subsequent line represented duration and what the endpoints represented ("short, like blinking your eyes" on the left and "long, like the time from waking up in the morning to going to bed at night" on the right) in between each of the remaining tasks and any time they indicated confusion.

Because each child was tested with four lines – a substantial battery especially for the youngest children – questions were limited to four items per line. Each child completed the series of lines in the same order: (1) numbers, (2) events, (3) time words, and (4) time words with numbers. Within each line, half the participants received the four stimuli in the order shown in Table 2, while the other half received the reverse order. As in Experiments 1 and 2, participants were presented with time word stimuli (lines 3 and 4) in the context of events that could take variable amounts of time, e.g., "[jumping] for a minute." The same action verbs were used as in Experiments 1 and 2.

Following completion of the four number line and timeline tasks, each participant was asked 3 follow-up questions: "How minutes are in an hour?", "How many hours are in a day?", and "How many seconds are in a minute?".

Table 2
Number-line and timeline stimuli in Experiment 3.

Line 1: Numbers	Line 2: Events	Line 3: Time words	Line 4: Number + time
4	Watching movie	Hour	2 hours
45	Washing hands	Second	6 hours
18	Trip to zoo	Minute	9 min
61	Eating lunch	Day	3 min

4.1.3. Coding and analyses

To analyze the time line and number line data, we first measured the distance (measured as centimeters from the 0 point) from the left endpoint to the intersection of the number line and each participant's pencil marks. Marks falling exactly on the left endpoint were recorded as 0.1 cm (to avoid divide-by-zero errors) and those falling exactly on the right endpoint were recorded as 17.0 cm.

To assess children's knowledge of the *rank ordering* of events and time words, responses to the four stimuli for each line were rank-ordered by increasing magnitude or duration. For duration words, the order was: 1 = second; 2 = minute; 3 = hour; 4 = day. For events, the adults' modal order ($N = 36$) was: 1 = washing hands ($N = 36$); 2 = eating lunch ($N = 35$); 3 = watching a movie ($N = 34$); 4 = going on a trip to the zoo ($N = 35$). For each estimated item which fell in the correct rank, the participant was awarded a 1, for each incorrectly ranked item, the participant was given a 0, resulting in a score out of 4 for each line.

To assess number estimation performance, participants' estimates (measured as centimeters from the 0 point) were converted to their corresponding values from 0 to 100 and plotted as a function of magnitude being estimated. The endpoints on the number line were 0 and 100, and the highest number children were asked to estimate was 61.

To assess children's knowledge of the relative durations of events and time words, we computed ratios between the estimates (termed "estimate ratios") for each possible pair of stimuli (e.g., min/sec, hour/sec, hour/min, day/sec, day/min, day/hour), and plotted the children's estimate ratios as a function of the mean estimate ratios provided by our adult sample.

As in Experiments 1 and 2, linear mixed-effects analyses were performed in R (R Core Team, 2013) using the *lme4* (Bates et al., 2013) extension.

4.2. Results and discussion

4.2.1. Rank ordering analysis

To corroborate the findings of Experiments 1 and 2, we first assessed children's knowledge of the rank ordering of numbers, events, and time words (e.g., second < minute < hour < day). For each task (numbers, events, time words, and time words with numbers), we calculated the proportion of items each child placed in the correct rank relative to the others, yielding an accuracy score for that line. Mean proportions of correctly ordered estimates for each age group on each task are shown in Fig. 3. We conducted a linear mixed effects analysis of children's accuracy in which with age and task were entered as fixed effects, and random effects included random intercepts for subjects and random

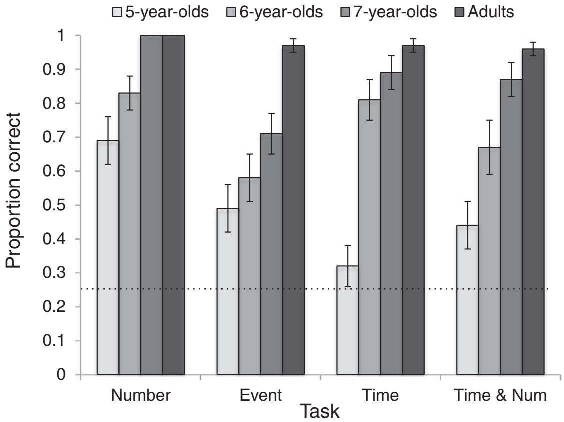

Fig. 3. Proportion of correctly ranked estimates for each age group in Experiment 3. Tasks: *Number* = number line estimation, *Event* = familiar event duration estimation (no time words), *Time* = duration word estimation (e.g., minute), and *Time & Num* = duration words modified by number words. Error bars indicate SEM. Dashed line indicates chance performance, one item falling in the correct rank.

by- subject slopes for the effect of task. This model revealed significant effects of age group and task, driven by a significant interaction between them, $\chi^2(9) = 37$, $p < 0.001$. All age groups performed best when ordering number words, indicating that children understand the number line paradigm and that their difficulty ordering duration expressions cannot be attributed to difficulty with numbers. Overall, children also demonstrated substantial understanding of the rank ordering of time words, despite the increased difficulty of this task relative to Experiments 1 and 2. Strangely, though the youngest group (5-year-olds) did not perform significantly above chance (0.25) when ranking unmodified duration words, $t(21) = 2.8$, n.s., this group performed above chance when time words were modified by (incongruent) number words, $t(21) = 2.8$, $p = .01$. However, because the stimuli were designed so that the number words gave incorrect cues to which expression denoted the longer duration (see Table 2; e.g., "9 seconds," despite having the largest number word, is the second-shortest duration), we cannot attribute this result solely to number word understanding. By age 7, children rank ordered duration words both with and without number words as well as adults (t's < 1.5, p's > .1).

Surprisingly, we found that children in all three age groups performed relatively poorly when ranking familiar events (described without time words) by increasing duration, and that even the oldest children performed much more poorly than adults, $t(25) = -4.4$, $p < 0.001$. Given their poorer ability to rank familiar events by duration, it seems highly unlikely that children's learning of the rank ordering of duration terms is mediated by knowledge of the approximate durations of events (e.g., that children learn "an hour" by mapping it to events described as "an hour", and noting that duration of those events).

4.2.2. Number estimation

The previous analysis tested children's ability to order expressions, without regard to whether these estimates were accurate. We next analyzed children's ability to accurately place numbers on the number line. Following previous number line estimation studies (e.g., Barth, Starr, & Sullivan, 2009; Booth & Siegler, 2006; Siegler & Opfer, 2003; Sullivan & Barner, 2014a), we assessed performance by plotting subjects' estimates as a function of the numbers being estimated. As the slopes of the best-fitting linear regression through the data for each age group (with subject as a random factor) approach 1, estimation becomes more accurate. As indicated by the steep slopes of their estimation functions (Table 3), we found that by age 7, children estimated numbers in this range no differently than adults, $t = 1.8$, n.s. Six-year-olds' performance was somewhat lower than that of adults, $t = 4.1$, $p < .001$, but still quite high ($\beta = 0.9$). Five-year-olds performed more poorly, but also produced estimates that were correlated with the magnitudes they were asked to estimate. These data are similar to those found in other developmental studies of number-line estimation for magnitudes up to ~50 (e.g., Barth & Paladino, 2011; Barth et al., 2009; Slusser, Santiago, & Barth, 2013; Sullivan & Barner, 2014a). Critical to the present study, the results suggest that failure to accurately estimate duration cannot be attributed to an inability to place quantities on a line or to associate number words with approximate magnitudes.

4.2.3. Duration estimation

Having established that all groups were relatively competent at placing number words accurately on a number line, we next evaluated children's ability to accurately estimate the durations of events, time words, and time words with numbers on timelines.

Table 3
Regression analysis of number-line estimation.

Task	Age group	β	SEM	R^2
Number-line	5 YO	0.61	0.11	0.22
	6 YO	0.90	0.05	0.60
	7 YO	0.99	0.02	0.83
	Adult	1.0	0.03	0.90

Note: A mixed ANOVA on children's number estimates found main effects of the magnitude being estimated, age group ($p < 0.05$), and a significant interaction between the two (all F's > 7.6, all p's < 0.05).

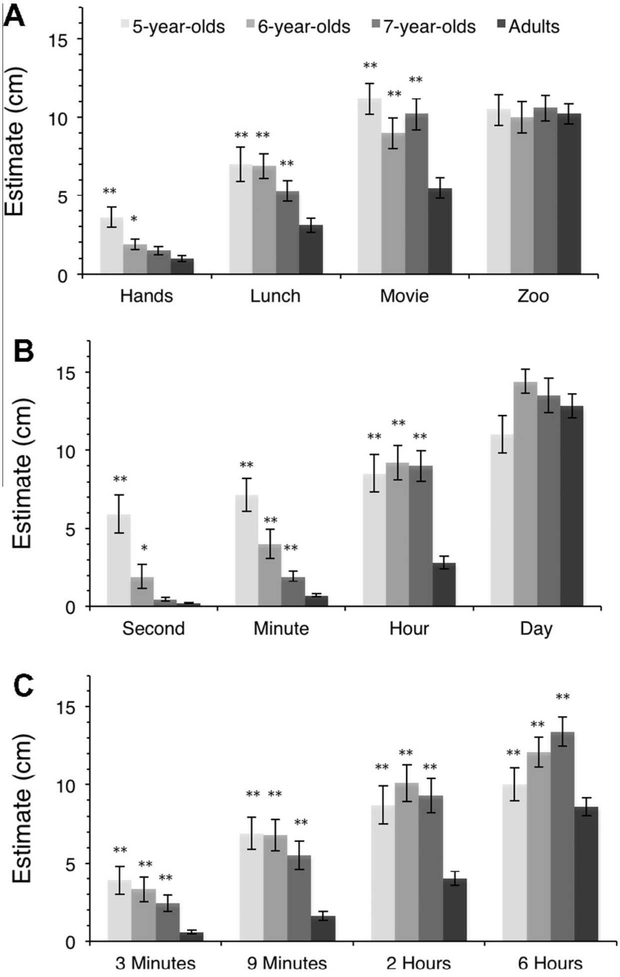

Fig. 4. Mean duration estimates for events (A), time words (B), and time words with numbers (C). Estimates (cm) were defined as the distance from the left endpoint of the line. Asterisks indicate the significance level of the difference between children's estimates and adults', $^{*}p < 0.05$, $^{**}p < 0.005$.

The mean estimates (distance from the left endpoint of the line) given by each age group for each temporal item are shown in Fig. 4. Children of every age generally overestimated duration substantially. Within each of the three duration estimation tasks, linear mixed-effects analyses of participants' raw duration estimates (with age group and item as fixed effects, random intercepts for subjects, and random by-subject slopes for the effect of item) revealed highly significant main effects of age and item (e.g., 'hour'), as well as significant interactions between them, all χ^2's < 25; all p's < 0.001. Here, the effect of item indicates that participants distinguished the terms from one another on the line; if they were simply guessing randomly, we would expect to see no effect. Although there were a few cases in which children's estimates did not differ significantly from adults' (see Fig. 4), overall children were poor at representing absolute durations on timelines.

To compare children's knowledge of durations of familiar events and those denoted by duration words, we divided each duration estimate given by each child by the mean estimate given by the adult sample on that item. A linear mixed-effects analysis of these normalized estimates (with age group and task as fixed effects, and random intercepts for subjects and random by-subject slopes for the effect of task) revealed effects of age group and task as well as a significant interaction, all χ^2's > 5.8, all p's < .001, and post-hoc tests showed that children's estimates of event duration were significantly more adult-like than those of duration words, both with and without numbers, t's > 4, p's < 0.001.

4.2.4. Relative duration analysis

In the previous analysis we found that children differed substantially from adults when estimating durations on timelines, and that they were poorer at estimating the durations indicated by time words like *minute* than those of familiar events. However, it is possible that some children, despite failing to represent absolute duration like adults, might nevertheless have knowledge of the approximate *relative* durations that they encode. For example, despite placing an hour at a different location on the time line than adults, they might also place a minute at a correspondingly different location, and thus show evidence of knowing that an hour is roughly 60 times longer than a minute.

To test this possibility, we calculated ratios between the durations estimated for each pair of stimuli (e.g., minute/second, hour/minute), and used these "estimate-ratios" as dependent measures in our analysis. To assess children's knowledge of the relative durations of events and time words, we plotted each group of children's estimate-ratios as a function of the adults' mean estimate-ratios (Table 4).[1] Similar to the number line analysis described above, greater steepness in the slope of that linear model indicates more adult-like performance, in this case reflecting that children spaced their estimates on the line more similarly to adults.

As indicated by their slopes in Table 4, children demonstrated strong understanding of the relative durations of familiar events and much poorer knowledge of duration words. Despite the fact that children's raw estimates of event durations differed from adults', by age 5 children appear to understand how these durations relate to one another (e.g., that it takes roughly three times as long to visit the zoo as to watch a movie). This result suggests that preschoolers' representations of familiar events include mappings to approximate durations. In contrast, young children demonstrated poorer understanding of the relative durations of time words, though their performance increased significantly with age. In particular, a pronounced improvement was observed in the 6-year-olds relative to the 5-year-olds, suggesting that children may begin to map duration words onto approximate durations around this age. However, when numbers modified duration words, all children performed much more poorly (Table 4), and there was only marginal improvement with age. The shallow slopes for both 5- and 6-year-olds, followed by only a modest gain at 7, indicate that a more sophisticated understanding of the proportional relationships between these terms – like that required to read and interpret a clock – emerges quite late. This finding has been borne out in studies on children's ability to tell time, which remains imperfect until late in the elementary school years (e.g., Friedman & Laycock, 1989).

Here, children demonstrated greater knowledge of both the absolute and relative durations of events than of time words. Note that this effect is opposite our findings in the rank ordering analysis discussed above, despite the fact that both analyses were performed on the same raw data. The rank ordering analysis indicated that children are better able to rank duration words than to rank events by increasing duration (Fig. 3). These contrasting findings suggest that, although children were more likely to place duration word estimates in the correct sequence, those placements were quite poorly "spaced out" along the line. For example, a child who "bunched up" all four item placements near one end of the line, with no order errors, might demonstrate poorer knowledge of relative duration than a child who made a reversal in order but had more appropriate item spacing overall.

4.2.5. Definition knowledge

By the time children begin to show rudimentary knowledge of the absolute durations of time words (age 6 or 7), most have entered school. In Grade 1, most elementary school curricula include instruction on the meanings of time words (Common Core State Standards Initiative, 2010). Thus our final question was what role children's knowledge of these definitions might play in their ability to estimate the durations of time words and to order them. We addressed this by asking children to

---

[1] Note that for time words it is possible to use the true ratios between items, e.g., 'day/second' = 86,400, as predictors of performance. We chose not to use this measure since for some items, like 'second' and 'minute', a difference could not be reliably represented on the line. Nevertheless, using this metric of performance generates the same pattern of results.

Table 4
Regression analysis of relative duration estimates (Exp. 3).

Task	Age group	β	SEM	R^2
Event	5	0.86	0.15	0.19
	6	0.85	0.13	0.25
	7	0.88	0.1	0.4
	Adult	1.0	0.05	0.64
Time word	5	0.15	0.03	0.15
	6	0.68	0.09	0.31
	7	0.87	0.09	0.44
	Adult	1.0	0.04	0.73
Time word w/number	5	0.25	0.05	0.16
	6	0.19	0.02	0.49
	7	0.46	0.04	0.5
	Adult	1.0	0.07	0.51

Notes. Estimate-ratios were regressed against the mean estimate-ratios produced by adults. The y-intercept was fixed at 0 for all regressions. Analyses of variance revealed significant effects of item in all three tasks, a significant effect of age in the time word task only, and interactions between these factors in the time word and time word with numbers task (all F's > 2.5, all p's < 0.01).

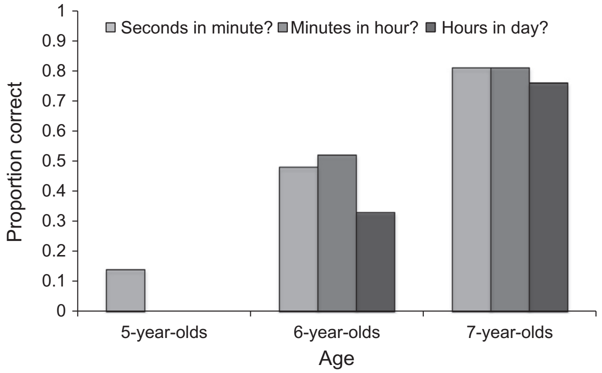

Fig. 5. Proportion of children in each age group who correctly answered three follow-up questions about the definitions of time words: "How many seconds are in a minute?," "How many minutes are in an hour?," and "How many hours are in a day?".

answer three final follow-up questions: "How many seconds are in a minute?", "How many minutes are in an hour?", and "How many hours are in a day?". The proportion of children in each age group who correctly answered each question is shown in Fig. 5. We observed drastic age differences in children's knowledge of duration word definitions (Fig. 5). While 75% of 5-year-olds knew no definitions and none knew more than one, 90% of 7-year-olds knew two or three definitions. The 6-year-old sample was more varied – with 55% knowing 0 or 1 definition, and 45% knowing 2 or 3.

Next, we tested the possibility that definition knowledge supported children's ability to estimate relative durations. In a linear mixed effects analysis of children's duration estimate-ratios, we entered definition knowledge (0–3 definitions) as a fixed effect, along with the adult mean estimate-ratio, age group, and task (events, time words, time words with numbers). We included random intercepts for subjects as well as by-subject random slopes for the effect of definition knowledge and the effect of task. This model revealed a main effect of the adult estimate and a significant interaction between definition knowledge and task, $\chi^2(2) = 12.9$, $p < 0.001$.

In follow-up analyses of the effect of definition knowledge specifically in the duration word estimation task, we found that when this factor was added to the model predicting estimate-ratios, there was a significant effect of definition knowledge $\chi^2(1) = 12.5$, $p < 0.001$. When age and definition knowledge were in the model, the effect of age was non-significant, $\chi^2(2) = 3.5$, $p = 0.18$, in contrast to models that did not include definition knowledge, $\chi^2(2) = 12.6$, $p < 0.005$). However, even when age was included, the effect of definition knowledge remained marginally significant, $\chi^2(1) = 3.5$, $p = 0.06$. Thus, knowledge of definitions completely explained differences between age groups on this task. This finding, which was not observed for events or for time words with numbers, suggests that children's improvement on the duration word estimation task may be related to increased knowledge of time word definitions between the ages of 5 and 7.

Lastly, we assessed the relation of definition knowledge to children's ability to correctly *rank* time words and event durations. For each task, we performed a linear mixed-effects analysis of children's proportion of correctly ranked items, including age group and definition knowledge as fixed effects. As random effects, we included intercepts for subjects, and by-subject random slopes for the effect of definition knowledge. This model revealed main effects of age, but no effects of definition knowledge on the accuracy of either time word or event ranking, χ^2's < 2.3, p's > 0.1. Thus, knowledge of the formal definitions of time words is not related to children's ability to rank them, consistent with our earlier finding, in Experiment 1, that even 4-year-olds perform above chance on simple duration-word magnitude discrimination.

5. General discussion

The purpose of these studies was to determine how children begin their acquisition of duration words, like *second*, *minute*, and *hour*, prior to being taught their precise definitions in grade school. First, we asked whether children assign interim meanings to these terms during the long delay between when they begin producing these words in speech, at age 2 or 3, and the time that they acquire their adult definitions, at around age 7. Secondly, we asked what these early meanings might look like, and how they might be learned. We focused on two alternatives: (1) the Duration Mapping hypothesis that children begin by individually associating duration words with approximate durations or (2) the Lexical Ordering hypothesis that they initially interpret these words based on their rank ordering within the lexical class of duration words. Consistent with the Lexical Ordering account, we found that initially children were able to contrast and rank duration words, but had little to no knowledge of their absolute durations. Our results indicate that proficiency in estimating the absolute time encoded by duration words emerges relatively late, and may even rely on formal instruction in grade school.

In Experiment 1, children completed a forced-choice task in which they were asked to identify which of two duration words represented the longer duration. For example, children were told that one character performed an action for an hour and another performed the same action for a minute, and were then asked which character did the action "more." We found that by age 4, children chose the correct character more often than expected by chance, and that there were no systematic differences across different words. As children grew older, their performance on these items improved across the board, again without significant differences between items, suggesting that their knowledge of the rank ordering emerges somewhat holistically.

However, as shown in Experiment 2, children's understanding of duration words appears to be initially limited to knowing their rank ordering. In Experiment 2, we used the same paradigm, but now inserted number words to test whether children were sensitive to the absolute durations of time words. Our logic was that if children know that an hour is roughly 50–60 times longer than a minute, then they should also know that 2 hours is more than 3 minutes. On critical trials, we found that 4-year-olds performed at chance and that even 6-year-olds – who have robust understanding of number words and counting – performed more poorly than on contrasts with congruent cues, e.g., 3 hours vs. 2 minutes. These findings indicate that, despite having knowledge of the rank ordering of these terms (e.g., day > hour > minute), children have relatively little understanding of their absolute durations.

Finally, in Experiment 3, we corroborated and extended these results by asking children to estimate durations using a spatial representation of time. We found that by age 5, children were relatively accurate when placing numbers on a number line, and, while they tended to overestimate the durations of familiar events (e.g., of 'washing your hands' and 'eating lunch'), they "spaced out" those estimates similarly to adults, indicating that they do understand their relative durations. However, children performed very poorly when asked to estimate the durations represented by time words like *minute* and *hour*. Not only were the absolute positions of children's estimates on the timeline unlike those of adults, but the relative distances between their estimates were also far from adult-like until the early school years. Furthermore, we observed that the 6- and 7-year-old children who knew the formal definitions of duration words (e.g., one minute equals sixty seconds) were much better able to represent their relative durations on timelines than those who did not.

These results indicate that prior to learning their definitions, preschoolers knew both that duration words indicate lengths of time and that some words indicate longer lengths of time than others. However, although children often understood that an hour is longer than a minute, they generally did not know *how much* longer an hour was. Thus early in development they failed on comparisons like *3 minutes* vs. *2 hours* and were unable to represent the appropriate positions or distances between these terms on timelines. Critically, none of these failures are predicted under the Duration Mapping account, in which duration words are individually mapped onto approximate durations. If such mappings were present, we would expect children to not only rank-order the words correctly, but also to estimate their approximate durations (as they could for number words). Thus, children's interim meanings for these words do not appear to be given by their individual relationships to duration perception. Instead, these meanings appear to be defined by their relations to other time words. Specifically, each word is understood by virtue of its position in a rank ordering of the other duration words known to the child, based on the inference that each of these terms denotes a different duration.

These findings are consistent with those of Shatz et al. (2010), who found that, before children are able to use or comprehend duration words in an adult-like way, they understand that these terms belong to a common lexical category (e.g., they answer questions about duration with duration words). Extending the findings of Shatz et al., we show that the lexical category that children form for duration words is not a simple grouping of these words, but rather a structured, ordered scale that reflects some knowledge of the relative temporal magnitudes of the words (e.g., that an hour is longer than a minute). Nevertheless, this scale is not the same as the scale eventually adopted by adults, which also includes information about the absolute duration represented by each word, and about the proportional relationships between them (e.g., that an hour is sixty times longer than a minute). Instead, it is a more coarse ordinal scale in which words are ranked by increasing duration, with little to no information regarding absolute duration or the proportional relationships among the various terms.

Duration words are only one of many ways we talk about time, but some evidence suggests that children may use similar strategies to learn other types of time words. By the time children learn duration words, they have also learned grammatical tense and several other types of time words, including sequence terms (e.g., *before* and *after*) and deictic terms (e.g., *yesterday* and *tomorrow*) (Ames, 1946; Busby & Suddendorf, 2005; Bloom, 1970; Grant & Suddendorf, 2011; Harner, 1981, 1982; Nelson, 1998; Weist, 1986; Weist & Buczowska, 1987). Though the literature indicates that these expressions are likely acquired earlier than duration words, children's production of deictic and sequence time words also precedes adult-like comprehension. In each case, there is an extended period of inaccurate use in speech and poor performance on laboratory comprehension tasks (e.g., Clark, 1971; Harner, 1975, 1982; Trosborg, 1982; Weist et al., 1991). Furthermore, within-domain errors are also observed with these other time words. For instance, a child might substitute *before* for *after* or *tomorrow* for *yesterday*, or use a term like *yesterday* to refer to the past in general rather than the specific previous day (Bloom, 1970; Harner, 1981). Such specific errors indicate that children understand that these words refer to the domains of sequential and deictic time before they know precisely what each word means. Building on this category knowledge, it is possible that children can also use lexical contrast to learn these other time words.

Prior work has also shown that young children have ordered, list-like lexical representations for larger groups of temporal words, including the days of the week and the months of the year. In both these cases, children seem to learn the list of terms by rote, and can recite them prior to being able to use them to locate events in time or reason about the temporal distances between them (Friedman, 1986, 1989, 1991). Much like we see for duration words, in early development, children seem to know simply that the days of the week and months of the year are sets of contrasting time markers, without knowing how each one relates to the passing of time. For example, a child might know that December comes after November without knowing that December is in the winter, or that it's when Christmas happens, or that it has 31 days. Similarly, a child might know that Tuesday follows Monday without having any idea whether Tuesday is today or tomorrow or yesterday (Friedman, 1977, 1982, 1986).

Of course, calendar words also differ from duration words in several important ways that make it less surprising that children initially learn calendar words as ordered lists. The days of the week and months of the year, for example, have little content beyond their relative ordering, and no apparent link to perception. The common category membership of these terms is conveyed in their lexical form – e.g., all the days of the week contain the suffix-*day*. And, perhaps most critically, children are explicitly taught to recite them in order from an early age. Like the letters of the alphabet (A, B, C...) and the count list (1, 2, 3...), the sequence of days (Monday, Tuesday, Wednesday...) is highly rehearsed, making the ordering of these words more transparent for the child. Indeed, when asked to recite the days of the week in our lab, children as young as 3 often burst into song. All of these factors make the learning problem children face when acquiring this set of time words much less complex that than involved in duration word acquisition. When catchy mnemonic devices are available, it is unlikely that children would adopt any other learning strategy.

In contrast, there are no morphological hints that English words for duration refer to time in their names, nor any indication in their morphology that they go together in any way. Children are much less likely to receive instruction from adults on the rank order of these words, because they are not defined in terms of their order. Thus it is unlikely that 4-year-olds could acquire their rank order via rote memorization. Further, given the large numerosities and complex mathematical principles (e.g., multiplying by 60) involved in their formal definitions, it is also highly unlikely that children as young as 4 actually learn the ordering of duration words by receiving instruction on their definitions in preschool. Instead, we suggest that both the common category and the rank ordering of duration words are inferred indirectly from their use in adult discourse.

Given the evidence that children form categories for abstract words such as duration words prior to learning their formal definitions, one might first ask how children initially determine that duration words refer to lengths of time. In the absence of early "word-to-world" mappings between a lexical item and its perceptual referent, it has been suggested that children rely primarily on cues from their linguistic input when learning "hard" (i.e., abstract) words (Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005). One such cue is the syntax of the sentence. In a process known as "syntactic bootstrapping," children use their knowledge about the grammatical role a novel word plays in a larger sentential context to narrow their hypothesis space when entertaining possible meanings for the term (Gleitman et al., 2005). While it is possible that this type of inference plays a role in duration word acquisition (e.g., for identifying their grammatical status as measure words), we suspect that information from the larger discourse context may play a greater role. As noted by Tare et al. (2008), duration words are used in child-directed speech in a relatively wide variety of contexts, and, although some of these usage contexts do not provide information about their precise durations (e.g., the idiomatic usage of "in a minute"), they nevertheless may help children to establish their common category. In particular, information about the rank ordering of duration words could be gleaned from contrastive usages in speech (e.g., "I thought he'd only be 15 minutes late, but it's been an hour already"). Studies in progress are exploring this possibility via corpus analyses of child-directed speech.

The model of time word learning suggested by these studies is consistent with a broader theory of how children learn the meanings of abstract domains of words. In the cases of number words, color words, and emotion words, for example, children initially identify the conceptual domain to which a set of terms relate, then infer how those terms interrelate, and only later map them onto their extra-linguistic (e.g., perceptual) referents (Davidson et al., 2012; Le Corre & Carey, 2007; Shatz et al., 2010; Wagner, Dobkins, & Barner, 2013; Widen & Russell, 2003; Wynn, 1990). In each of these cases, children quickly identify a set of words that refer to a common conceptual domain, and begin using the words in speech in a non-adult-like way. According to Carey (2009), such case studies are examples of what she calls "Quinian bootstrapping," whereby "an explicit structure is learned initially without the meaning it will eventually have, and at least some relations among the explicit symbols are learned directly in terms of each other" and "the ordering...exhausts their initial representational content" (pp. 329).

In the case of number words, for example, there is evidence that though children have access to approximate nonverbal representations of number from birth (e.g., Izard et al., 2009; Xu & Spelke, 2000; Xu et al., 2005), they do not initially map number words onto these representations (Davidson et al., 2012; Le Corre & Carey, 2007; but see Wagner & Johnson, 2011). Instead, they first memorize a subset of their count list, then gradually learn the individual meanings of a subset of those items (e.g., 1–2–3) over the course of many months, and eventually learn that the count list can be used to identify the cardinality of large sets (see Carey, 2009; Davidson et al., 2012; Le Corre & Carey, 2007; Sarnecka & Lee, 2009; Wynn, 1990, 1992). Only after learning to count do they begin showing systematic evidence of mapping number words onto approximate magnitudes (Davidson et al., 2012; Le Corre & Carey, 2007), a process that likely involves creating a global structure mapping between the count list and the structure of nonverbal number representations (Sullivan & Barner, 2012, 2014b). As with duration words, children appear to learn the structural relationships between number terms, prior to associating these terms with nonverbal representations, despite the fact that those representations are present from birth.

Like duration words and number words, children produce color words and recognize that they belong to a lexical class prior to learning their adult meanings (e.g., Backscheider & Shatz, 1993; Sandhofer & Smith, 1999; Wagner et al., 2013). Also, while early studies argued that children produce color words for many months before learning what the domain of words refers to, more recent work indicates that children know that color words represent color from the time they begin using them. Similar to what we've shown here for time, children begin by acquiring proto-meanings for color words that differ from the meanings understood by adults, and generally are broad overextensions (Wagner et al., 2013). According to Wagner et al., children acquire the precise adult-like meanings of color words by learning how color words relate to each other. Early, overly broad meanings for words like *green* and *red* are narrowed as children add words like *blue*, *pink* and *orange* to their vocabularies. Thus, much like the cases of time and number, color word learning involves not simply mapping words to perception, but also learning about how words within a lexical class relate to one another.

In conclusion, although the mapping between duration words and approximate durations does not appear to be intuitive for children, we find that children nevertheless learn quite a bit about the relationships between these terms in the preschool years, and assign proto-meanings to them based on the inference that they denote different durations. The three experiments presented here support the Lexical Ordering hypothesis, indicating that children learn the lexical category for time words as well as the ordered structure of that category prior to learning their formal definitions. However, our results also suggest that many children do not map these words onto precise representations of duration until after they learn their formal definitions.

Acknowledgments

We thank Eleanor Chestnut, Katherine Kimura, Saima Malik Moraleda, and Karen Bejar for their assistance with subject recruitment and data collection. Tyler Marghetis, Esther Walker, Jessica Sullivan, and members of the Language and Development Lab provided helpful comments. We are grateful to the Montessori Institute of San Diego Children's House, Hawthorne Children's Development Center, Kinderhouse Montessori School, La Jolla United Methodist Church Nursery School, Miramar College Child Development Center, World of Wonders Preschool, Honey Bear Preschool and Child Care Center, Del Mar Montessori, St. James Lutheran, and the Reuben H. Fleet Science Center, where data was collected. Most of all, we thank all the children and families who participated, without whom this research would not be possible. This work was funded by a grant to D.B. from the James S. McDonnell Foundation.

References

Ames, L. B. (1946). The development of the sense of time in the young child. *The Pedagogical Seminary and Journal of Genetic Psychology, 68*(1), 97–125.

Au, T. K. F., & Markman, E. M. (1987). Acquiring word meanings via linguistic contrast. *Cognitive Development, 2*(3), 217–236.

St. Augustine (398/1992). *Confessions* (H. Chadwick, Trans.). Oxford: Oxford University Press.

Backscheider, A. G., & Shatz, M. (1993). Children's acquisition of the lexical domain of color. In *What We Think, What We Mean, and How We Say it: Papers from the Parasession on the Correspondence of Conceptual, Semantic, and Grammatical Representations* (Vol. 29, p. 11).

Barth, H., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*, 125–135.

Barth, H., Starr, A., & Sullivan, J. (2009). Children's mappings of large number words to numerosities. *Cognitive Development, 24*(3), 248–264.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4 classes. R package version 1.0-5*.

Bender, A., & Beller, S. (2014). Mapping spatial frames of reference onto time: A review of theoretical accounts and empirical findings. *Cognition, 132*, 342–382.

Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*(1), 189.

Boroditsky, L. (2011). How languages construct time. In Dehaene & Brannon (Eds.), *Space, time and number in the brain: Searching for the foundations of mathematical thought*. 978-0-12-385948-8. Elsevier.

Brannon, E. M., Suanda, S., & Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Developmental Science, 10*(6), 770–777.

Brown, R. (1973). *A first language: The early stages*. Boston: Harvard U. Press.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.

Busby, J., & Suddendorf, T. (2005). Recalling yesterday and predicting tomorrow. *Cognitive Development, 20*(3), 362–372.

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus*, 59–68.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Casasanto, D., Fotakopoulou, O., & Boroditsky, L. (2010). Space and time in the child's mind: Evidence for a cross-dimensional asymmetry. *Cognitive Science, 34*(3), 387–405.

Catania, A. C. (1970). Reinforcement schedules and psychophysical judgments: A study of some temporal properties of behavior. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules*. New York: Appleton-Century-Crofts.

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.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language, 17*, 417–431.

Clifton, R. K. (1974). Heart rate conditioning in the newborn infant. *Journal of Experimental Child Psychology, 18*(1), 9–21.

Colombo, J., & Richman, W. A. (2002). Infant timekeeping: Attention and temporal estimation in 4-month-olds. *Psychological Science, 13*, 475–479.

Common Core State Standards Initiative (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28*(1), 125–127.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123*(1), 162–173.

De Villiers, J. G., & De Villiers, P. A. (1978). *Language acquisition*. Harvard University Press.

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature, 266*(5604), 718–719.

Droit-Volet, S., Clément, A., & Wearden, J. (2001). Temporal generalization in 3-to 8-year-old children. *Journal of Experimental Child Psychology, 80*(3), 271–288.

Droit-Volet, S., Tourret, S., & Wearden, J. (2004). Perception of the duration of auditory and visual stimuli in children and adults. *The Quarterly Journal of Experimental Psychology, 57A*, 797–818.

Droit-Volet, S., & Wearden, J. H. (2001). Temporal bisection in children. *Journal of Experimental Child Psychology, 80*(2), 142–159.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.

Fitzgerald, H. E., & Brackbill, Y. (1968). Interstimulus interval in classical pupillary conditioning. *Psychological Reports, 23*(2), 369–370.

Friedman, W. J. (1977). The development of children's understanding of cyclic aspects of time. *Child Development*, 1593–1599.

Friedman, W. J. (1982). Conventional time concepts and children's structuring of time. *The developmental psychology of time*, 171–208.

Friedman, W. J. (1986). The development of children's knowledge of temporal structure. *Child Development*, 1386–1400.

Friedman, W. J. (1989). The representation of temporal structure in children, adolescents and adults. *Advances in Psychology, 59*, 259–304.

Friedman, W. J. (1990). Children's representations of the pattern of daily activities. *Child Development, 61*, 1399–1412.

Friedman, W. J. (1991). The development of children's memory for the time of past events. *Child Development, 62*(1), 139–155.

Friedman, W. J., & Laycock, F. (1989). Children's analog and digital clock knowledge. *Child Development*, 357–371.

Fuson, K. C. (1988). Children's counting and concepts of number. Springer-Verlag Publishing.

Fuson, K. C., & Hall, J. W. (1983). The acquisition of early number word meanings: A conceptual analysis and review. In H. P. Ginsburg (Ed.), *The development of mathematical thinking*. New York: AcademicPress.

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

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animaltiming. *Psychological Review, 84*, 279–325.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development, 1*(1), 23–64.

Grant, J. B., & Suddendorf, T. (2011). Production of temporal terms by 3-, 4-, and 5-year-old children. *Early Childhood Research Quarterly, 26*, 87–96.

Harner, L. (1975). Yesterday and tomorrow: Development of early understanding of the terms. *Developmental Psychology, 11*(6), 864.

Harner, L. (1981). Children's understanding of time. *Topics in Language Disorders, 2*(1), 51–66.
Harner, L. (1982). Talking about the past and the future. In W. J. Friedman (Ed.), *The developmental psychology of time*. Academic Press.
Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences, 106*(25), 10382–10385.
James, W. (1890). *The principles of psychology* (Vol. 1). New York: Holt.
Kant, I. (2009/1781). In P. Guyer (Ed.), *Critique of pure reason*. Cambridge University Press.
Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.
Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*(2), 395–438.
Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology, 39*(5), 795.
Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense large-number discrimination in human infants. *Psychological Science, 14*(5), 396–401.
McCormack, T., Brown, G. D. A., Maylor, E. A., Darby, R. J., & Green, D. (1999). Developmental changes in time estimation: Comparing childhood and old age. *Developmental Psychology, 35*, 1143–1155.
McTaggart, J. (1908). The unreality of time. *Mind, 17*, 457–474.
Nelson, K. (1998). *Language in cognitive development: The emergence of the mediated mind*. Cambridge University Press.
Piaget, J. (1969). *The child's conception of time*. Routledge & K. Paul.
Pouthas, V., Droit, S., & Jacquet, A. Y. (1993). Temporal experiences and time knowledge in infancy and early childhood. *Time & Society, 2*(2), 199–218.
Provasi, J., Rattat, A. C., & Droit-Volet, S. (2011). Temporal bisection in 4-month-old infants. *Journal of Experimental Psychology: Animal Behavior Processes, 37*(1), 108.
R Core Team (2013). *R: A language and environment for statistical computing.*
Rivière, V., Darcheville, J. C., & Clément, C. (2000). Rapid timing of transitions in inter-reinforcement interval duration in infants. *Behavioural Processes, 52*(2), 109–115.
Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology, 35*(3), 668.
Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge in early childhood. *Journal of Experimental Child Psychology, 103*, 325–337.
Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6*(3), 357–379.
Shatz, M., Tare, M., Nguyen, S. P., & Young, T. (2010). Acquiring non-object terms: The case for time words. *Journal of Cognition and Development, 11*, 16–36.
Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237–250.
Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General, 142*(1), 193.
Stubbs, A. (1968). The discrimination of stimulus duration by pigeons. *Journal of the Experimental Analysis of Behavior, 11*(3), 223–238.
Sullivan, J., & Barner, D. (2012). How are number words mapped to approximate magnitudes? *Quarterly Journal of Experimental Psychology, 66*, 389–402.
Sullivan, J., & Barner, D. (2014a). The development of structural analogy in number-line estimation. *Journal of Experimental Child Psychology, 128*, 171–189.
Sullivan, J., & Barner, D. (2014b). Inference and association in children's early numerical estimation. *Child development, 85*(4), 1740–1755.
Tare, M., Shatz, M., & Gilbertson, L. (2008). Maternal uses of non-object terms in child directed speech: Color, number, and time. *First Language, 28*, 87–100.
Trehub, S. E., & Thorpe, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 43*(2), 217.
Trosborg, A. (1982). Children's comprehension of 'before' and 'after' reinvestigated. *Journal of Child Language, 9*(02), 381–402.
VanMarle, K., & Wynn, K. (2006). Six-month-old infants use analog magnitudes to represent duration. *Developmental Science, 9*(5), F41–F49.
Wagner, K., Dobkins, K., & Barner, D. (2013). Slow mapping: Color word learning as a gradual inductive process. *Cognition, 127*(3), 307–317.
Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition, 119*(1), 10–22.
Weist, R. M. (1986). Tense and aspect. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 356–374). Cambridge, England: Cambridge University Press.
Weist, R. M., & Buczowska, E. (1987). The emergence of temporal adverbs in child Polish. *First Language, 7*(21), 217–229.
Weist, R. M., Wysocka, H., & Lyytinen, P. (1991). A cross-linguistic perspective on the development of temporal systems. *Journal of Child Language, 18*(01), 67–92.
Whorf, B. L. (1956). The relation of habitual thought and behavior to language. In J. Carroll (Ed.), *Language, thought, and reality*. MIT Press.
Widen, S., & Russell, J. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology, 39*, 114–128.
Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*(2), 155–193.
Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology, 24*(2), 220–251.
Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11.
Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8*(1), 88–101.