Title
The development of structural analogy in number-line estimation

Authors
Sullivan, Jessica
Barner, David

Publication Date
2014-07-01

Peer reviewed

Journal of Experimental Child Psychology 128 (2014) 171–189

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

The development of structural analogy in number-line estimation

Jessica Sullivan [a,*], David Barner [b]

[a] Department of Psychology, Skidmore College, Tisch Learning Center, Saratoga Springs, NY 12866, USA
[b] Department of Psychology, University of California, San Diego, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received 6 September 2013
Revised 20 May 2014

Keywords:
Estimation
Numerical cognition
Analogy
Number-line
Math
Spatial cognition

ABSTRACT

Recent studies have revealed that making number-line estimates requires not only number knowledge but also a host of other cognitive skills. Here, we argue that a fundamental component of number-line estimation is the act of relating the target number being estimated to another numerical reference point (e.g., a previous estimate, the endpoint of the line) and then extending this relation to the spatial domain—in other words, that children recruit analogical reasoning skills when estimating. Because such analogical comparisons require both the selection of a numerical reference point and the comparison of that reference point with the target number, we aimed to understand *which* reference points children use and *how* they use them. To this end, we tested whether and how 5-, 6-, and 7-year-olds used their previous estimates to constrain subsequent estimates. We found that children used their previous estimates as reference points, that older children used reference points differently than younger children, and that the ability to access previous estimates limited our youngest participants' ability to perform well on our number-line estimation task. We conclude that the analogical reasoning component of number-line estimation is substantial and shapes children's earliest estimation performance.

* Corresponding author. Fax: +1 518 580 5319.
E-mail address: jsulliv1@skidmore.edu (J. Sullivan).

http://dx.doi.org/10.1016/j.jecp.2014.07.004
0022-0965/

Introduction

Number-line estimation—that is, the ability to place numbers on a physical line in correspondence to their relative magnitude—is often used to test theories of number knowledge, math skill, and cognitive development more broadly (Booth & Siegler, 2006; Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008; Moeller, Pixner, Kaufmann, & Nuerk, 2009; Siegler & Booth, 2004; Siegler & Opfer, 2003; Siegler & Ramani, 2008). Estimation performance is predictive of math achievement (Booth & Siegler, 2008; Ramani & Siegler, 2008; Schneider, Grabner, & Paetsch, 2009; Siegler & Ramani, 2008) and memory for numbers (Thompson & Siegler, 2010), suggesting that the skills and knowledge measured by number-line estimation tasks are important to general academic success. Although most experiments on number-line estimation focus on how changes in estimation ability relate to the development of number knowledge, some recent studies have shown that number-line estimation depends not only on number knowledge but also on a cluster of additional capacities not directly related to number (Anobile, Stievano, & Burr, 2013; Barth & Paladino, 2011; Kolkman, Hoijtink, Koresbergen, & Leserman, 2013; Slusser, Santiago, & Barth, 2013). As a result, developmental changes in number-line estimation (and their correlation with other important cognitive factors) may stem in large part from changes to non-numerical capacities.

Many studies have found that children's ability to accurately place numbers on a number line improves gradually over development and that this improvement is characterized by a shift toward accurate linear estimation behavior within children's familiar number range (e.g., Booth & Siegler, 2006; Siegler & Opfer, 2003; Slusser et al., 2013). These changes to estimation performance depend, in part, on changes to children's understanding of the verbal number system. As children gain familiarity with a particular numerical range (e.g., 0–100), estimation improves in that range (but not necessarily for unfamiliar numbers, e.g., 101–1000). However, as noted above, improvements to estimation behavior might not indicate improvements to number knowledge alone; recent work has also related the development of number-line estimation ability to constructs as diverse as executive function and spatial proportion reasoning (Barth & Paladino, 2011; Kolkman et al., 2013; Slusser et al., 2013). These studies raise the possibility that improvements to estimation behavior arise not only from improvements to number knowledge but also from improvements to more domain-general cognitive skills. These studies also raise the more general point that estimation tasks might not just measure the accuracy of children's mental number line but may also measure other types of knowledge and skills.

In the current study, we explored the role of one such ability—analogical reasoning—in supporting number-line estimation. How might estimation ability require analogical reasoning skills? To understand the role of analogical reasoning in estimation performance, it is necessary to understand the processes involved in analogy and how these might apply to number-line estimation (for discussions, see Cantlon, Cordes, Libertus, & Brannon, 2009; Carey, 2009; Sullivan & Barner, 2012). Very generally, an analogy involves picking out a relation between entities in one domain and then applying this relation to entities in a separate domain. For example, consider the analogy that "a mitten is to a hand as a sock is to a foot" (or, in standard analogical notation, mitten:hand::sock:foot). This analogy involves two components. The first is a within-domain comparison (e.g., mitten:hand) that is used to extract the relation(s) between the entities in one domain (e.g., that a mitten goes on a hand, that mittens keep hands warm). The second is an across-domain comparison that applies the relation picked out by the within-domain comparison to entities in the new domain (e.g., that a sock goes on a foot, that socks keep feet warm).

Number-line estimation tasks involve precisely this kind of analogy. Because both the length of number lines and the range of values they represent can vary from one line to the next, the only way to construct an estimate is to relate the number being estimated to other numerically meaningful reference points. In other words, on any particular trial, the appropriate location for an estimate can be determined only by first comparing the number being estimated with other numerical reference points on the number line and by then translating the resulting numerical relation into an analogous spatial representation. For example, consider an estimate of 50 that is located 2 inches from the 0 point on a number line. Determining whether this estimate is accurate is impossible without knowing the location of another numerical reference point such as the numerical upper bound of the line (e.g., 100 vs. 1,000,000). So, as in the mitten:hand analogy described above, the ability to make number-line estimates depends on establishing a within-domain relation between items in the number list and extending this relation to spatial points on the number line (see Barth & Paladino, 2011, for a discussion of the spatial component of this analogy).

In this study, we sought to understand how analogical processes contribute to changes in estimation performance. As noted above, when children are asked to estimate a number, they must make a within-domain numerical comparison between the number being estimated and other numerical reference points on the number line. This involves at least two steps. First, children must select a numerical reference point to be used as a point of comparison. Second, they must assess the relation between that reference point and the number being estimated. Only once children have taken these two steps can they complete the analogy by representing the relation between the two numbers spatially on the number line. Thus, to understand the role of analogical reasoning in number-line estimation, it is crucial to understand both (a) which reference points children select when estimating and (b) which numerical relations they pick out between these reference points.

With respect to reference point selection, children have two main options. One option is to use reference points that are provided to them directly by the experimenter. For example, when estimating 50 on a 0–100 number line, children might select 100 as an appropriate reference point because it is the endpoint of the number line. To complete the within-domain numerical comparison, they then use this reference point to calibrate their estimate by picking out some relation (described in detail below) between 100 and 50. Past studies indicate that older children and adults use reference points such as the endpoint and midpoint of the number line to calibrate their estimates (Barth & Paladino, 2011; Siegler & Opfer, 2003; Sullivan, Juhasz, Slattery, & Barth, 2011; for related findings in dot array estimation tasks, see Izard & Dehaene, 2008; Sullivan & Barner, 2012, 2014). However, in some cases it might not be possible to use midpoints and endpoints as reference points because using these experimenter-provided points often requires knowledge of relatively large numbers (e.g., 100) that some young children lack (Barth & Paladino, 2011).

Alternatively, in an experiment with multiple trials, children could make use of their own previous estimates—for example, by recalling their previous response of 10 during a trial in which they are asked to make an estimate for 50 (Sullivan et al., 2011; Vul, Sullivan, & Barner, 2013). This strategy might allow children with limited numerical knowledge to use reference points from their familiar number range. However, this strategy is also constrained, in this case by children's ability to access a previous estimate and use it to guide a numerical comparison. Because number-line experiments typically ask children to make one estimate per number line, children's estimates from previous lines can often be accessed only via memory. If children use their previous estimates to calibrate their subsequent estimates, then the ability to encode, recall, or mentally manipulate the numerical magnitude—or spatial location—of their previous estimate may limit their ability to make accurate estimates. Consequently, if children are unable to use experimenter-provided reference points (e.g., because these numbers are too large) and are unable to access responses from previous trials, then they may fail to make accurate estimates despite having strong internal representations of how numerals encode number. Currently, it is unknown whether young children use their previous estimates to construct new number-line estimates and to what degree their early problems with number-line estimation are related to domain-general cognitive difficulties in accessing previous responses when calibrating estimates. In this study, we explored the question of which reference points children use when making number-line estimates.

A separate question that we explored in this study is which numerical *relations* (between reference points and the number being estimated) children make use of when making estimates. As noted earlier, analogies can encode a number of different relations between entities. For example, in the hand:glove example, a child might focus on the fact that a hand goes inside a glove or, alternatively, that gloves keep hands warm. Similarly, in the case of estimation, individual children may encode different relations between numerals, with consequences for their ability to make accurate adult-like estimates. Here, we focused on two possible relations: ordinality and relative distance.

The first within-domain relation that children might construct is an ordering between the number being estimated and the numerical reference point. For example, a child who relies on ordinality to estimate the location of 50 (using the endpoint of the line 100 as a reference point) should first note that 50 is smaller than 100 and, therefore, place his estimate somewhere to the left of the endpoint. A different child who also recruits knowledge of ordinality but uses her previous estimate (e.g., of 10) as a reference point will place her estimate of 50 somewhere to the right of where she had placed 10 because 50 is larger than 10. Using the ordinal relation between the number being estimated and the selected numerical reference point to calibrate estimates is an unsophisticated strategy; it does not guarantee accurate responding. However, it also does not require knowledge about the relative distances between numbers in the count list, making it a potentially viable strategy for estimating, especially for young children who do not yet have sufficient mastery of the number system to accurately construct more sophisticated relations (discussed below).

The second within-domain relation that children might construct is the relative distance between the target number being estimated and a numerical reference point. For example, a child who has knowledge of the relative distance between numbers might note that 50 comes halfway between the beginning of the count list and 100 and use this information to place his estimate at the midpoint of the line. A different child who uses her previous estimate (e.g., of 10) might note that 50 refers to a quantity that is five times greater than is referred to by 10 and, therefore, place 50 five times farther from the start of the number line than she had placed 10. Although this strategy is more advanced than ordinality alone (because it involves both placing numbers in order *and* representing their relative distances), it also does not guarantee accurate responding, especially if children begin from a reference point that is not accurately situated on the number line (e.g., a previous estimate).

To investigate the role of analogy in number-line estimation, we tested how children select and reason about numerical reference points when estimating. To do so, we conducted an experiment that differed from previous studies in three critical ways. First, we asked whether children attended to the numerical ordering of new estimates relative to their previous estimates. Critically, it is possible to provide inaccurate responses while still providing ordinal responses. However, analyses typically used to assess estimation are not designed to detect ordinal responding in the absence of accuracy. Ours is the first study to test children's use of previous estimates as reference points during number-line estimation tasks and whether children respect the ordering of these estimates (see Sullivan & Barner, 2014, for some evidence of this strategy in array estimation tasks). Specifically, we asked whether individual estimates were ordinal relative to previous estimates (independent of their accuracy). Doing this allowed us to assess the extent to which children—at all ages—made analogical comparisons between the number being estimated and previous estimates.

Second, we asked whether children attended to the relative distance between new estimates and their previous estimates. We tested this in two ways. First, we simulated estimation data for participants who relied only on the ordering of adjacent numbers when generating estimates. We then asked whether the simulated data were similar to our participants' actual estimates (consistent with reliance on trial-to-trial ordinality alone) or whether their estimates were better explained by the use of both ordinality and relative distance information. Next, we introduced a "distribution manipulation" through which some children were asked to make estimates of number drawn from a distribution of relatively smaller numbers, whereas others were asked to make estimates from a distribution of relatively larger numbers. This manipulation was designed to test children's use of relative distance information when relating previous estimates to later ones. As reported by Sullivan and colleagues (2011), participants tend to underestimate large numbers and overestimate small numbers. When overestimation or underestimation happens early in a task, participants who use previous estimates to calibrate later ones will carry this error forward throughout the remainder of the task. However, early differences in estimation should not carry through to later trials if (a) children exclusively use stable reference points on the number line to calibrate their estimates or (b) children use previous estimates but only to represent the ordering of numbers and not the relative distance between them (a prediction corroborated by the above-mentioned simulations). Thus, this manipulation allowed us to test whether children represented the relative distance between early estimates and later ones by measuring whether their estimates—across the entire task—differed across distribution conditions.

Third, we reasoned that, as in other cases of quantity comparison (e.g., Bryant & Trabasso, 1971), if children make use of previous estimates to constrain subsequent ones (by comparing the target number being estimated with previous estimates), then their estimation performance should depend on the their ability to access previous responses. This view of estimation is especially plausible in light of recent research showing that children's ability to remember and update mental representations online is a strong predictor of estimation performance (Kolkman et al., 2013). Typical estimation tasks require children to make estimates one at a time. If children do not use their previous estimates as reference points, then this property of the number-line estimation task should be irrelevant. However, if children *do* use their previous estimates as reference points, then not having access to previous estimates could limit children's performance; they may have difficulty in remembering or accessing the numerical magnitudes that they previously estimated (thereby limiting their ability to make a within-domain numerical comparison), and they may have difficulty in remembering or accessing the location of the previous estimate (thereby limiting their ability to analogically extend their numerical comparison to the spatial domain). To test whether having visual access to previous estimates influences estimation performance, we provided some participants with visual access to their previous estimates and compared their estimation behavior with that of children who received a standard number-line estimation task where they were asked to make estimates one at a time. This methodological choice built on previous studies that have asked children to place multiple estimates on a single line (Siegler & Booth, 2004). By making several changes to the previous paradigm (e.g., testing our participants on a wide range of numbers, not providing a training period, considering measures other than accuracy), we were able to leverage this method to test the role of analogical reasoning in early estimation.[1] Thus, this is the first study to test the role of analogy in children's early estimation ability by considering developmental changes in (a) which reference points children use and (b) which relations children pick out between estimate and reference point.

Method

Participants

A total of 85 children participated. Of this sample, 77 children completed at least 24 trials and were included in the final analyses: 26 5-year-olds, 25 6-year-olds, and 26 7-year-olds (range = 5;0–7;11 [years;months]).

Materials

Stimuli consisted of horizontal black number lines that were 23 cm long. Individual lines were centered on paper measuring 4.25 $\times$ 11 inches. Printed on the left of each number line was the numeral 0, and printed on the right was the numeral 100. The numbers to be estimated were presented auditorily and ranged from 3 to 97. There were no visual cues as to the location of the midpoint of the line (Barth & Paladino, 2011).

Procedure

Participants were shown the number line and were told, "This is a number line. See? It goes from 0 all the way to 100," while the experimenter gestured from left to right across the length of the line. The experimenter continued, "Each number has its own special place on the number line. Today, you're going to show me where certain numbers go on the number line. Look! Zero goes here [gesture to the left-most endpoint] and 100 goes here [gesture to the right-most endpoint]. And all of the other

[1] Siegler and Booth (2004) tested 5-year-olds only with very small numbers (0–10), whereas we tested 1 to 100, which allowed us to test the use of analogy for numbers outside children's counting range. In addition, Siegler and Booth gave children training phase and 66 pretest trials. We were interested in spontaneous estimation strategies and, thus, avoided such training in our study. Finally, unlike their study, we did not restrict analyses to measures of accuracy alone but also included ordinality, which allowed us to find much richer knowledge in 5-year-olds than previously reported.

numbers have their own special places on the number line. I'm going to give you a pencil, and your job will be to draw an up-and-down line to show me where each number goes. Are you ready?" Participants were then given 24 estimation trials.[2] On each trial, the number to be estimated was provided, and children were given a new differently colored pencil to mark each answer (to differentiate estimates when they were marked on the same sheet). The task took approximately 20 min to complete.

To test whether the accessibility of previous estimates influenced children's estimates, participants were randomly assigned to one of two conditions: a Single Estimate condition or a Multiple Estimate condition. In the Single Estimate condition, as with previous studies of children's number-line estimation, participants made estimates for numbers one at a time, marking each estimate on a new number line (see Booth & Siegler, 2006; Siegler & Opfer, 2003). In the Multiple Estimate condition, participants made estimates one at a time but provided multiple estimates on the same number line (e.g., Siegler & Booth, 2004). To avoid cluttering the number line, estimates for the first 12 trials were recorded on one number line and estimates for the last 12 trials were recorded on a separate number line (only one number line was visible at a time). Twelve children (14% of our test population) asked to be reminded of a previous estimate (e.g., "What number was the pink line?," "Can I see the other sheet?"); their requests were honored in the Multiple Estimate condition ($n = 5$) but were denied in the Single Estimate condition ($n = 7$).

Participants estimated numbers drawn from one of two possible distributions. In the Small Number Distribution, 24 numbers were selected between 1 and 100 such that 4 were smaller than 10: 3, 4, 6, 8, 12, 14, 17, 18, 21, 24, 25, 29, 33, 39, 42, 48, 52, 64, 72, 79, 81, 84, 90, and 96 (Booth & Siegler, 2006). The Large Number Distribution contained the 24 numbers generated by subtracting the Small Number set from 100 (Barth & Paladino, 2011; Sullivan et al., 2011). There were two possible trial orders within each distribution. These were pseudo-random permutations of the selected numbers arranged so that the first few trials contained a small number (for those in the Small Number Distribution set) or a large number (for those in the Large Number Distribution set).

Children's estimation behavior was also qualitatively coded online for evidence of reference point use, counting, and other strategies. Preliminary analyses indicated that these measures were not related to any of the data reported here; thus, they are not included.

Simulations

We conducted several simulations of estimation data in order to formalize our predictions about the empirical effects of (a) using previous estimates to calibrate later ones and (b) our distribution manipulation. Simulations were conducted using custom MATLAB code or in Excel (both available on request). Each simulated experiment contained 24 participants randomly assigned to one of our two distribution conditions (and also to one of our two trial orders). Each simulation was iterated 1000 times.

Trial-to-trial ordinality

This simulation randomly generated an estimate between 0 and 100 for the first trial[3] and then followed trial-to-trial ordinality x% of the time (where x was the percentage of ordinal responses at each age group in our real data; see Results). For example, if on Trial 1 an estimate was randomly selected as 25, then on Trial 2 the response for a large number would fall between 25.01 and 100, whereas the response for a smaller number would fall between 0 and 24.99, thereby respecting ordinality in each case. We ran 1000 iterations of this simulation for each age group's ordinality rate (5-, 6-, or 7-year-olds).

[2] Approximately 60 of the participants were given the opportunity to complete a second set of 24 trials in the opposite condition. Due to significantly higher rates of error across all age groups and numerous experimenter notes of inattention during the second 24 trials, data for these trials were not analyzed further.

[3] Note that if the number requested on the first trial is relatively large, this means that the model is probabilistically likely to underestimate that first number. Also note that after the first trial, an estimate of a large number is probabilistically likely to be followed by a request for a smaller number. In this way, estimating from a large number distribution can lead to systematic underestimation if participants use previous estimates to calibrate later ones.

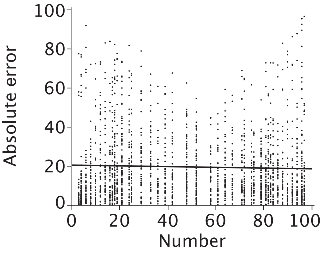

Fig. 1. Absolute value of amount of error, plotted by the number being estimated. Points are individual data points; line is best-fit line.

Trial-to-trial relative distance

This simulation generated a first estimate by taking the target number being estimated (e.g., 18) and randomly selecting an estimate that was ±[0–40%][4] of the target number (e.g., 18 ± [0–7.2]). We selected the range of 0 to 40% error because that is most likely to converge at a percentage absolute error in estimation (PAE) of 20%, which is what we found in our data (see Fig. 1).

To generate the second estimate, this simulation calculated the ratio of the target number on Trial 2 to the target number on Trial 1 (e.g., if the target number on Trial 2 was 36, then the ratio would be 2:1) and then multiplied the first estimate by that ratio. In this way, the relative distance between estimates was preserved, regardless of the accuracy of the initial estimate, and each subsequent estimate was contingent on the immediately preceding estimate.

Independently generated estimates

We ran two simulations in which estimates were generated independently from one another (e.g., without referencing previous estimates). For the first simulation, on each trial the estimate was a randomly selected number that was ±[0–40%] of the target number.

The second simulation took into account participants' bias to overestimate small numbers and underestimate large ones (Sullivan et al., 2011). If the number being estimated was larger than 50, then the program generated an estimate that was 0 to 40% below the number being estimated; if it was smaller than 50, then the program generated an estimate that was 0 to 40% above the number being estimated.

Analyses

Children's responses were converted to their numerical estimate equivalents. Indecipherable responses were excluded ($n = 9/1848$ trials). Responses that were located immediately to the right of the number line's endpoint were included in the final analyses ($n = 28/1848$ trials) because these were frequently accompanied by children's explanations (e.g., "This one has to be off the list"). These responses resulted in some estimates that were larger than 100 (see also Cohen & Blanc-Goldhammer, 2011, for a discussion of how the bounds of a number line can constrain estimates in undesirable ways and why the assessment of numerical knowledge can be facilitated by using unbounded number-line tasks). Analyses excluding these 28 trials were also conducted, with identical effects to those reported below.

Our analyses focused on two measures of estimation performance. First, we used regressions to analyze the relation between each estimate and the number being estimated (e.g., Barth, Starr, & Sullivan, 2009; Booth & Siegler, 2006; Lipton & Spelke, 2005; Siegler & Opfer, 2003). This type of analysis allows us to test the relation between the numbers being estimated and children's estimates and

[4] We also simulated other amounts of error (e.g., 0–10% error) for this and all other simulations in which error rates were specified in the simulation. Results were quantitatively and qualitatively similar to those reported here. When testing for an interaction between distribution and magnitude (see below), we found that as error decreases, the use of trial-to-trial relative distance information is even *more* likely to yield an Interaction, whereas all models are *less* likely to do so.

is useful because we can construct more complicated models that predict children's estimates from the target number being estimated *and* other factors (e.g., visual access condition).

We also measured whether children's estimates respected the ordinality of the count list. A trial was labeled as ordinal if children provided an estimate in the correct direction relative to a previous estimate regardless of its accuracy (e.g., by providing a larger estimate on trial n than on trial $n-1$ if and only if a larger number was requested on trial n than on trial $n-1$). Our main ordinality test asked whether any given estimate was ordinal relative to the immediately preceding estimate (thereby comparing trial n with trial $n-1$).

With the exception of binomial comparisons with chance and chi-square analyses, all analyses reported below were conducted using the LME4 package of R (Bates & Sarkar, 2007; R Development Core Team, 2010). All models were linear mixed models with participant considered a random factor. Condition (Single Estimate vs. Multiple Estimate) and distribution (Small Number Distribution vs. Large Number Distribution) were considered fixed factors. Ordinality scores resulted in binomial data and, therefore, were subjected to binomial logit analyses. Simulated data and real data were analyzed using the same code. We report parameter estimates (β), MCMC (Markov chain Monte Carlo) p value estimates, and standard error estimates. All results reported below are of real (child-generated) data unless noted otherwise.

Results

Estimation performance and replication of past data

We predicted participants' estimates from a model containing age and the magnitude of the number being estimated. Consistent with previous research, there was an effect of age such that older children provided different estimates than younger children ($\beta = 11.5$, $SE = 1.6$, $p < .0001$), an effect of magnitude indicating that estimates were related to the number being estimated ($\beta = .66$, $SE = .11$, $p < .0001$), and an interaction between magnitude and age indicating age-related differences in the specific relation between the target number and children's estimate ($\beta = -.20$, $SE = .02$, $p < .0001$).

Next, we predicted participants' estimates from the numerical magnitude they were estimating, separating participants by age. Here, β can be interpreted as a simple slope measure where the closer β is to 1, the closer estimates are to what we called "adult-like" performance (although even children with a slope of 1 may differ in important ways from adults). Predictably, 5-year-olds performed the worst, although they still made estimates that were linearly related to the target magnitude ($\beta = .36$, $SE = .03$, $p < .0001$). Among the older children, 6-year-olds' estimates had a slope closer to 1, indicating more adult-like performance ($\beta = .57$, $SE = .02$, $p < .0001$), and 7-year-olds performed extremely well ($\beta = .74$, $SE = .02$, $p < .0001$). Thus, younger children's estimates were somewhat inaccurate and did not display an adult-like linear relation between number and estimates.[5]

Finally, we asked whether PAE increased with the magnitude of the target number, where PAE was defined by the absolute difference between the number being estimated and the child's estimate. In general, when estimation tasks draw on the approximate number system (ANS), estimation error increases with number (Dehaene, 1997; Whalen, Gallistel, & Gelman, 1999). However, past studies have found that number-line estimates do not show this signature of the ANS (e.g., Siegler & Opfer, 2003) and, thus, that estimates depending on analogical numerical comparisons are unlikely to be linked to the ANS. We tested whether the PAE increased with number (indicating that estimates were supported by the ANS). Consistent with past reports, we found that it did not ($\beta = -.007$, $SE = .013$, $p > .50$; see Fig. 1), suggesting that the ANS did not support participants' number-line estimates.

[5] To ensure that our data replicated recent reports of children's estimation, we also assessed the degree to which proportional reasoning models of estimation could account for deviations from linear performance found in our data. We found that, as reported previously (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Sullivan et al., 2011), deviations from linearity could be partially accounted for by the error introduced in proportional judgments. We also tested logarithmic fits and found that our youngest participants tended to be fit well by these curves. Because our interest was *not* in the particular shape of the estimation function involved but rather in children's use of analogy and working memory, we do not discuss these models further.

Use of previous estimates and reliance on ordinality relations

Critical to the current proposal—that developmental changes to estimation behavior may be linked to changes in the types of reference points used during estimation tasks—we tested (a) whether children compared their earlier estimates with later estimates and (b) how children used those reference points. To this end, we analyzed children's ordinality scores, which represented whether children preserved numerical ordering relations across adjacent trials. As described above, for each child, we computed whether a particular trial's estimate was ordinal relative to the immediately preceding estimate. So, if a larger number was requested on trial n than on trial $n - 1$, a child's response was counted as ordinal if and only if it was larger for trial n than for trial $n - 1$. Because it is impossible to compute ordinality for the first trial of the task (because there is no preceding trial), this trial was not scored.

Our logic was that children who provide inaccurate estimates could be doing so for one of two reasons. First, they could be performing randomly (e.g., out of lack of numerical knowledge or confusion). Alternatively, poor estimators may lack knowledge of the relative distance between numbers but may still represent the ordering of the target number relative to their previous estimate. If children represent the relative ordering of number words and use this to guide their number-line mappings, then they should demonstrate above-chance levels of ordinality.

First, we asked whether children provided ordinal responses. In a binomial logit regression, we predicted the ordinality of estimates from age and found that the likelihood of ordinal responses increased significantly as a function of age ($\beta = 1.02$, $SE = 0.16$, $p < .0001$). Although young children's estimates were less likely to be ordinal than older children's estimates, all participants demonstrated high levels of ordinality, with 5-year-olds providing ordinal responses on 69.6% of all trials, 6-year-olds doing so on 84.1% of trials, and 7-year-olds doing so on 93.3% of trials (see Fig. 2). There was a significantly greater proportion of ordinal trials within each age group than would be expected by chance alone (chance = .50, all binomial ps < .0001), and this was true for each condition separately (Single Estimate and Multiple Estimate, all binomial ps < .01). Thus, although the youngest children in our study were relatively inaccurate estimators, they nonetheless produced highly ordinal responses.

Our analysis of ordinality revealed that all age groups provided highly ordinal estimates. Importantly, this high level of ordinality did not simply reflect high levels of accuracy. Although there was a strong relation between estimation error and ordinality in our dataset, only approximately half of the variability in ordinality was accounted for by PAE (R^2s = .51 for 5-year-olds, .68 for 6-year-olds, and .38 for 7-year-olds; see Fig. 3). The same pattern—that typical measures of estimation accuracy account for only part of the variability in ordinality—was also found for other measures of estimation proficiency such as estimation slope, linear R^2 (the R^2 of the linear fit between the number being estimated and a participant's estimate), and log R^2 (the R^2 of the logarithmic best fit between the number being estimated and a participant's estimate). The accuracy measure that accounted for the most variability in ordinality scores was R^2 for the participant's linear fit (R^2s = .70 for 5-year-olds, .78 for 6-year-olds, and .60 for 7-year-olds). Other accuracy measures were less predictive of ordinality

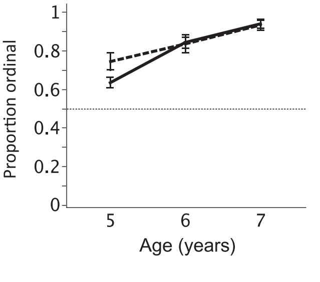

Fig. 2. Proportion of ordinal responses for 5-, 6-, and 7-year-olds. The dashed line indicates performance in the Multiple Estimate condition, the solid line indicates performance in the Single Estimate condition, and the dotted line indicates chance performance. Error bars are standard errors.

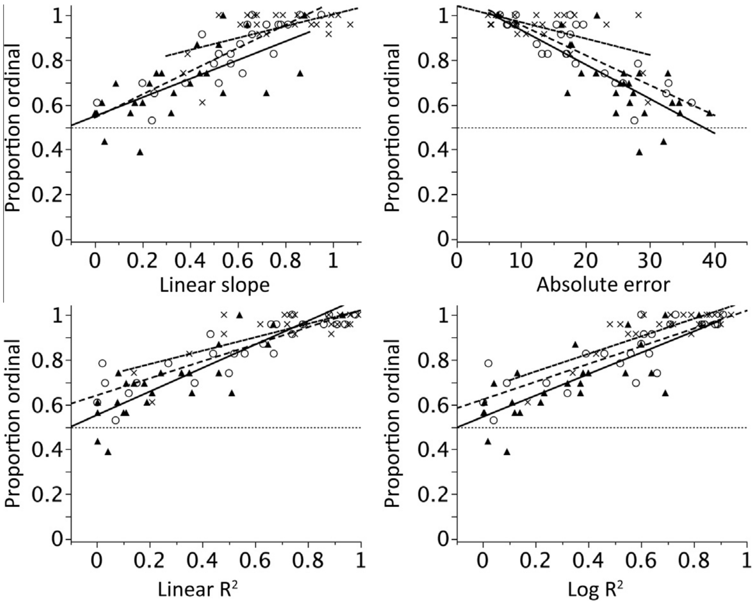

Fig. 3. Proportion of ordinal responses plotted by four different estimation measures. Each point is a child; triangles are 5-year-olds, circles are 6-year-olds, and crosses are 7-year-olds. The solid regression line indicates best fit for 5-year-olds, the dashed line indicates best fit for 6-year-olds, and the patterned line indicates best fit for 7-year-olds. The dotted line at 0.5 is chance ordinality performance.

(absolute value of the distance of linear slope from 1: R^2s = .46 for 5-year-olds, .73 for 6-year-olds, and .32 for 7-year-olds; log R^2: R^2s = .59 for 5-year-olds, .69 for 6-year-olds, and .68 for 7-year-olds). Thus, probably because ordinality does not require accuracy, it picks out variability that is not captured by existing measures of estimation ability.

The dissociation between estimation accuracy and ordinality is most evident when examining the performance of individual participants. For example, the seven participants with the highest estimation error (five 5-year-olds and two 6-year-olds) all had above-chance ordinality scores, and some participants with perfect ordinality scores nonetheless showed substantial estimation error (of the 15 participants with perfect ordinality scores, their average PAEs ranged from just under five to just over 28, with a group mean of 12.37). Among the 44 participants who had very high levels of estimation error (PAEs between 15 and 30), 14 had ordinality scores greater than 95%, whereas only four of these participants showed ordinality performance below or near chance. Thus, although estimation error and ordinality were related, they are two fundamentally distinct phenomena (see Fig. 3).

Although ordinality was generally very high, most children made at least some estimates that were non-ordinal. We next asked whether these non-ordinal responses were randomly distributed across the number line (e.g., due to poor memory for previous responses or other random errors) or were instead restricted to particular numbers (e.g., due to relative unfamiliarity with large and less familiar numbers). To test this, we analyzed the relation between the numerical magnitude being requested on a particular trail and the likelihood that a participant provided an ordinal response. There was no effect of numerical magnitude for any of our age groups (5-year-olds: $\beta = -.005$, $SE = .003$, $p = .10$; 6-year-olds: $\beta = -.004$, $SE = .004$, $p = .32$; 7-year-olds: $\beta = -.003$, $SE = .006$, $p = .96$), suggesting that children were no less likely to provide ordinal estimates for large numbers than for small numbers (see Fig. 4).

Together, these analyses reveal four properties of the development of reference point selection and use during number-line estimation tasks. First, the fact that ordinality was above chance for all age groups and in all conditions and independent of accuracy suggests a reliance on previous trials to

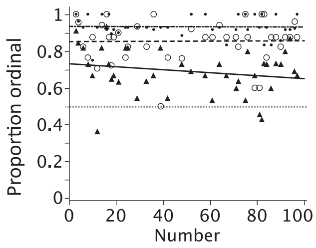

Fig. 4. Proportion of ordinal responses, plotted by the number being estimated. Each point is the group mean for that number; triangles are 5-year-olds, circles are 6-year-olds, and dots are 7-year-olds. The solid regression line indicates best fit for 5-year-olds, the dashed line indicates best fit for 6-year-olds, and the patterned line indicates best fit for 7-year-olds. The dotted line at 0.5 is chance ordinality performance.

calibrate subsequent estimates. Second, these analyses show that children make ordinal comparisons even at the earliest ages tested and even in cases where their estimates were otherwise inaccurate. Third, older children's estimates were more likely to be ordinal than younger children's estimates. Fourth, ordinality was unrelated to the numerical magnitude of the number being estimated, suggesting that limitations to children's ordinal responses do not stem from lack of knowledge of a particular numerical range but instead stem from other factors (e.g., the accessibility of previous estimates, discussed below) that affected all trials equally.

Although we have shown that children attend to the ordinality of numbers when estimating, one might wonder whether children who relied only on trial-to-trial ordinality could nonetheless provide estimates that were linearly related to the target number being estimated (as our participants did). This might be of special importance for 5-year-olds; can participants who provide ordinal estimates only 69.9% of the time actually provide estimates that are linearly related to the target number being estimated? To address this question, we turn now to our simulated estimation data. As described

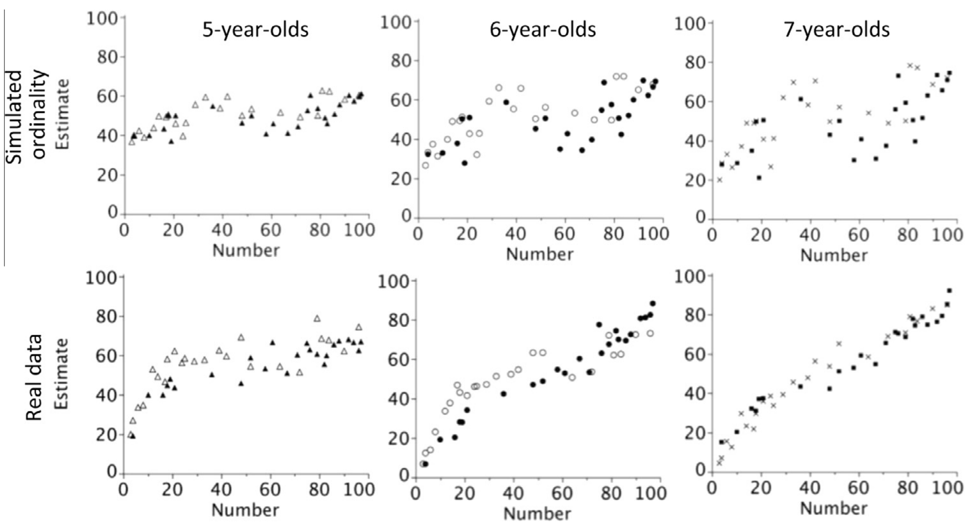

Fig. 5. Estimation performance for our simulation of ordinality-based responses (top row) and real data (bottom row). Simulated data represent estimates generated by providing ordinal estimates at the rates typical of each age group; data in the top row represent the mean of all 1000 simulated experiments in a given age range, and data in the bottom row represent the means for all real participants in a given age range. Triangles (left column) are 5-year-olds, circles (middle column) are 6-year-olds, and crosses/squares (right column) are 7-year-olds. Filled markers indicate Large Number Distribution data, and unfilled markers indicate Small Number Distribution data.

above, we ran simulations of estimation data generated via trial-to-trial ordinality. Our simulation of 5-year-olds generated its first estimate randomly and then generated ordinal estimates 69.9% of the time. Did these simulated 5-year-olds provide estimates that were linearly related to the target number being estimated? They did in the majority of simulations (see Fig. 6 below). This suggests that trial-to-trial ordinality alone is sufficient to support the generation of estimates that are linearly related to the target magnitude and that this is true even in populations with relatively low rates of ordinality. In fact, measures of estimation variability (computed by taking the mean estimate for each number averaged across all iterations of our simulations) appear relatively similar in our real 5-year-old data and our ordinality-based simulated 5-year-old data (mean simulated linear $R^2 = .40$, real linear $R^2 = .57$), although even among 5-year-olds the ordinality-only simulation does not appear to account for estimates of small numbers (under ∼10; Fig. 5). Simulations of our 6- and 7-year-olds' ordinality rates led to a linear relation between target magnitude and estimate 99.8% and 100% of the time, respectively. However, in these cases the simulated estimates were substantially more variable than the responses we found in our data.

Use of relative distance information

To provide ordinal estimates, children only need to represent the ordering of the numbers and the location of previous estimates. However, when adults make number-line estimates, they express both ordinal and relative distance relations between previous trials and later trials (Sullivan et al., 2011; Vul et al., 2013). And, as shown in Fig. 5, although ordinality explains 5-year-olds' data relatively well, simulations based on ordinality alone are more variable and less accurate than our real 6- and 7-year-olds' data. This raises the possibility that an important developmental change in estimation ability is the shift from reliance on primarily ordinality-based analogical comparisons to reliance on relative distance comparisons.

To investigate this, we asked whether participants were sensitive to our distribution manipulation. The basic logic of the distribution manipulation is as follows: If participants use their previous estimates to constrain subsequent estimates, then differences in estimates that occur early in an estimation task should carry through to later trials, as observed in previous studies of adults (Sullivan et al., 2011). For example, consider two participants. The first is in the Large Number Distribution condition

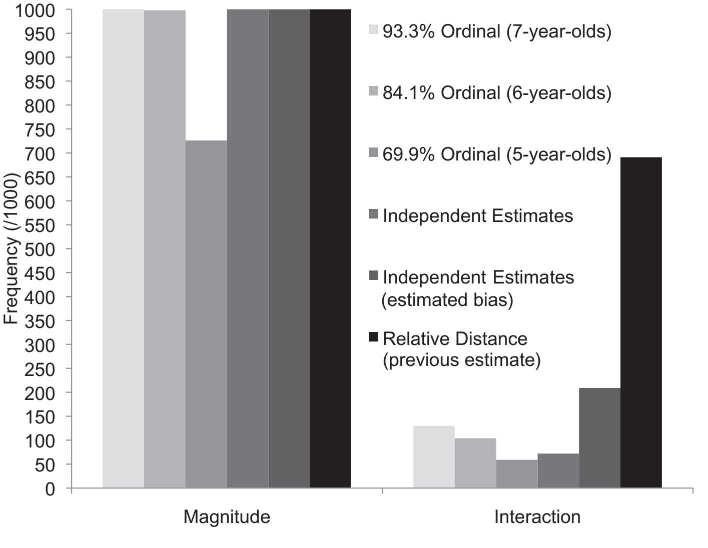

Fig. 6. Simulated estimation data. The *y* axis represents the number of iterations for which there was a significant ($|t| > 1.96$) effect of magnitude or an interaction between magnitude and distribution on simulated estimates.

and is asked to estimate 96 relatively early in the estimation task (and is likely to underestimate it, perhaps as 90), and the other is in the Small Number Distribution condition and is asked to estimate 4 (and is likely to overestimate it, perhaps as 6). On the next trial, both participants are asked to estimate 48. If both participants rely on their previous estimates and do so using relative distance information, then the participant in the Large Number Distribution might place his estimate of 48 at 45 (because 96:48::90:45), whereas the participant in the Small Number Distribution might place her estimate of 48 at 72 (because 4:48::6:72). In this way, small errors that are likely to emerge based on the magnitude of the number first estimated should carry through to an entire estimation task, but only if previous estimates constrain later ones and only if these estimates take account of relative distance information (as in the example above and as we demonstrate below using simulated data).

Critically, the largest distribution effects should be found at the midpoint of the line. This is for two separate but related reasons. First, estimates that are near the endpoints of the line are more constrained (e.g., I can underestimate 50 by 10, but I cannot underestimate 7 by 10). Thus, when making estimates near the endpoints of the number line, it will not always be possible to carry error through from a previous trial due to the constraints imposed by having fixed endpoints on the number line. As a result, we should expect smaller effects of distribution near the endpoints (see Sullivan et al., 2011, for evidence of this in adults). Second, an effect of distribution should be most likely if participants use their previous estimates to constrain later ones and if they do so by representing relative distance information (see Fig. 6).[6] But participants are *less* likely to use previous estimates to constrain estimates that are near experimenter-provided reference points (in this age range, the relevant reference points are the two endpoints [see Barth & Paladino, 2011]; adults' reference points are slightly different [see Sullivan et al., 2011]). These two observations suggest that insofar as children use their previous estimates to constrain later ones via relative distance comparisons, we should find an interaction between distribution and magnitude such that the effect of distribution is largest near the midpoint of the line.

We first conducted an analysis predicting children's estimates from magnitude, age, distribution, and all of their interactions. This analysis found an effect of magnitude ($\beta = -.83$, $SE = .17$, $p < .0001$), an effect of age ($\beta = -14.02$, $SE = 2.21$, $p < .0001$), an interaction between magnitude and age ($\beta = .23$, $SE = .05$, $p < .0001$), a marginal effect of distribution ($\beta = -34.41$, $SE = 19.80$, $p = .08$), and no other effects. For this analysis, the absence of a significant interaction between distribution and magnitude does not indicate that there was no effect of distribution; the effects of our distribution manipulation were expected to be small (Sullivan et al., 2011) and may have been overwhelmed by the large differences in estimation performance we expect across age groups. Furthermore, the distribution effect is unlikely to appear if some age groups relied heavily on ordinality. Because of this, we also conducted planned analyses separately by age group.

In these analyses, we predicted estimate from magnitude, distribution, and their interaction. Predictably, all age groups showed an effect of magnitude, demonstrating that their estimates were linearly related to the number they were estimating even when the distribution condition was taken into account (5-year-olds: $\beta = .34$, $SE = .05$, $p < .0001$; 6-year-olds: $\beta = .50$, $SE = .03$, $p < .0001$; 7-year-olds: $\beta = .79$, $SE = .03$, $p < .0001$).

However, not all groups were affected by distribution information. For 5-year-olds, there was no significant effect of distribution ($\beta = -6.53$, $SE = 4.68$, $p > .15$) and no interaction between distribution and magnitude ($\beta = .002$, $SE = .06$, $p > .95$). For 6-year-olds, there was an effect of distribution ($\beta = -16.25$, $SE = 4.98$, $p < .01$) and an interaction between distribution and magnitude ($\beta = .23$, $SE = .06$, $p < .0001$), suggesting that their estimates differed as a function of both the numerical magnitude being estimated and the particular distribution they were estimating. For 7-year-olds, there

[6] It is possible—but substantially less likely—for effects of distribution to emerge if a participant uses previous estimates to calibrate later ones but does so via ordinality alone. This is because if the Large Number Distribution participant described above estimates 96 as 90, then his estimate of 48 cannot be larger than 90 (the expected value, as calculated by averaging all possible estimates smaller than 90, is 45). In contrast, the participant in the Small Number Distribution who estimated four as six would need to provide an estimate larger than six to preserve ordinality, and the expected value for this estimate would be 53.5. In this way, even participants who rely on ordinality alone could be influenced by distribution; in our simulation based on 7-year-olds' ordinality rates, an interaction between distribution and magnitude emerged in 13% of experiments, whereas in our simulation of 5-year-olds' data, only 5.9% showed an interaction.

also was an interaction between distribution and magnitude ($\beta = -.11$, $SE = .04$, $p < .025$), but there was no main effect of distribution ($\beta = 3.99$, $SE = 4.80$, $p > .25$).

Although it is clear that 6- and 7-year-olds showed an interaction between magnitude and condition, one might wonder whether the best explanation of these data is really that older children use trial-to-trial relative distance information to calibrate their estimates. To further probe the source of the interaction between magnitude and condition, we turn to our simulated data. When we simulated 1000 experiments in which participants were randomly assigned to one of our distribution conditions and used their previous estimates to calibrate later ones and did so by attending to relative distance, an interaction between distribution and magnitude emerged 69.1% of the time (Fig. 6). No other simulation yielded such a high rate of interaction. Although it is theoretically possible to find an interaction in data generated via any of the estimation strategies that we simulated, based on the fact that we found significant interactions in *both* groups of older children, and based on the data provided by our simulation, the most likely explanation seems to be that the older children, but not 5-year-olds, used their previous estimates to calibrate later ones and did so based on the relative distance between estimates.

Availability of previous estimates

We have shown that children make use of their previous estimates when estimating the location of a target number by making a comparison between previous estimates and the target number. To do this, they must be able to access the numerical magnitude and location of previous estimates. However, in typical estimation tasks, estimates are produced one at a time and the previous estimates are not visually available to participants. This raises the possibility that children's ability to produce accurate estimates may be limited by their ability to access information from previous estimates.

To test this, we constructed a model predicting participants' estimates from the numerical magnitude being estimated, age, and visual access condition (Single Estimate vs. Multiple Estimate). This model asks whether (a) individual estimates are predicted by the target number, (b) this relation is modulated by having visual access to previous estimates, and (c) this effect was mediated by age. Critically, an interaction between condition and magnitude (meaning that participants gave smaller estimates for smaller numbers and larger estimates for larger numbers in one of the two conditions) would provide evidence that having visual access to previous estimates improves estimation performance. Predictably, we found an effect of magnitude ($\beta = -.43$, $SE = .14$, $p < .01$), an effect of age ($\beta = -9.23$, $SE = 2.16$, $p < .0001$), and an interaction between age and magnitude ($\beta = .17$, $SE = .02$, $p < .0001$). Again, this shows that participants provided estimates that were related to the target magnitude and that older children made more accurate estimates than younger children. Critical to our main question, we also found an interaction between condition and magnitude ($\beta = -.56$, $SE = .22$, $p < .05$), suggesting that participants in the visual access condition provided more accurate estimates than participants in the no access condition. We also found a three-way interaction among condition, age, and magnitude ($\beta = .09$, $SE = .04$, $p < .05$), suggesting that the benefit of having visual access to previous estimates was restricted to children of a certain age.

To better understand these effects, we next analyzed each age group separately. Again, across all age ranges, magnitude remained a significant predictor of estimation performance (5-year-olds: $\beta = .41$, $SE = .04$, $p < .0001$; 6-year-olds: $\beta = .55$, $SE = .04$, $p < .0001$; 7-year-olds: $\beta = .75$, $SE = .03$, $p < .0001$). Although older children performed better than young children, at each age group participants' estimates were related to the target number being estimated.

Critically, only 5-year-olds showed an interaction between condition and magnitude ($\beta = -.18$, $SE = .06$, $p < .01$); these youngest participants were more likely to provide smaller estimates for smaller numbers and larger estimates for larger numbers in the Multiple Estimate condition (where it was easy to access previous estimates) relative to the Single Estimate condition (where it was hard to access previous estimates). This resulted in more accurate performance in the Multiple Estimate condition than in the Single Estimate condition (slope relating estimates to the target number for the Single Estimate condition: $\beta = .21$; Multiple Estimate condition: $\beta = .38$; here, β can be interpreted as a simple slope measure with adult-like performance as $\beta = 1$). We found a similar benefit in 5-year-olds for other measures of estimation accuracy (PAE: Single Estimate condition = 28.4; Multiple Estimate

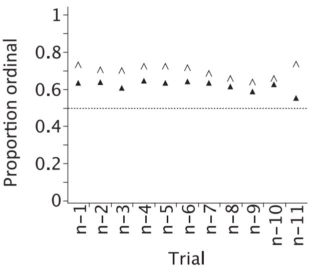

Fig. 7. Proportion of 5-year-olds' ordinal responses plotted by trial, where $n-1$ represents the proportion of responses on trial n that are ordinal relative to trial $n-1$ and where $n-7$ represents the proportion of responses on trial n that are ordinal relative to trial $n-7$. Filled triangles are data from the Single Estimate condition, and carats are data from the Multiple Estimate condition.

condition = 23.58; linear R^2: Single Estimate condition = .16; Multiple Estimate condition = .34; log R^2: Single Estimate condition = .19; Multiple Estimate condition = .39).

Our 6-year-olds showed a main effect of condition ($\beta = -8.50$, $SE = 4.30$, $p < .05$) but no interaction ($\beta = .05$, $SE = .05$, $p > .25$). This suggests that participants in the Multiple Estimate condition made, on average, larger estimates than those in the Single Estimate condition, but without an interaction we cannot conclude that having visual access to previous estimates improved 6-year-olds' estimates. Finally, 7-year-olds showed neither an effect of condition ($\beta = -.48$, $SE = 4.81$, $p > .90$) nor an interaction ($\beta = -.008$, $SE = .04$, $p > .80$). This suggests that these older children, whose estimates tended to be quite accurate and linear even in a standard number-line estimation task (e.g., Booth & Siegler, 2006), do not show improved performance from having visual access to previous estimates, likely because their performance was already quite strong.

Because our 5-year-olds were the least likely to provide accurate estimates but still seemed to rely heavily on previous estimates when estimating (as evidenced by the effect of the visual access condition and by their above-chance rates of ordinal responding), we asked one final question of our data. We asked whether these children provided estimates that were in the correct direction (ordinal) not only relative to the immediately preceding estimate ($n-1$) but also relative to *all* previous estimates ($n-2$, $n-3$, . . ., $n-11$). As in our initial ordinality analysis, we coded each trial as 1 if it was in the correct direction relative to the previous trial $n-x$ and as 0 if it was in the incorrect direction. We did this independently for each of the 11 preceding trials (so that, e.g., a failure to provide an ordinal response on trial $n-8$ relative to trial n did not preclude a successful ordinal response on trial $n-9$ relative to trial n). Importantly, 5-year-olds' estimates were more ordinal relative to each of the 11 previous trials tested than would be expected by chance (50%) alone (all binomial ps < .001). For each of the 11 previous trials, responses in the Multiple Estimate condition were more likely to be ordinal than those in the Single Estimate condition (see Fig. 7), although this pattern did not reach significance for all comparisons, likely because power decreased as the distance from n increased.[7] Together, the effects of condition on accuracy and ordinality suggest that even our youngest participants used their knowledge of ordinality to relate the target number being estimated to all previous trials and that they were best able to do this when they had visual access to their previous estimates.

Discussion

In this study, we investigated whether the analogical reasoning demands of estimation tasks influence children's number-line estimation performance. We reasoned that if number line estimation is

[7] Chi-square tests found significant effects of Single Estimate versus Multiple Estimate on ordinality at $p = .093$ for $n-2$, $p = .025$ for $n-3$, $p = .058$ for $n-4$, $p = .032$ for $n-5$, $p = .092$ for $n-6$, $p = .26$ for $n-7$, and $p = .29$ for $n-8$. However, computations of ordinality are subject to increasing data loss as the trial distances become larger (e.g., to compute ordinality for $n-11$, comparisons can be made only for data points after Trial 11, whereas $n-1$ comparisons can make use of all data except the first trial). Thus, nonsignificant tests involving larger intertrial distances may be underpowered.

fundamentally analogical, then estimation should require selecting a numerical reference point (e.g., a point on the line or a previous estimate), identifying a relation between that reference point and the target number being estimated, and extending this numerical relation to the spatial domain. To test how the analogical demands of number-line estimation tasks affect the developmental trajectory of estimation, we asked (a) which reference points children use and (b) how they use these reference points when estimating. We showed that by 5 years of age, although children are very inaccurate estimators, they are also not random and have begun to use analogical processes to guide their estimates. Specifically, we found that children attend to the numerical ordering of the target number and their previous estimates—and respect this ordering information when placing estimates on the number line. We also found that number-line estimation changes (a) as children learn to identify and order reference points on the number line, (b) as they learn to represent the relative distance between these points, and (c) as they become better able to access previous estimates when they are not otherwise visually available. These data suggest that important changes in the development of number-line estimation are related to the analogical demands of number-line tasks.

Counter to the view that poor estimation performance reflects a lack of familiarity with a particular numerical range (e.g., Lipton & Spelke, 2005; Siegler & Opfer, 2003), we found that even our youngest participants made ordinal responses—meaning that their estimates were in the correct direction relative to previous estimates—and that they did so even when they failed to make accurate estimates. Amazingly, this was true at all magnitudes tested, suggesting that children made use of ordinality across the entire number-line. This finding demonstrates that very young children may have a greater understanding of the relation between numbers within the count list than had previously been documented (see also Barth et al., 2009; Sullivan & Barner, 2014).

We also showed that the accuracy and ordinality of 5-year-olds' estimates improved when these children had visual access to previous estimates. This suggests that 5-year-olds' difficulties with estimation are not purely due to a lack of numerical knowledge but are also partially the result of difficulty in accessing previous estimates. As noted earlier, a critical component of the analogical comparison required by estimation is relating a target number to a reference point. If children use their previous estimates as reference points, then their ability to access these previous estimates could limit estimation performance. Consistent with this, we found that our youngest participants made better estimates when they had visual access to their previous estimates. Critically, having visual access to previous estimates should *not* have influenced estimation performance if children did not use their previous estimates as points of comparison. Although the current study did not directly test which cognitive mechanisms limited our youngest participants from using their previous estimates when these estimates were not visually accessible, possible explanations include working memory, domain-general processing ability, and the ability to update mental representations. Future studies should explore these possibilities.

We also found that attested changes to estimation performance between 5 and 7 years of age (e.g., Booth & Siegler, 2006; Siegler & Opfer, 2003;) may stem in part from developmental changes to children's ability to construct sophisticated analogical numerical comparisons. For example, we found that the ability to make comparisons on the basis of the relative distance between numbers (and not just their ordinality) develops gradually between 5 and 7 years of age. Evidence for this came from two sources. First, our simulated estimation data showed that estimates generated by following trial-to-trial ordinality alone are relatively similar to 5-year-olds' actual estimates but are relatively dissimilar to the estimates provided by our 6- and 7-year-olds. Second, we found evidence that 6- and 7-year-olds' estimates were affected by which numbers these children estimated first in the experiment but that this was not the case for 5-year-olds. As shown by additional simulations, the most likely explanation of this finding is that only the older children represented the relative distance between previous estimates and new estimates. Taken together, these data suggest that improvements in estimation performance between 5 and 7 years of age stem from multiple sources. Older children not only are better at accessing and manipulating previous estimates but also attend more to the relative distance between numbers (and not just ordinality).

Although this study showed that children make number-line estimates by relating a target number to numerical reference points (e.g., previous estimates), several questions remain about the development of this process. First, the cause of the shift from using information about ordinality to making estimates on the basis of relative distance information remains unknown. On the one hand, children might have knowledge of the relative distance between numbers from a very young age but simply are unable to deploy this knowledge. This might be because they fail to appreciate its relevance to the task, because respecting distance relations imposes additional domain-general demands on working memory, attention, and the like (e.g., Anobile et al., 2013), or because children initially have trouble in transferring information regarding the distance between numerals to the spatial domain. On the other hand, children might initially lack sophisticated knowledge of the relative distance between numbers and acquire this knowledge only after prolonged exposure to the number system and counting routine. Consistent with this, it seemed that even our 5-year-olds used relative distance information for the smallest numbers estimated (see Fig. 4). These two possibilities point to very different mechanisms of estimation change and to different pictures of early numerical knowledge. Future research should disambiguate between these two possibilities.

Another question left open by our study—and our result that 5-year-olds' estimation performance is affected by access to past estimates—is how memory capacity and number-line estimation performance are related. Recent studies, like ours, suggest that estimation and memory for numbers may be related but leave open the mechanism that governs this relation. In one study, Thompson and Siegler (2010) showed that children who had better memory for number words were more likely to be linear estimators—a result that was taken as evidence that memory for numbers is improved by the development of an adult-like system of number representations. However, another explanation for these data is that children who have better memory for numbers may be better able to provide accurate estimates because estimation involves analogical processes that depend critically on memory—that is, accessing and manipulating past estimates. Although our data do not directly test the role of working memory in estimation, our finding that 5-year-olds performed better when they had visual access to their previous estimates is consistent with the idea that memory for numbers is a prerequisite for estimation rather than a product of mature number representations (see Bryant & Trabasso, 1971; see also Siegler & Booth, 2004).

This study suggests that structural analogy is likely fundamental to number-line estimation; children compute within-domain numerical relations between target numbers and numerical reference points when estimating and then analogically extend these relations to the spatial domain (marking their numerical estimates on a spatial number line). Although past reports have detailed the cognitive consequences of the spatial comparisons required by this analogy (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Slusser et al., 2013), ours is the first to test the numerical component of this analogy and to show that children use their own previous estimates as reference points. We showed that children as young as 5 years make analogies when estimating and that both the types of relations that guide these analogies and the ease with which children compute these analogies change over development.

These findings not only contribute to an explanation of the developmental transitions in estimation ability described in this study but also are consistent with past findings in the literature. For example, our data are broadly consistent with reports that young children use experimenter-provided reference points (e.g., midpoint, endpoint) to calibrate their estimates and that children's reference point use affects their estimation (Barth & Paladino, 2011; Slusser et al., 2013; Sullivan et al., 2011). In addition, although these data suggest that estimation tasks depend on non-numerical capacities required by analogy, they are also consistent with a role for developing numerical capacities that change in parallel (e.g., Siegler & Opfer, 2003). Thus, our study suggests that estimation is a multifaceted process that draws on multiple cognitive capacities (Thompson & Opfer, 2010) and that, therefore, cannot be interpreted as a simple pure proxy for the development of numerical representations.

In summary, we have argued that number-line estimation tasks are fundamentally analogical; estimation depends critically on the ability to relate a target number to a previous estimate and then extend this relation to the spatial domain. Our data show that, due to the analogical component of number-line estimation tasks, estimation abilities may be limited by factors that are not directly related to numerical knowledge. This suggests that previous work may have underestimated young children's estimation abilities and that 5-year-olds may possess more sophisticated knowledge of the number system than previously believed. In addition, these findings paint a new picture of the mechanisms underlying estimation performance by demonstrating that analogical comparison guides young children's estimates.

Acknowledgments

Thanks go to Eleanor Chestnut, Kevin Smith, Katie Wagner, and members of the LADLab team for helpful comments on this manuscript. Thanks also go to Ada Tong and Michael Sim for help in gathering and coding data, to the staff and families at the Reuben H. Fleet Science Center for their collaboration, and to the schools and day-care facilities that made this research possible. Data for this study were collected while J.S. was a student at the University of California, San Diego. This work was supported by a Jacobs Graduate Fellowship and a National Science Foundation (NSF) Graduate Research Fellowship to J.S. and by a James S. McDonnell Foundation grant to D.B. None of the funding sources played any role in study design, data collection, analyses and interpretation of data, writing of this article, or the decision to submit the manuscript for publication.

References

Anobile, G., Stievano, P., & Burr, D. (2013). Visual sustained attention and luminosity sensitivity correlate with math achievement in children. *Journal of Experimental Child Psychology, 116*, 380–391.
Barth, H., & Paladino, A. (2011). The development of numerical estimation in children: 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*, 248–264.
Bates, D., & Sarkar, D. (2007). *LME4: Linear mixed-effects models using S4 classes (R package, Version 0.9975-12)*. . http://CRAN.R-project.org.
Booth, J., & Siegler, R. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 41*, 189–201.
Booth, J., & Siegler, R. (2008). Numerical magnitude representations influence math learning. *Child Development, 79*, 1016–1031.
Bryant, P., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature, 232*, 456–458.
Cantlon, J., Cordes, S., Libertus, M., & Brannon, E. (2009). Comment on "Log or linear? Distinct intuitions of the number scale in Western Amazonian indigene cultures". *Science, 323*, 38b.
Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.
Cohen, D., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review, 18*, 331–338.
Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number-line and familiarity with numbers in 5- to 9-year-old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology, 99*, 1–17.
Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*, 1221–1247.
Kolkman, M., Hoijtink, H., Koresbergen, E., & Leserman, P. (2013). The role of executive functions in numerical magnitude skills. *Learning and Individual Differences, 24*, 145–151.
Lipton, J., & Spelke, E. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76*, 978–988.
Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology, 103*, 503–515.
Ramani, G., & Siegler, R. (2008). Promoting broad and stable improvements in low income children's numerical knowledge through playing number board games. *Child Development, 79*, 375–394.
R Development Core Team (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Schneider, M., Grabner, R., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in Grades 5 and 6. *Journal of Educational Psychology, 101*, 359–372.
Siegler, R., & Booth, J. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428–444.
Siegler, R., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237–243.
Siegler, R., & Ramani, G. (2008). Playing linear numerical board games promotes low income children's numerical development. *Developmental Science, 11*, 655–661.
Slusser, E., Santiago, R., & Barth, H. (2013). Developmental changes in numerical estimation. *Journal of Experimental Psychology: General, 142*, 193–208.
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. (2014). Inference and association in children's numerical estimation. *Child Development, 85*, 1740–1755.
Sullivan, J., Juhasz, B., Slattery, T., & Barth, H. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin and Review, 18*, 557–563.
Thompson, C., & Opfer, J. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81*, 1768–1786.

Thompson, C., & Siegler, R. (2010). Linear numerical-magnitude representations aid children's memory for numbers. *Psychological Science, 21*, 1274–1281.
Vul, E., Sullivan, J., & Barner, D. (2013). Slow drift of individuals' magnitude-to-number mapping. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
Whalen, J., Gallistel, C., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*, 130–137.