



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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



Levels of number knowledge during early childhood

Barbara W. Sarnecka*, Michael D. Lee

Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA

ARTICLE INFO

Article history:

Received 6 September 2008

Revised 6 February 2009

Available online 5 April 2009

Keywords:

Counting

Estimation

Give-N

Guessing

Knower levels

Number words

Performance error

ABSTRACT

Researchers have long disagreed about whether number concepts are essentially continuous (unchanging) or discontinuous over development. Among those who take the discontinuity position, there is disagreement about how development proceeds. The current study addressed these questions with new quantitative analyses of children's incorrect responses on the Give-N task. Using data from 280 children, ages 2 to 4 years, this study showed that most wrong answers were simply guesses, not counting or estimation errors. Their mean was unrelated to the target number, and they were lower-bounded by the numbers children actually knew. In addition, children learned the number-word meanings one at a time and in order; they treated the number words as mutually exclusive; and once they figured out the cardinal principle of counting, they generalized this principle to the rest of their count list. Findings support the 'discontinuity' account of number development in general and the 'knower-levels' account in particular.

© 2009 Elsevier Inc. All rights reserved.

Introduction

The human mind can represent many different kinds of information. A basic developmental question is whether these representational resources change over time. Everyone believes that learning adds new information to existing representational systems. But only some people believe that learning can also result in the creation of new representational systems, with the ability to represent new kinds of information.

Nowhere is this debate more salient than in the domain of number. Large, approximate number concepts (i.e., analog-magnitude number concepts) are clearly shared, not only between human infants and adults but also between humans and other species (Brannon, 2002; Brannon, Abbott, & Lutz,

* Corresponding author. Fax: +1 949 824 7235.

E-mail address: sarnecka@uci.edu (B.W. Sarnecka).

2004; Cantlon & Brannon, 2007; Dehaene, 1997; Flombaum, Junge, & Hauser, 2005; Gallistel, 1990; Lipton & Spelke, 2003; McCrink & Wynn, 2004; Meck & Church, 1984; Wood & Spelke, 2005; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005). The same is true of parallel individuation—the ability to track small sets of up to three or four individuals at a time (Antell & Keating, 1983; Feigenson & Carey, 2003, 2005; Feigenson, Carey, & Hauser, 2002; Hauser & Carey, 2003; Leslie, Xu, Tremoulet, & Scholl, 1998; Starkey & Cooper, 1980; Uller, Carey, Huntley-Fenner, & Klatt, 1999; Wynn, 1992a). Recent studies even suggest that singular/plural—the privileged distinction between individuals and sets that structures so many human languages—is shared by non-human primates (Barner, Wood, Hauser, & Carey, 2008).

Some scholars have argued that counting (specifically, knowledge of the principles that structure counting) is another example of a shared capacity. For adults, the meaning of any cardinal number word (e.g., “two,” “five,” “eighteen”) is determined by that word’s ordinal position in the list (e.g., even without speaking Russian, one knows that the fiftieth word in the Russian count list means 50). It is possible that children have the same understanding. In other words, it is possible that large, exact-number concepts are essentially continuous over development (Cordes & Gelman, 2005; Gelman & Gallistel, 1978; Gelman, Meck, & Merkin, 1986; Greeno, Riley, & Gelman, 1984).

Others however, have cited counting as an example of developmental discontinuity. In these accounts, children initially derive the cardinal meanings for “one” through “three” or “four” not from counting, but rather from a limited-capacity system based on parallel individuation. This system is limited in capacity because it can handle sets of only up to three or four individuals. To represent “five” and higher numbers, children must master counting. Thus, the discontinuity in these accounts comes from the change in representational systems; the limited-capacity system provides children with the cardinal meanings of “one” through “four,” and a counting system provides meanings for all higher number words.

Some scholars consider the continuity/discontinuity question to be settled. Indeed, studies going back several decades have argued that children initially fail to understand counting (e.g., Baroody & Price, 1983; Briars & Siegler, 1984; Frye, Braisby, Lowe, Maroudas, & Nicholls, 1989; Fuson, 1988, 1992; Klahr & Wallace, 1976; Wagner & Walters, 1982; Wynn, 1990, 1992b). But even for those who believe that number development is discontinuous, there remains the question of how to describe and explain the discontinuities.

The most developed account to date is what we call the knower-levels account (Carey, 2001, 2004; Carey & Sarnecka, 2006; Condry & Spelke, 2008; Le Corre & Carey, 2007; Le Corre, Van de Walle, Brannon, & Carey, 2006; Wynn, 1992b). According to this proposal, the child learns the cardinal meanings of “one” through “four” one at a time and in order. (The term *knower-level* refers to the child’s progress on this front. The child starts out as a *pre-number knower* and then progresses through the *one-knower*, *two-knower*, *three-knower*, and [for some children] *four-knower* levels.) At this point, the cardinality associated with each number word is still represented via the limited-capacity system.

A conceptual breakthrough happens when the child figures out the cardinal principle of counting (Gelman & Gallistel, 1978). In procedural terms, this principle states that when we count a set of items, the last number word we say indicates the cardinality of the whole set. In conceptual terms, this principle means that the cardinal meaning of any number word is fixed by the word’s ordinal position in the list.

Children who have figured this out are called *cardinal principle knowers* (sometimes abbreviated *CP knowers*). The knower-levels account claims that when children learn this principle, they quickly generalize it to all of the words in their count list. Thus, although they learned the meanings of “one,” “two,” and “three” (and possibly “four”) one at a time, they learn the meanings of “five” and higher number words all at once (Carey, 2001, 2004; Wynn, 1990, 1992b; see also Klahr, 1984; Klahr & Wallace, 1976).

Another claim associated with the knower-levels account is that children view number-word meanings as mutually exclusive. For example, two-knowers know that “one” means 1, and “two” means 2. But they further assume that no other number words mean 1 or 2; and that the words “one” and “two” cannot refer to any other set sizes (Sarnecka & Gelman, 2004; Wynn, 1992b).

Thus, there are at least two points of debate. Some scholars remain interested in the broad question of continuity/discontinuity: Do children initially represent cardinal number word meanings through a

counting process, or through a limited-capacity (i.e., parallel individuation/subitization) process? Other scholars accept the discontinuity view, but not necessarily the knower-levels view: Are knower-levels a real phenomenon? Does it really make sense to sort children in this way, or to report the results of studies by knower-level rather than age? The point of the current article is to show that these long-standing questions are answerable given enough data and the right analyses. The key is to look at children's incorrect answers (not just the correct ones) and to understand the distinction between performance errors (i.e., errors of counting or estimation) and guessing.

Performance error versus guessing

The current study uses the Give-N task (Bermejo, 1996; Condry & Spelke, 2008; Frye et al., 1989; Le Corre & Carey, 2007; Le Corre et al., 2006; Schaeffer, Eggleston, & Scott, 1974; Wynn, 1990, 1992b). In this task, the child is asked to produce a set of some number of items. For example, the experimenter might ask the child to “give three bananas” (from a bowl of 15 toy bananas) to a toy monkey. If the child does not give 3 bananas, then the answer is wrong. Wrong answers happen for one of two reasons. The first reason is performance error. If the child knows that “three” means 3, but gives some other number of bananas, it is because he or she made a mistake in counting or estimation. The second source of wrong answers is ignorance, which results in guessing. If the child does not know that “three” means 3, then he or she simply needs to guess how many items to give.

Performance error (for both counting and estimation) is well understood, and the wrong answers follow predictable distributions, as illustrated in Fig. 1 (see, e.g., Barth, La Mont, Lipton, & Spelke, 2005; Barth et al., 2006; Cordes, Gelman, Gallistel, & Whalen, 2001; Dehaene, 1997; Gallistel, 1990; Huntley-Fenner, 2001; Platt & Johnson, 1971; Whalen, Gallistel, & Gelman, 1999). Qualitatively, these predictions are the same for both children and adults, although children may be less accurate overall. The key thing to notice is that the wrong answers do not fall haphazardly. They are grouped symmetrically around the target number. So, for example, when people make errors in counting or estimation, they

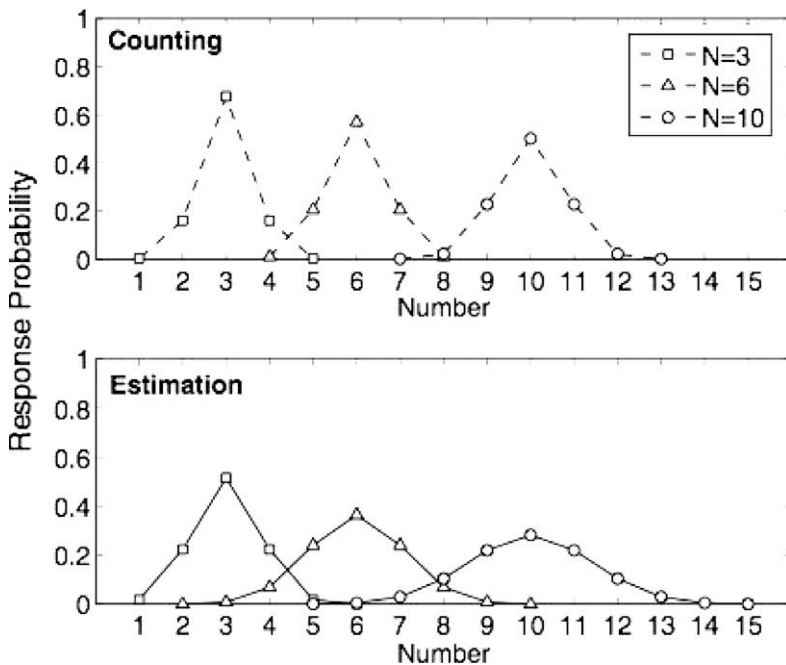


Fig. 1. Illustration of hypothetical response curves predicted by two types of performance error: counting error (top) and estimation error (bottom). To minimize visual clutter, the figure shows only the target number words “three,” “six,” and “ten.”

still give twice as many objects, on average, when they are asked for “six” as when they are asked for “three.”

Guessing is different from performance error. Unlike performance errors, guesses are unrelated to the target. If the child really does not know what set size he or she is supposed to produce (because the child does not know the cardinal meaning of the number word requested), then that information cannot affect his or her answer. For example, imagine that you (the reader) are a participant in a Give-N study administered in Japanese. Imagine further that all you know of the Japanese number list is *ichi*, *ni*, *san* (one, two, three). If you are asked for any other number (e.g., *go* or *jyuu*), then you simply need to guess how many objects to give, and your guess (unlike performance error) will be unrelated to the number requested. There is no reason for you to give twice as many objects when they ask you for *jyuu* as when they ask you for *go* because you do not know that *jyuu* means “ten” and *go* means “five.”

Thus, quantitative analyses can distinguish performance error from guessing. If most of children’s Give-N errors are performance errors (i.e., they tend to fall near the target, equally above and below it), then we can conclude that the continuity view is correct; children understand number words in essentially the same way as adults, but children simply make more mistakes in counting or estimating. On the other hand, if most of children’s Give-N errors are guesses (i.e., they are unrelated to the target), then we can conclude that the discontinuity view is correct; number development involves real conceptual changes, not just improvements in accuracy. Deciding between these alternatives was the goal of the first part of our analysis.

The second part of our analysis tested the claims of a particular discontinuity proposal—the knower-levels account. Specifically, we tested the knower-levels claims that (a) children learn the first few number words in order; (b) children treat the number words as mutually exclusive; and (c) when children figure out the cardinal principle of counting, they quickly generalize this knowledge to the rest of their count list.

Method

We examined these questions by analyzing previously published Give-N data from three different samples. The Give-N task procedure was the same in all cases, but samples differed in the number of trials per child, in the target numbers for which children were asked, and in the other counting tasks that children completed.

Dataset 1

Data were collected from 82 monolingual speakers of English, ages 2 to 4 years (mean = 3 years 7 months, range = 2 years 11 months to 4 years 6 months), tested at preschools in Irvine, California, or at a university cognitive development lab in Cambridge, Massachusetts. Children completed the Give-N task as part of their participation in other studies. Each child in Dataset 1 also completed an intransitive counting task where the experimenter simply asked the child to “count to ten.” Our analysis includes only those children who counted to “ten” perfectly. (An additional 7 children participated in the original studies but were excluded from this analysis because they did not count to “ten” perfectly.) Thus, we can be sure that every child in Dataset 1 was familiar with the number word list through “ten.”

Some of the children ($n = 64$) also completed a transitive counting task where they were asked to count arrays of 5 and 10 objects. The mean score on the array of 10 was 8.3 if no errors are allowed or was 9.5 if one skip or double-count error is allowed. Because the current analysis distinguishes performance errors from outright guessing, it does not rely on children being able to count perfectly. But we include this information to give the reader a sense of where the children were in terms of procedural counting skill. Detailed descriptions of both counting tasks can be found in Sarnecka and Carey (2008).

For the Give-N task, the child was presented with a bowl of 15 identical small toys (e.g., rubber bananas approximately 3 cm long) and was asked to place some number of the toys on a plate for a stuffed animal. (e.g., “Can you give three bananas to the monkey?”) After the child finished placing objects on the plate and slid the plate across the table to the stuffed animal, the experimenter asked

a single follow-up question of the form “Is that N ?” (e.g., “Is that *three*?”). If the child said no, then the experimenter returned the plate to the child and restated the original prompt. Each child completed 4 to 21 trials of this task (mean = 12.4 trials). Numbers requested included “one,” “two,” “three,” “four,” “five,” “six,” “eight,” and “ten.”

For the children in Dataset 1, the number “six” was requested only if the child had correctly given 5 when asked for “five” on a previous trial. Children in this dataset who had not produced sets of 5 correctly were never asked to give “six.” This produced a somewhat different distribution of errors for “six” than for other target numbers (see Results section below). Trials for “six” in Dataset 1 were the only subset of data affected in this way.

Dataset 2

These data were collected from 162 children, ages 2 or 3 years (mean = 3 years 2 months, range = 2 years 9 months to 3 years 7 months), living in Ann Arbor, Michigan (70 children); Kobe, Japan (48 children); or St. Petersburg, Russia (44 children). All children were tested in their native language by a female native speaker.

Each child completed 9 to 15 trials of the Give- N task (mean = 14.7 trials). The numbers requested were “one,” “two,” “three,” “five,” and “six.” These children were not given a count-to-ten task.

Dataset 3

These data were collected from 36 children, ages 2 to 4 years (mean = 3 years 3 months, range = 2 years 7 months to 4 years 1 month), tested in communities around Boston, Massachusetts. Of these children, 21 were native English speakers tested in English and 15 were native Russian speakers tested in Russian.

Each child completed 26 to 36 trials of the Give- N task (mean = 35.6 trials). The numbers requested were “one,” “two,” “three,” “four,” “five,” and “ten.” These children were not given a count-to-ten task.

Results

The children understood and cheerfully complied with the Give- N task. In particular, there was no evidence that the follow-up question (“Is that N ?”) caused children to doubt their answers or self-correct. Children stuck with their original answers on more than 99.8% of trials, and no child ever changed a correct answer. For detailed descriptions of the children’s behavior (e.g., pointing, counting aloud), as well as their performance on other number tasks, see Sarnecka and Carey (2008), Sarnecka and Gelman (2004), and Sarnecka, Kamenskaya, Yamana, Ogura, and Yudovina (2007).

In the following analyses, we first take up the question of whether the incorrect Give- N responses are guesses or performance errors, and then we test the specific claims of the knower-levels theory.

Performance error versus guessing, part 1: Incorrect responses did not cluster around the target

Performance errors cluster around the target, such that the mean incorrect response for any target is the same as the target number itself (e.g., the mean incorrect response on trials asking for “five” is always 5, the mean incorrect response on trials asking for “ten” is always 10). Guesses, by contrast, are unrelated to the target. To make this analysis maximally powerful, it is desirable to pool the data from all three datasets. However, performance error is possible only if children have memorized the number word list, and this could raise questions about Datasets 2 and 3, where children were not asked to recite the list. Thus, we present the analysis twice: once for all three datasets together and again for Dataset 1 by itself.

The mean incorrect response for each number word is given in Table 1. For Datasets 1, 2, and 3 pooled, the mean incorrect response was different from the target number in every case. More important, the mean incorrect responses were uncorrelated with the targets ($r = .16$). If children systematically under- or overcounted, or if they systematically under- or overestimated, then their wrong

Table 1
Summary of incorrect responses.

Target number	Total trials for that target number	Number and percentage of trials garnering incorrect responses	Mean incorrect response	Is the mean incorrect response different from the target number?
<i>Datasets 1, 2, and 3 pooled</i>				
One	831	51 (6)	7.94	$t = 8.37, p < .001$
Two	803	206 (26)	4.37	$t = 8.05, p < .001$
Three	870	441 (51)	3.93	$t = 5.21, p < .001$
Four	341	188 (55)	4.75	$t = 2.72, p < .01$
Five	823	621 (75)	5.42	$t = 2.32, p < .05$
Six	534	447 (84)	4.72	$t = 6.97, p < .001$
Eight	(all trials occurred in Dataset 1 [see below])			
Ten	265	223 (84)	5.50	$t = 15.41, p < .001$
<i>Dataset 1 only</i>				
One	130	10 (8)	11.60	$t = 6.28, p < .001$
Two	102	16 (16)	9.07	$t = 5.32, p < .001$
Three	169	54 (32)	6.81	$t = 5.32, p < .001$
Four	126	60 (48)	8.07	$t = 6.87, p < .001$
Five	153	75 (49)	8.51	$t = 6.57, p < .001$
Six	74	24 (32)	6.54	$t = 1.17, p = .25, ns$
Eight	52	39 (75)	8.84	$t = 0.47, p = .64, ns$
Ten	54	43 (80)	8.72	$t = 1.99, p = .10, ns$

Note. Percentages are in parentheses.

answers would still be correlated with the targets, even if the means were not the same. The pattern we see in these data indicates that children were not making performance errors at all; they were simply guessing.

The mean for “one” was higher than for the other targets because many children who failed to give 1 object simply gave all 15. (They did the same on trials asking for other numbers, but the mean for “one” was most affected because there were so few errors for “one” overall.)

If we consider Dataset 1 alone, the mean incorrect responses were still poorly correlated with the targets ($r = .31$). In this dataset, the mean incorrect responses for “six,” “eight,” and “ten” actually did not differ from the targets. But this was not the clustering of answers around the target predicted by performance error because the mean incorrect responses for targets “two” through “ten” all were very similar. Rather, the t tests for “six,” “eight,” and “ten” were not significant because the mean incorrect response in Dataset 1 was around 7 to 9 items for all targets above “one.” (For a discussion of why the means in Dataset 1 were higher overall than in Datasets 2 and 3, see “Did children treat number words as mutually exclusive?” subsection below.)

An interesting point about Dataset 1 concerns the responses for “six.” The mean incorrect response for “six” was a little lower than the responses for “four,” “five,” “eight,” and “ten” (although this difference did not reach statistical significance using a one-way analysis of variance [ANOVA]), $F = 1.06$, $p = .38$, *ns*. This makes sense when we consider that children were asked for “six” objects only after they had successfully given 5 for “five” (see Method section above). Because the limited-capacity system works only up to 3 or 4 items, sets of 5 and above must be counted. Therefore, children who give 5 items for “five” were either cardinal-principle knowers or lucky guessers. (Distinguishing these two was the reason for asking children to give “six” after they had succeeded with “five” in the original study.) Because the children who were asked for “six” were more likely than other children to be cardinal-principle knowers, their errors were more likely to be performance errors. Thus, the mean of those errors was closer to the target.

Performance error versus guessing, part 2: Incorrect responses were lower-bounded by children’s knowledge, not symmetrical about the target

Another prediction of performance error (true of both counting and estimation error) is that incorrect responses should be symmetrical: Overcounting and undercounting should happen

equally often, as should overestimation and underestimation. For example, consider the errors that you (the reader) might make on the Give-N task. You know what all of the number words mean, so your errors will all be performance errors, not guesses. If you were asked to produce sets of “five” over and over again, you might occasionally give 4 or 6 items (or even 3 or 7 items) by mistake. And the chance of you giving 4 would be equal to the chance of you giving 6, just as the chance of you giving 3 would be equal to the chance of you giving 7. This is true because performance error is symmetrical.

On the other hand, to return to an earlier example, imagine that you were presented with the Give-N task in Japanese, and you could count to only three in Japanese (*ichi, ni, san*). If someone asked you to give *go*, you might give 4, 5, 6, 7, 8, 9, or more items. But you probably would not give 1, 2, or 3 items because you already know the Japanese words for those cardinalities. In other words, your guesses would be *lower-bounded* by the number words you know. Thus, you would be much more likely to give 7 items than 3 items, even though these are equidistant from the target *go* (five).

Note that this prediction holds true only if children consider number-word meanings to be mutually exclusive, a claim that has also been made in the knower-levels literature (Wynn, 1992b; see also Sarnecka & Gelman, 2004). Thus, with this analysis, we simultaneously test two things: (a) whether most wrong answers are performance errors or guesses and (b) whether children treat the number words as mutually exclusive.

To do this, we identified the group of children who performed perfectly on each target number. For example, a child who always gave 3 items when asked for “three,” and never gave 3 items when asked for any other number, was considered to be ‘perfect on 3’. Note that this sorting made no assumptions about how the child performed at any other number. A child might perform perfectly on 3 but make mistakes on 1, 2, 4, 5, 6, 8, and 10. Another child might perform perfectly on all numbers. Both children would be included in the group ‘perfect on 3’.

Fig. 2 illustrates an example. This figure shows data for all children who were perfect (P) on 3 in the trials where they were asked (A) for “five.” The *x* axis lists all of the possible set sizes children could produce, and the *y* axis shows the number of times each set size was produced. There are no responses of 3 because they were selected out (all of these children performed perfectly on 3). Responses of 5 do not appear because they were not errors (“five” was the number asked for). White columns on the graph show wrong answers that were not counted in our analysis. Black columns show the answers that were counted.

As was the case for most target numbers, many responses fell in the range of 4 to 6 (see Table 1). For our purposes however, the relevant comparisons are between pairs of equidistant responses where the lower response overlaps with the child’s perfect number (in this example, 3). Thus, the relevant pairs for this example are responses of ± 3 or ± 4 from the target. As Fig. 2 shows, there were five such errors, all falling above the target, none below it. This analysis was repeated for all possible pairs of ‘perfect’ and ‘asked’ numbers.

It is important to note that we conducted this analysis in a way that was actually biased *against* the knower-levels explanation. We excluded errors that lacked potential matches below the target (e.g., responses of 10–15 in Fig. 2). However, we included responses that lacked potential matches *above* the target, because these might provide important evidence against the knower-levels account. For example, if a child was asked for “ten” items and gave only 2, we counted that error. We did so even though an equivalent error in the other direction (i.e., a response of 18) was not possible because there were only 15 items in the bowl.

Final counts

The final counts for each dataset were as follows. Dataset 1 contained 28 errors above the target (1 or more for 19 of the 82 children) and 0 errors below it. Dataset 2 contained 31 errors above the target (1 or more for 17 of the 162 children) and 1 error below it. This error was made by a child who performed perfectly on 2 but gave 1 object on another trial asking for “three.” Dataset 3 contained 11 errors above the target (1 or more for 6 of the 36 children) and 12 errors below it, all produced by the same child. This child performed perfectly on trials asking for “one,” “two,” “three,” and “four” but

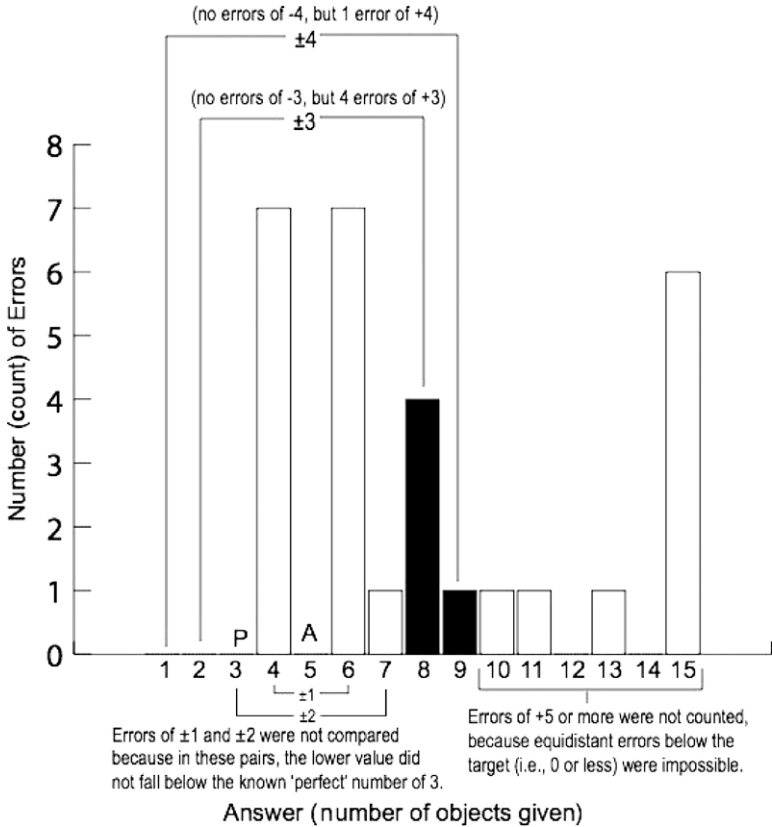


Fig. 2. Illustration of the analysis showing that children's incorrect answers were lower-bounded by the numbers they knew (rather than being symmetrical about the target). "P" stands for "perfect," indicating that this group of children performed perfectly on the number 3 (i.e., they always gave 3 objects when asked for "three" and never gave 3 otherwise). "A" stands for "asked," indicating that the children were asked for "five" items. This is merely an illustrative example; the analysis was repeated for all possible pairs of numbers.

gave 2 objects in all 12 trials asking for "ten." Because this error is not explained by either account (i.e., it does not look like guessing and does not look like counting or estimation error), we conclude that the child probably mistook the word "ten" for the word "two." Aside from this child, there were no below-target errors in Dataset 3.

The grand total, then, was 13 below-target errors (or only 1 if the child with the apparent "ten/two" confusion is discounted). There were 70 above-target errors—a clearly asymmetrical result.

These results demonstrate two things. First, they provide convergent evidence that most of the incorrect responses in these datasets are simply guesses, not errors of counting or estimation. Second, when children are asked for numbers they do not know, their guesses are lower-bounded by the numbers they do know. In other words, children do treat the number words as mutually exclusive, as the knower-levels account has claimed.

Testing the knower-levels account

So far, we find strong support for the view that number development is discontinuous. Next, we test the claims of a specific discontinuity theory—the knower-levels account.

Did children treat number words as mutually exclusive?

Above, we found evidence for the knower-levels claim that children treat the number words as mutually exclusive. This mutual-exclusivity constraint means that a child's guesses about unknown number words are lower-bounded by the meanings of the number words he or she knows.

This lower-boundedness also explains why the average wrong answer for each target was higher in Dataset 1 than in Datasets 1, 2, and 3 pooled. If we look at the highest number on which each child performed perfectly, we find that the mean highest perfect number for children in Dataset 1 was 3.94, whereas the mean for all datasets pooled was 2.15. In other words, the children in Dataset 1 simply knew more number-word meanings. This is not surprising given that they were an average of 5 months older.

The mutual-exclusivity constraint means that when children make guesses about numbers whose cardinal meanings they do not know (e.g., "eight"), they avoid giving set sizes whose names they *do* know. (This is the lower bound placed on their guesses.) For example, because children in Dataset 1 were more likely to know "three" and "four," they were less likely to produce sets of 3 or 4 when asked for "eight." This resulted in a higher mean guess for Dataset 1.

Did children learn the number word meanings in order?

The central claim of the knower-levels theory is that children learn the number-word meanings in order. In other words, if a child knows that "four" means 4, he or she should also know that "three" means 3, "two" means 2, and "one" means 1. To examine this, we started with each child's highest perfect number and looked at how the child performed on trials asking for lower numbers. For example, if a child's highest perfect number was 4, we look at how that child performed on trials asking for "one," "two," and "three."

A breakdown of all children by age and highest perfect number appears in Fig. 3. Note that our criterion for knowing a number required a child to give 100% correct responses and no incorrect responses for that number. More typical knower-levels studies give a child credit for knowing a number if he or she has twice as many correct responses as incorrect responses for that number (e.g., Wynn, 1992b). For our purposes here, the stricter criterion of 100% correct is preferable because it does not gloss over any errors (and errors are the focus of our analysis). For a sorting of these children into knower levels (based on the typical criteria), see Sarnecka and Carey (2008), Sarnecka and Gelman (2004), and Sarnecka and colleagues (2007).

Once we exclude from the analysis children who did not perform perfectly at any number ($n = 57$) and those whose highest perfect number was 1 ($n = 82$), we are left with 141 children. Of these, 134 children (95%) showed a perfect pattern of learning the number words in order. That is, they performed perfectly on every trial for every target number below their highest perfect number. Of the remaining 7 children, 6 made only a single mistake on a lower number trial. For example, one child's highest perfect number was 3. On one trial asking for "two," she gave 5 objects. On all of the other trials asking for "two," she gave 2 objects. The final child (whose highest perfect number was 10) made two mistakes: On one trial asking for "three" she gave 4, and on another trial asking for "five" she gave 6. Overall, these data strongly support the claim that children learn the number word meanings in order.

Do children immediately generalize the cardinal principle to the rest of their counting list?

According to the knower-levels view, sometime after becoming a three- or four-knower, a child figures out the cardinal principle of counting. This allows the child to assign cardinal meanings to all of the other number words in his or her counting list (Wynn, 1992b; see also Carey, 2001, 2004; Carey & Sarnecka, 2006).

The current data are not ideal for testing this claim because most children were never asked for more than "five" or "six" items. However, a small subset of children was asked for "five," "ten," and sometimes "eight" items. (The numbers requested depended on the study in which the child was enrolled and had nothing to do with the child's performance.)

From among these children, we identify those who performed perfectly at any high target (i.e., 5, 8, or 10). There were 8 such children. According to the knower-levels view, we can reasonably assume

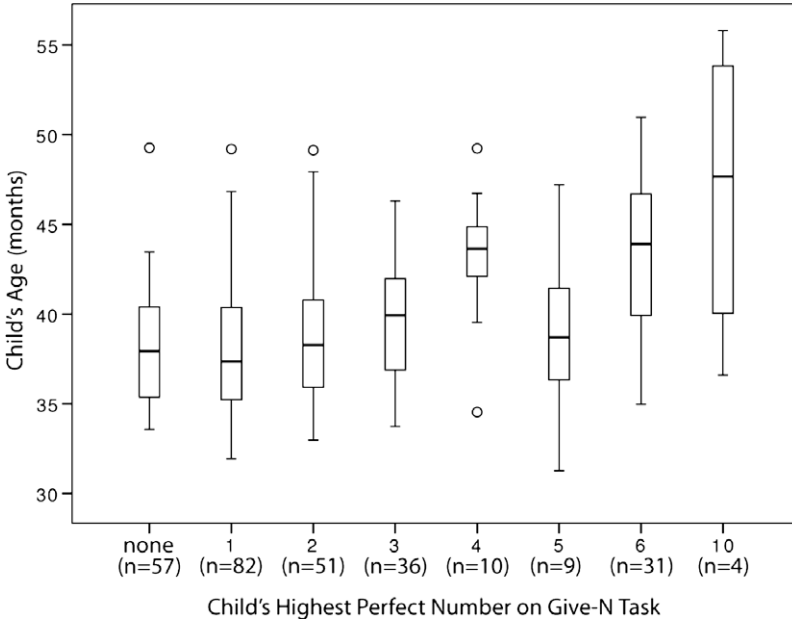


Fig. 3. Breakdown of participants by age and highest perfect number. Note that not all children were tested on the same set sizes (see Method). For most children, the highest number requested was “five” or “six.”

Table 2

Performance of cardinal principle knowers on trials asking for “five,” “eight,” and “ten” objects.

Number requested by experimenter	Number of objects given by child														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Five (38 trials)					37	1									
Eight (9 trials)							1	8							
Ten (39 trials)									2	29	3				5 ^a

Note. Values in cells indicate the numbers of responses. Blank cells indicate zero responses.

^a All five of these responses were made by the same child, who gave 10 objects on one trial asking for “ten” and gave all 15 objects on the other five trials.

that a child who performs perfectly at 5 or any higher number is a cardinal-principle knower. Thus, any errors made by these children should be performance errors rather than guesses.

Results (see Table 2) support the knower-levels theory. Incorrect responses were rare, and when they occurred they seemed to be performance errors (i.e., they fell immediately above or below the target). The exception was a single child who gave 10 objects on one trial asking for “ten” but gave all 15 objects on the other five trials asking for “ten.” By successfully producing a set of exactly 10 objects once, this child demonstrated that he could do it. His choosing to dump out the whole bucket on subsequent trials might indicate noncompliance rather than ignorance.

We can also check this knower-levels claim by going in the opposite direction. We can identify what seem to be counting or estimation errors (i.e., errors that fall close to high-number targets) and ask whether the children who made them were more likely to be cardinal-principle knowers. This method has the advantage of including children who did not perform perfectly at any high number.

We looked at errors falling just above or below high number-targets—trials asking for “eight” and getting responses of 7 or 9, and trials asking for “ten” and getting responses of 9 or 11. Altogether, there were 23 such errors in the dataset. The average highest perfect number of children making these

errors was 3.00, and the average highest perfect number of children making all other errors was 1.31. This difference was significant, $t = 3.57, p < .001$.

The result is the same if we include errors ± 2 from the target (i.e., responses of 6, 7, 9, or 10 for “eight”; responses of 8, 9, 11, or 12 for “ten”). In this case, there were 34 relevant errors and the average highest perfect number of the children making them was 2.68, as compared with 1.31 for all other children. Again, the difference was statistically significant, $t = 3.98, p < .001$. In other words, the children who made performance errors did indeed know more numbers (i.e., they had higher knower-levels) than the sample at large, a finding consistent with the knower-levels view.

Discussion

Overall, our results strongly support the view that number development is discontinuous and involves real conceptual change. Specifically, we find support for the claims of the knower-levels account of number development. All of the children in Dataset 1 knew how to count to “ten,” but most of them did *not* know the cardinal meanings of all the number words up to “ten.” If they had known what the words meant, their errors would have been performance errors (mistakes in counting or estimation). But in fact, their errors were unrelated to the target number, indicating that they (most children on most trials) were simply guessing.

But they made *educated* guesses; their guesses excluded the cardinalities they could identify. For example, children who knew the meanings of “one,” “two,” “three,” and “four” did not offer 1, 2, 3, or 4 items when guessing the meaning of any higher number word. In other words, their guesses were lower-bounded by the numbers they knew. Performance errors (unlike educated guesses) are symmetrical. They are not lower-bounded by a person’s knowledge, because they do not reflect ignorance. The lower-boundedness of children’s guesses provides evidence for two claims: (a) the claim that number development is discontinuous and (b) the claim (associated with the knower-levels theory) that children treat the number words as mutually exclusive even before they know the words’ exact cardinal meanings.

These findings also support the knower-levels claim that children learn the number-word meanings in order. The children in this study overwhelmingly performed perfectly up to some number and then made mistakes for every number above that. This is a powerful demonstration of one of the central claims of the knower-levels account.

Finally, our data provide some evidence that once children induce the cardinal principle of counting, they quickly generalize that principle to the rest of the words in their count list. The children who performed perfectly on 5 also performed well on 8 and 10, and their errors looked like performance errors rather than guesses. Similarly, the errors in the study that looked like performance errors (i.e., those that fell within ± 2 of the high-number targets “eight” and “ten”) were made by children with higher knower-levels than the rest of the sample. Because most of the data in this study came from children who had not yet induced the cardinal-principle of counting, they are not ideal for testing claims about cardinal-principle knowers. However, the (relatively few) data points we do have support the knower-levels account.

Our conclusions from this study are both theoretical and methodological. On the theoretical side, we conclude that the knower-levels framework describes a real phenomenon in development. Children do learn the cardinal meanings of the words “one,” “two,” “three,” and perhaps “four” one at a time and in order. During this period, they really do not know what cardinalities the higher number words denote, except that they are different from the cardinalities denoted by the low number words. This finding lends validity to those studies that have used the knower-levels framework in the past.

On the methodological side, these results suggest that future studies of number development in young children should include a knower-levels diagnostic task such as Give-N and should analyze results by knower-level (controlling for age), as well as by age (controlling for knower-level). As Fig. 3 shows, there is great individual variation in the ages at which children progress through the knower-levels, and any particular developmental change could be related to age, knower-level, or both. Knower levels give researchers a relatively precise way to describe the knowledge state of an

individual child. This enables much more powerful analyses of data, and much more valid comparisons of groups of children, both within and across labs.

Acknowledgments

This research was supported by National Institute of Child Health and Human Development (NICHD) Grant 00234342 to the first author. Collection of Dataset 1 was supported by National Science Foundation (NSF) REC-0337055 to Elizabeth Spelke and Susan Carey. Collection of Dataset 2 was supported by a fellowship from the University of Michigan Culture and Cognition Program to the first author and by NICHD Grant HD36043 to Susan Gelman. Collection of Dataset 3 was supported by National Institutes of Health (NIH) Grants R01 HD038338 and 1T32MH067556-01A1 to Susan Carey. We thank the children and families who participated in the original studies and the preschools that hosted that research. We also thank University of California, Irvine (UCI) Cognitive Development Lab managers Helen Braithwaite and Emily Carrigan and research assistants John Cabiles, Alexandra Cerutti, Jyothi Ramakrishnan, Sarah Song, Dat Thai, and Gowa Wu for their help with data collection.

References

- Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, *54*, 695–701.
- Barner, D., Wood, J., Hauser, M., & Carey, S. (2008). Evidence for a non-linguistic distinction between singular and plural sets in rhesus monkeys. *Cognition*, *107*, 603–622.
- Baroody, A. J., & Price, J. (1983). The development of the number-word sequence in the counting of three-year-olds. *Journal for Research in Mathematics Education*, *14*, 361–368.
- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Nonsymbolic arithmetic in adults and young children. *Cognition*, *98*, 199–222.
- Barth, H., La Mont, K., Lipton, J., & Spelke, E. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 14116–14121.
- Bermejo, V. (1996). Cardinality development and counting. *Developmental Psychology*, *32*, 263–268.
- Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, *83*, 223–240.
- Brannon, E. M., Abbott, S., & Lutz, D. (2004). Number bias for the discrimination of large visual sets in infancy? *Cognition*, *93*, B59–B68.
- Briars, D., & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology*, *20*, 607–618.
- Cantlon, J. F., & Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biology*, *5*, 2912–2919.
- Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind & Language*, *16*, 37–55.
- Carey, S. (2004). Bootstrapping and the origins of concepts. *Daedalus*, *133*, 59–68.
- Carey, S., & Sarnecka, B. W. (2006). The development of human conceptual representations. In M. Johnson & Y. Munakata (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 473–496). New York: Oxford University Press.
- Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, *137*, 22–38.
- Cordes, S., Gallistel, C. R., Gelman, R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*, 698–707.
- Cordes, S., & Gelman, R. (2005). The young numerical mind: When does it count? In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 127–142). New York: Psychology Press.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, *6*, 568–584.
- Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, *97*, 295–313.
- Feigenson, L., Carey, S., & Hauser, M. D. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, *13*, 150–156.
- Flombaum, J. I., Junge, J. A., & Hauser, M. D. (2005). Rhesus monkeys (*Macaca mulatta*) spontaneously compute addition operations over large numbers. *Cognition*, *97*, 315–325.
- Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development*, *60*, 1158–1171.
- Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer Verlag.
- Fuson, K. C. (1992). Relationships between counting and cardinality from age 2 to age 8. In J. Bideaud, C. Meljac, & J. Fischer (Eds.), *Pathways to number: Children's developing numerical abilities* (pp. 127–149). Hillsdale, NJ: Lawrence Erlbaum.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Gelman, R., Meck, E., & Merkin, S. (1986). Young children's numerical competence. *Cognitive Development*, *1*, 1–29.
- Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology*, *16*, 94–143.
- Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: Examinations of content and format. *Cognitive Psychology*, *47*, 367–401.

- Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': Numerical estimation by 5- to 7-year-olds. *Cognition*, 78, B27–B40.
- Klahr, D. (1984). Transition processes in quantitative development. In R. Sternberg (Ed.), *Mechanisms of cognitive development* (pp. 101–139). San Francisco: W. H. Freeman.
- Klahr, D., & Wallace, J. G. (1976). *Cognitive development: An information-processing view*. Hillsdale, NJ: Lawrence Erlbaum.
- 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, 395–438.
- Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology*, 52, 130–169.
- Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing “what” and “where” systems. *Trends in Cognitive Sciences*, 2, 1–10.
- Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large number discrimination in human infants. *Psychological Science*, 14, 396–401.
- McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction in infants. *Psychological Science*, 15, 776–781.
- Meck, W. H., & Church, R. M. (1984). Simultaneous temporal processing. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 1–29.
- Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: Effects of error contingencies. *Learning and Motivation*, 2, 386–414.
- Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, 108, 662–674.
- Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition*, 92, 329–352.
- Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, J. B. (2007). From grammatical number to exact numbers: Early meanings of “one”, “two”, and “three” in English, Russian, and Japanese. *Cognitive Psychology*, 55, 136–168.
- Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology*, 6, 357–379.
- Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, 210, 1033–1035.
- Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development*, 14, 1–36.
- Wagner, S. H., & Walters, J. (1982). A longitudinal analysis of early number concepts. In G. Foreman (Ed.), *Action and thought: From sensorimotor schemes to symbolic operations* (pp. 137–161). New York: Academic Press.
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10, 130–137.
- Wood, J. N., & Spelke, E. S. (2005). Infants' enumeration of actions: Numerical discrimination and its signature limits. *Developmental Science*, 8, 173–181.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155–193.
- Wynn, K. (1992a). Addition and subtraction by human infants. *Nature*, 358, 749–750.
- Wynn, K. (1992b). Children's acquisition of number words and the counting system. *Cognitive Psychology*, 24, 220–251.
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74, B1–B11.
- Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, 8, 88–101.