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



# Does Learning to Count Involve a Semantic Induction?

Kathryn Davidson, Kortney Eng, and David Barner

University of California, San Diego

Abstract

We tested the hypothesis that, when children learn to correctly count sets, they make a semantic induction about the meanings of their number words. We tested the logical understanding of number words in 84 children that were classified as ‘‘cardinal-principle knowers’’ by the criteria set forth by Wynn (1992). Results show that these children often do not know (1) which of two numbers in their count list denotes a greater quantity, and (2) that the difference between successive numbers in their count list is 1. Among counters, these abilities are predicted by the highest number to which they can count and their ability to estimate set sizes. Also, children’s knowledge of the principles appears to be initially item-specific rather than general to all number words, and is most robust for very small numbers (e.g., 5) compared to larger numbers (e.g., 25), even among children who can count much higher (e.g., above 30). In light of these findings, we conclude that there is little evidence to support the hypothesis that becoming a cardinal-principle knower involves a semantic induction over all items in a child’s count list.

**Introduction**

As adults, our knowledge of numeral meanings (1, 2, 3, . . .) is unbounded. We know – for any number word that we can think of – that if two sets are labeled with the same numeral – e.g., *fifty-nine* – then both sets contain the same number of items. We also know that if one item is added to either set, then the label for this set is the next word in the count list – i.e., *sixty*. And finally, we know that for any two number words, the word that comes later in the count list refers to a larger quantity. Explaining how children come to know these facts about *all possible numbers* is a problem, since they are only exposed to a finite set of numbers, and exhibit limited knowledge of numbers during early stages of language acquisition. By some accounts, children must make a semantic induction in acquisition, at which point they generalize the above number knowledge to all numerals they have learned, and to all numbers that they will ever learn (Sarnecka & Carey, 2008; Wynn, 1990, 1992). Often, this inductive leap is said to occur when children become so called ‘‘cardinal principle knowers’’, at around the age of 3 and a half. Here, we examine the empirical basis for this hypothesis and whether learning to count involves a semantic induction.

In the early stages of acquisition, children's knowledge of numerals is limited and is mainly procedural in nature. Before they learn any numeral meanings, children first learn to recite the words in a list (e.g., ‘‘*one, two, three...*’’). As part of this procedure, they learn that numerals are always recited in the same order (i.e., the stable order principle) and that each numeral should be said when pointing to a different item (i.e., the one-to-one principle; Gelman & Gallistel, 1978). Sometime after 2 years of age, children begin to acquire meanings for words in their count list, beginning with *one*: ‘‘one-knowers’’ can give one object when asked, and can correctly label a set as one. However, although these children can recite higher numbers (e.g., 5 or 10), and know that these higher words contrast in meaning (Condry & Spelke, 2008; Wynn, 1992), they appear to lack meanings for the rest of the words in their count list (Wynn, 1990, 1992, etc.). After this one-knower stage, children become ‘‘two-knowers’’, and are able to distinguish *one* and *two* from each other and from the rest of the numbers in their count list. Next they learn *three*, (three-knowers) and then (sometimes) *four* (four-knowers). At each of these stages children are known as ‘‘subset- knowers’’ since they have exact meanings for only a subset of the words in their count list. Finally, sometime between the ages of 3-and-a-half and 4, children discover that counting can be used to generate sets of the correct size for any word in their count list. These children are referred to as cardinal principle knowers (CP-knowers), since they appear to understand how counting represents cardinalities.

Previous studies have reported qualitative differences between subset-knowers and CP-knowers in a constellation of different tasks that test knowledge of numerals beyond *four*. CP-knowers, but not subset-knowers, can successfully give sets of 5 or more objects when asked, and can identify sets of 5 or more in a forced choice task (Condry & Spelke, 2008; Le Corre, Van de Walle, Brannon, & Carey, 2006; Wynn, 1990, 1992, etc.). When a set beyond the child’s counting range has an item removed and replaced with a different item, CP-knowers understand that it retains its cardinality and thus that the same number word applies (Lipton & Spelke, 2005). Finally, when a large set is labeled (e.g., as *five*) and one item is added, CP-knowers are more likely than subset-knowers to respond correctly when asked if there are now six or seven objects (Sarnecka & Gelman, 2004; Sarnecka & Carey, 2008). Although these tasks place varying processing demands on children, they all test the understanding of how to label sets with number words.

By some accounts, these differences are the product of a semantic induction that occurs at the moment that children become CP-knowers (Carey, 2009; Condry & Spelke, 2008; Le Corre & Carey, 2007; Piantadosi, Tenenbaum, & Goodman, in press; Sarnecka & Carey, 2008; Wynn, 1990, 1992). Under this view, when children become CP-knowers they do not merely learn another blind procedure – e.g., that the last word in a count is the ‘‘right’’ response to a request like ‘‘how many?’’ Instead, becoming a CP-knower involves making an observation about the semantic relation between numerals, and thus learning why counting generates correct responses. This view is clearly articulated by Sarnecka and Carey (2008)[1]:

> ‘‘The cardinal principle is often informally described as stating that the last numeral used in counting tells how many things are in the whole set. If we interpret this literally, then the cardinal principle is a procedural rule about counting and answering the question ‘how many.’ . . .Alternatively, the cardinal principle can be viewed as something more profound – a principle stating that a numeral’s cardinal meaning is determined by its ordinal position in the list. This means, for example, that the fifth numeral in any count list – spoken or written, in any language – must mean five. And the third numeral must mean three, and the ninety-eighth numeral must mean 98, and so on. If so, then knowing the cardinal principle means having some implicit knowledge of the successor function – some understanding that the cardinality for each numeral is generated by adding one to the cardinality for the previous numeral.’’ (p. 665)

Although it seems beyond doubt that children must eventually acquire knowledge of the successor function, it is less clear whether becoming a CP-knower is the moment at which this

[1] Similarly, according to Carey (2009), becoming a CP-knower involves making the induction that ‘‘if ‘x’ is followed by ‘y’ in the counting sequence, adding an individual to a set with cardinal value x results in a set with cardinal value y’’ (p. 327). Thus, becoming a CP-knower involves an induction based on a semantic observation, rather than the mastery of another procedure for labeling or generating sets.

occurs. Most studies that argue for a semantic induction do so either on the basis of children's ability to correctly generate sets by counting (using the Give-a-Number task), or by comparing large and possibly heterogeneous groups of CP-knowers to groups of children that are subset-knowers. The first form of evidence is problematic because children may merely have learned a procedure for using counting to generate sets without understanding how the procedure gets its content, and thus how number words are semantically related to one another. As previous studies have noted, prior to becoming CP-knowers children learn a procedure for correctly labeling sets by counting, which clearly lacks the content that it has for adults. When asked to report ''how many'' items are in a set, many children who are classified as subset-knowers correctly count and report the last word in their count, as though they know its meaning (Frye, Braisby, Lowe, Maroudas, & Nicholls, 1989; Fuson, 1988; Sarnecka & Carey, 2008). Thus, despite the appearance of having made an induction about how counting represents number, these children clearly lack knowledge of the cardinality principle. It is equally possible, therefore, that a similar situation is true in the case of the Give-a-Number task. Although children have clearly learned a procedure for generating sets via counting, the count list may still lack the semantic content of the adult number word system.

The second form of evidence – i.e., broad differences between subset-knowers and CP-knowers – is problematic because it leaves open the possibility that some CP-knowers actually do not understand the logic of counting at all, and that differences between groups are driven by more advanced CP-knowers (e.g., those more counting experience). For example, in their study of the successor principle, Sarnecka and Carey (2008) found that overall CP-knowers were more likely than subset-knowers to judge that adding one item to a set with *five* items resulted in *six* rather than *seven* items. Although CP-knowers as a group were significantly better at this task than subset-knowers (who almost always failed), there were many CP-knowers who performed randomly on the task. Further, the study did not test whether these children had actually performed an induction over the whole count list, but instead tested only very low numbers like *five* and *six*. Thus, the study left open (1) whether becoming a CP-knower is the moment at which children make a semantic induction about how counting represents number, and (2) whether the learning that occurs at this moment is truly an induction about number words as a class, or instead is restricted to a small set of highly familiar words with which children have extensive experience.

Here, we explored whether becoming a CP-knower by Wynn’s (1990, 1992) criteria involves a semantic induction. We contrasted two alternative possibilities. On the one hand, becoming a CP-knower may indeed involve making an inference about how all number words represent number. Children may realize that for any number word x, its successor y denotes a set of x + 1, and thus may be able to apply this knowledge to all number words in their count list. Apparent variability in children’s understanding of this principle may be due to error in the tasks that are used to assess knowledge, and thus may not be meaningfully related to children’s true expertise with counting. On the other hand, becoming a CP-knower may not involve a semantic induction; in some sense these children may not know the cardinal principle at all, but instead may have acquired another blind procedure for applying numerals to sets (in this case, generating sets using the counting procedure to satisfy requests like ‘‘Give me six fish’’). Learning this procedure may be one of several steps that children take to acquiring the successor principle and applying it to all known numbers. For example, learning to generate and label sets may be a precursor to discovering the relationships between numbers, rather than the opposite. On this view, differences in number knowledge should exist among CP-knowers, and should be predicted by other measures that assess their expertise with numerals. Also, children who have not made a semantic induction may perform differently for different numbers within their count list, rather than exhibiting knowledge for all numbers that they know.

To test these questions, we identified a large group of young CP-knowers (*N* = 84) using Wynn's criteria, and evaluated their semantic understanding of large numbers with four tasks. Two tasks tested children's understanding of the logical relationship between numbers. The first, the Unit task, has been used by previous studies to assess children's knowledge of the successor principle, and to provide support for the semantic induction hypothesis (e.g., Sarnecka & Carey, 2008). The second, the More task, measures knowledge of the ‘‘later greater'' principle – i.e., that words later in the count list denote larger quantities. The two remaining tasks test children's relative expertise and experience with numerals by testing (1) how high they can count, and (2) their ability to perform rapid estimates in the absence of counting, a skill that is known to emerge only after several months of experience as a CP-knower (Le Corre & Carey, 2007). These measures allowed us to ask whether experience using the count list was predictive of children's understanding of how numerals are logically related, or whether becoming a CP-knower is sufficient for making a semantic induction.

## Method

### Participants

Participants were 84 children aged 3;4 to 5;3 (M = 4;6) from the greater San Diego area. Fifty-six children were recruited using a departmental infant and toddler database and were tested in the UCSD laboratory, and 28 children were tested in local daycares. All but 4 of the 84 children spoke English as their first language. Of those four children, two were English/Spanish bilinguals, one learned Spanish as a first language, and one learned Arabic. Critically, all 84 children who were retained for analysis were CP-knowers in English. An additional 14 children were excluded from the discussion that follows: eight were not classified as CP-knowers based on the Give-a-Number task, three failed to complete the tasks, and we were unable to determine ‘‘mapper’’ level for the remaining three due to insufficient data (see the Fast Dots task, below).

**Procedures and Stimuli**

Each testing session lasted 20–25 min. The experimenter sat across from the child at a small table. Five tasks were administered to each child: (1) the Give-a-Number task; (2) the Fast Dots task; (3) the More task; (4) the Unit task; (5) a Counting task (to determine how high each child could count).[2] There were two task orders (1–2–3–4–5; 1– 2–4–3–5).

**Give-a-Number task.** The Give-a-Number task was adapted from Wynn (1992) and was used to test whether children were CP-knowers. Materials included a red plastic circle and 10 plastic fish. The experimenter asked the child, ‘‘Can you put *N* fish in the red circle and tell me when you’re done?’’ After each trial, the experimenter asked, ‘‘Is that *N* fish? Can you count them for me please?’’ If the child recognized that he/she made an error, the experimenter asked, ‘‘Can you fix it so there are *N* fish in the red circle, and then tell me when you’re done fixing it?’’

Testing began with the number four. On each trial, when the child successfully placed *N* fish in the red circle, the fish were removed and the child was subsequently asked to place $N + 1$ fish on the red circle. If the child failed to give *N* fish, the experimenter asked the child to place $N - 1$ fish on the red circle on the next trial. Children were deemed an *N* knower if they could give the correct quantity for *N* at least 2 of 3 times that *N* was requested. Children were considered a CP-knower if they could give a correct quantity 2 of 3 times for the numbers 4, 5, 6,

[2] Various versions of a one-to-one correspondence task were also piloted, but are not reported here. In general, CP-knowers performed very poorly on the piloted tasks.

7 and 8.

**Fast Dots task.** This task was modeled after Le Corre and Carey (2007) and examined children's ability to estimate small and large quantities. The task was selected to act as a proxy for children's experience using number words, and was used in analyses as a predictor of children's performance on the Unit and More tasks, described below. Stimuli were presented rapidly (1 s stimulus duration) on a laptop computer screen, one set at a time. The sets contained the following numbers of dots: 1, 2, 3, 5, 6, 8, 10, 12, 16, 20, 25, 30, or 51. To maintain interest in the task, the stimuli were presented to children as different kinds of objects: ‘‘apples’’ (green dots), ‘‘stars’’ (yellow dots), ‘‘raindrops’’ (blue dots), and ‘‘tomatoes’’ (red dots). The sets of objects were presented repeatedly in this order or in the reverse order; the order in which different numerosities appeared was randomized across trials. Each child saw a total of 42 trials. For half of the trials the amount of colored area remained constant across trials. For the other half of trials the size of the dots remained constant.

Children were told that dots would flash quickly on the screen, and that they should try to say how many dots they saw. The experimenter emphasized that the estimates did not have to be perfect, and demonstrated by guessing on four practice trials. For these trials, the experimenter said, ‘‘Hmm…I think that was maybe…thirty dots,’’ but always said the correct number. After the four practice trials, the experimenter said, ‘‘Good job! I think you're ready to play.’’ After each trial, the children's response was recorded before the next trial was flashed. Children's responses were capped at 100 to minimize the effect of outlier estimates on results (e.g., If a child estimated 230, the answer was counted as 100, although a note was made of the child's actual answer).

**More task.** This task tested whether children knew that later words in the count list represent larger quantities. Children were shown small toy gift boxes that differed in color. To begin, the experimenter explained, ‘‘We're going to play pretend with these presents. We're going to pretend that these boxes have stickers in them!’’ On each trial, two of the boxes were placed in front of the child. Then, the experimenter said, ‘‘This box has *N* stickers and this box has *M* stickers. Which box has more stickers?’’ while pointing to each box, respectively. After the child responded, the boxes were replaced with a new pair of boxes and a new trial began. A total of 12 trials were presented. Small quantities were: 5 vs. 6; 5 vs. 7; 6 vs. 9; and 8 vs. 9 (we also tested 2 vs. 4 and 3 vs. 8 as a method check, since these had numbers within the subset knower range). Large quantities were: 21 vs. 27; 22 vs. 25; 23 vs. 28; 23 vs. 24; 24 vs. 26; and 25 vs. 26.

**Unit task.** This task was adapted from Sarnecka and Carey (2008) to test children's knowledge of the successor principle. Specifically, we asked whether children know that adding one item to a set corresponds to moving up one word in the count list (e.g., adding one item to a set labeled as *five* results in *six*). We contrasted this with a change of two items, which corresponds to moving up two words in the count list (e.g., adding two items to *five* results in *seven*). Unlike Sarnecka and Carey, who tested children with only small numbers (e.g., 5 + 1) we assessed children's judgments for small (4, 5), medium (either 13 or 14, and 15), and large numbers (24, 25). Materials were a red covered box and a pile of multi-colored beads. To begin, the experimenter said, ‘‘I'm going to put some beads in the box, then I want you to tell me how many there are.’’ The experimenter proceeded to place a number of beads in the box while saying, ‘‘I am putting *N* beads in the box.’’ Then a memory check was performed, with the experimenter closing the lid of the box and asking, ‘‘How many beads are in the box?’’ If the child answered incorrectly, the experimenter said, ‘‘Oops! Let’s try this again!’’ and began the trial again. When the child gave the correct answer, the experimenter opened the lid and said, ‘‘Good! Now watch!’’ and then proceeded to add, one at a time, either 1 bead or 2 beads to the box. The experimenter then closed the lid and asked, ‘‘Now are there $N + 1$ or $N + 2$ beads in the box?’’

The first 19 children were tested with small and large numbers only: they saw two $N + 1$ trials and two $N + 2$ trials for both the small and large numbers, for a total of 8 trials. Medium numbers were added for the subsequent 19 children, who received two $N = 1$ and two $N = 2$ trials for the small, medium, and large numbers, for a total of 12 trials. Finally, for the remaining 46 children, the task was shortened by reducing the number of times each quantity was tested to one $N = 1$ and one $N = 2$ trial each for small, medium, and large numbers, for a total of 6 trials. Trials were randomized with the constraint that a particular transition was never repeated at different set sizes (e.g., no child was tested with both 5 + 1 and 25 + 1, although they did hear 5 + 1 and 25 + 2). Reported scores are always the average for each child, such that all children were weighted equally.

**Counting task.** This task tested children’s ability to recite the count list, and was used as a proxy for experience with numbers and counting. The experimenter told the child, ‘‘For this game, I want you to count as high as you can, and I’ll tell you when to stop.’’ The child was allowed two errors and was asked to stop after the second error. The experimenter recorded both errors and the highest number that the child could count to, which was defined as the highest number reached before the child’s second error. If two errors were not made by the time the child reached 100, the experimenter asked him/her to stop, ‘‘Great job! You can stop counting now.’’

## Results

Preliminary analyses revealed no significant difference between the children tested in the lab and those tested in preschools. Therefore, data from these groups were merged for the analyses that follow.

**Give-a-Number task.** Only children who were classified as CP-knowers were included in the study (*N* = 84).

**Counting task.** On the Counting task, children's highest count ranged from 9 to 100 with a mean of 36.6 (*SD* = 24.7). The median was 29, and the mode was 39. In the following analyses, we focus on how children interpreted words that were in their count list. For these analyses we divided the sample into low counters (children whose highest count was 10–19; *N* = 17[3]), medium counters (highest count 20–29; *N* = 28) and high counters (highest count equal or greater than 30, *N* = 36). This grouping ensured that the ''small numbers'' (6–9) in the following tasks were all within the counting range of the low counters (9–19, *M* = 14), the ''medium numbers'' (11–19) were all within the counting range of the medium counters (20–29, *M* = 25), and the ''large numbers''
(21–29) were all within the range of the high counters (30–100, *M* = 57).

Not surprisingly, highest count was positively correlated with age ($r(80) = 0.25$, $p < 0.001$). Average ages were as follows for each group: low counters (*M* = 4;4; Range = 3;4–5;3); medium counters (*M* = 4;5; Range = 3;6–5;3); high counters (*M* = 4;7; Range = 4;0–5;0). Three children were unable to complete the counting task due to time constraints, and thus their data were excluded from analyses that reference highest count.

---

[3] There was one child who counted only to nine, but we included his data with the low counters. Since no trials tested the number 'ten', all of the smallest numbers were still within his count list.

**Unit task.** Central to the understanding of the positive integers is the successor principle – e.g., that adding 1 item to 5 results in a set of 6. If becoming a CP-knower involves a semantic induction over the count list, then all CP-knowers should understand that for any word in their list, *n*, adding one item results in a set labeled by its successor, $n + 1$. Not only should this be true of all CP-knowers, but it should also be true for their performance on all numbers in their count list. To the extent that it is not, then the evidence for a semantic induction is significantly weakened.

We found two main results that draw into question the induction hypothesis. First, many CP-knowers, and especially less proficient counters, failed to show any knowledge of the successor principle (via the Unit task) even for very small numbers that were well within their counting range (e.g., 4–5). Naturally, we did not expect children to succeed at the Unit task for numbers outside their counting range, and consistent with this, one-sample *t*-tests showed that low counters' performance did not differ from chance (50%) for medium numbers ($t(11) = 1.92$, $p = 0.08$) or for large numbers ($t(16) = 0.61$, $p > 0.1$). Similarly, medium counters performance did not differ from chance for large numbers ($t(27) = 1.10$, $p > 0.1$).[4] However, contrary to the predictions of the semantic induction hypothesis, low counters and medium counters also performed poorly for numbers within their counting range. Fig. 1a–c shows the number of Low, Medium, and High Counters who responded correctly between 0% and 100% of the time for numbers within their counting range. As a group, low counters' performance did not differ from chance for small numbers ($t(16) = 1.16$, $p > 0.1$). As can be seen in Fig. 1a, the most common pattern for low counters was to answer only 50% of the trials correctly, which is also the pattern that would be expected by random guessing (recall that it was a binary forced choice task).

[4] Note that the degrees of freedom vary across trial types due to the addition of medium numbers after the first 19 subjects had already participated, as discussed above.

Although medium counters were somewhat better than chance as a group for small numbers ($t(27) = 2.92$, $p < 0.01$) and for medium numbers ($t(20) = 2.23$, $p < 0.05$), the modal pattern of responses was again to answer correctly on only 50% of the trials. Thus, among low and medium counters most children showed no sign of understanding the successor principle for even the lowest numbers. Higher counters performed much better overall: as a group, high counters performed significantly better than chance on small number trials ($t(35) = 8.00$, $p < 0.001$), and Fig. 1c shows that many high counters answered 100% of the small number trials correctly. High counters' performance was also better than chance for large numbers ($t(35) = 4.72$, $p < 0.001$) but not for the challenging medium numbers – i.e., the ‘‘teens’’ – which are irregular relative to the rest of the count list ($t(28) = 1.33$, $p > 0.1$). Still, even among high counters the most frequent outcome for medium and large numbers was 50% performance. Thus, although Sarnecka and Carey (2008) report that CP-knowers, as a group, performed better than chance on the Unit task, the more fine-grained analysis presented here reveals that many CP-knowers do not have knowledge of the successor principle for even the smallest numbers, and that even many high counters (who could count past 30) lack knowledge of the successor principle for numbers beyond 10. These data suggest that knowledge of the successor principle does not arise automatically from becoming a CP-knower, but that this semantic knowledge may be acquired later in development.

Our second main finding, already alluded to, was that children's performance differed significantly for numbers of different sizes within their counting range (Fig. 2). This point is clearest for the high counters (since the larger numbers tested were beyond the counting range of low and medium counters, and thus do not provide a meaningful test). For high counters, an analysis of variance with Trial Type (small vs. medium vs. large numbers) as a within subjects factor found a significant effect of Trial Type ($F(2,101) = 6.39$, $p < 0.01$). Although high counters could count to 57 on average (Range = 30–100; Mode = 39) they performed significantly better for small numbers than for either medium numbers ($t(28) = 2.70$, $p < 0.01$), or large numbers ($t(35) = 1.99$, $p < 0.05$).[5] This is important because it shows that, even for very skilled counters, knowledge of the successor relation differs significantly according to the particular numbers tested, even when the numbers are comfortably within the child's counting range. Previous studies, which tested only very small numbers, may have therefore overestimated the knowledge of these children, arguing for an induction over the child's count list when in fact their knowledge is restricted to only a subset of small numbers in this list.

It is important to note that when performance for all children and all numbers was combined, children's mean accuracy on the Unit task for small numbers was 72%. This overall result is comparable to the finding of Sarnecka and Carey (2008) who reported that CP-knowers performed at 67% accuracy as a group when making judgments for small numbers involving adding one to 4 or 5. Despite this, we do not believe that these data provide support for a semantic induction. Sarnecka and Carey (2008) argued that the Unit task tested a key component of becoming a CP-knower, since CP-knowers performed significantly higher than subset-knowers. However, many CP-knowers in their study did not exhibit knowledge of the successor

---

[5] The fact that the medium numbers showed lower accuracy than high numbers we attribute to the particular situation in English, mentioned briefly above, that the count list has regular transitions for small numbers (*four, five*) and large numbers (*twenty-four, twenty-five*), while medium numbers follow a different pattern (*fourteen, fifteen*).

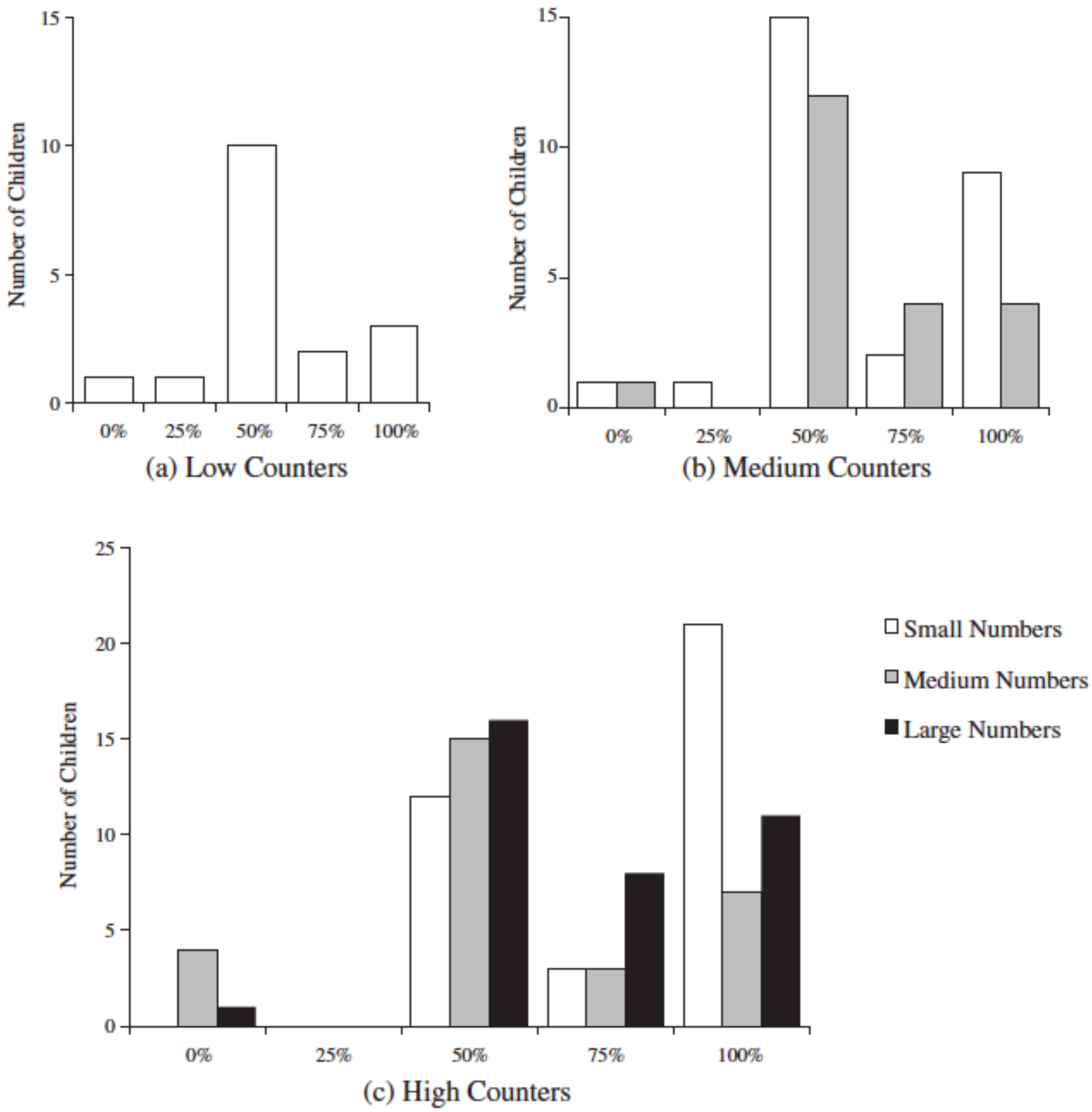


**Fig. 1.** (a–c) Number of low, medium, and high counters who responded correctly on 0–100% of the four trials in the Unit task for set sizes within their count range.

principle. In fact, only 16 out of 29 CP-knowers answered correctly on more than half of the trials, suggesting that many were randomly guessing (which may, in fact, have also produced some false positives who guessed correctly twice by chance). Overall, 11 of their children provided correct responses 75% of the time, and 5 reached 100% accuracy. Around the same number of children – 13 – chose correctly either 25% of the time (*N* = 1) or 50% of the time (*N* =

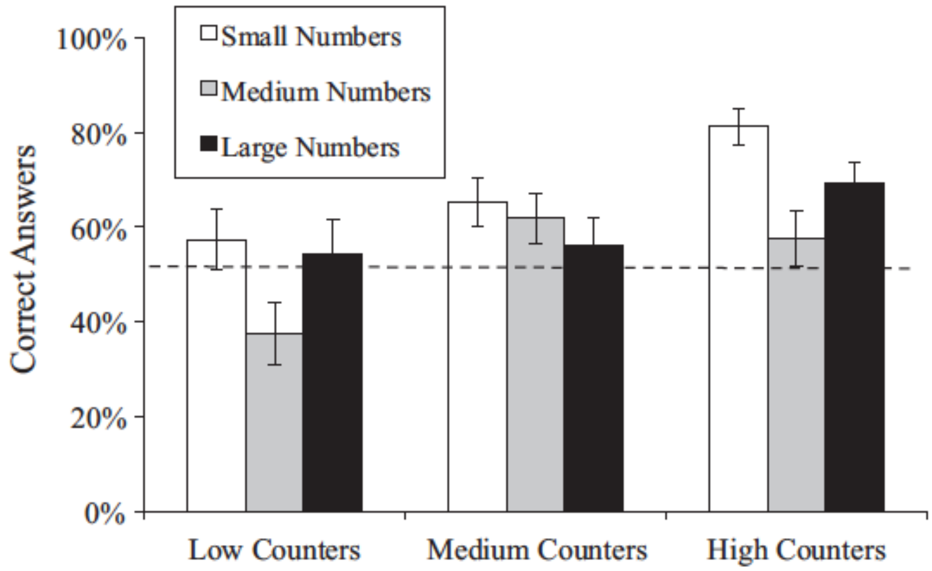


**Fig. 2**. Average percentage of correct answers by count group on the Unit task with small, medium, and large numbers.

12). Thus, Sarnecka and Carey may have overestimated the knowledge that results from becoming a CP-knower, since they did not explore whether differences between children were meaningfully related to experience counting (or simply noise in the task), and did not test numbers beyond five. Our data indicates that many CP-knowers perform quite poorly on the Unit task and that their performance is significantly correlated with counting expertise. Also, although more proficient counters perform better for small numbers, they have difficulty with larger numbers within their count list.

At the very least, these data show that the positive evidence for a semantic induction is much weaker than previously thought. To make the case for such an induction, stronger evidence would be required. However, the data presented here and previously may not rule out that such an induction takes place. Instead, it's possible that the tasks used here and in previous studies somehow mask children's competence. One possibility, for example, is that children may fail at the Unit task because they are unable to judge, for any number *N*, what its successor is in the count list. An analogy with the alphabet makes this clear – although adults rarely stumble in reciting the alphabet, it is more difficult to generate a randomly selected letter's successor – e.g., ''What comes after V?'' (for discussion, see Klahr, Chase, & Lovelace, 1983; Scharroo, Leeuwenberg, Stalmeier, & Vos, 1994).[6] Something like this ability is required for performing the Unit task – children must know not only how to count to *N*, but also to identify which of two alternatives is *N*'s successor even when not counting.

To ensure that children could identify a numeral's successor as required for the Unit task, we tested a separate group of 21 age-matched children (ages 3;6 to 5;4; *M* = 4;7) on a ''What Comes After *N*?'' task, which was identical in structure to the Unit task. All 21 children were classified as CP-knowers based on performance on the Give-a-Number task. Their highest count (*M* = 30.5; Range = 10–100) was determined as before and served as a basis for comparing children across the Unit and What Comes After N tasks (for the latter, we identified 10 low counters, 5 medium counters, and 6 high counters). To begin the task the experimenter said, ''Ok, now you're going to have to help me remember the order of those numbers [in the count list]!'' Then, for each number *N* that was tested in the Unit task – i.e., small (4, 5), medium (14, 15), and large numbers (24,25) – the experimenter asked, ''What number comes after *N*, *N* + 1 or *N* + 2?'' Each number was tested twice, and the alternatives were counterbalanced across trials, such that *N* + 1 was presented first once and *N* + 2 was presented first once. If performance on the What Comes After *N* task is comparable to performance on the Unit task, then children's difficulty with the Unit task might be attributed to an inability to access and reason about the ordering of numbers in the count list (i.e., when starting from an arbitrary position in the list, as

[6] We thank an anonymous reviewer for this observation.

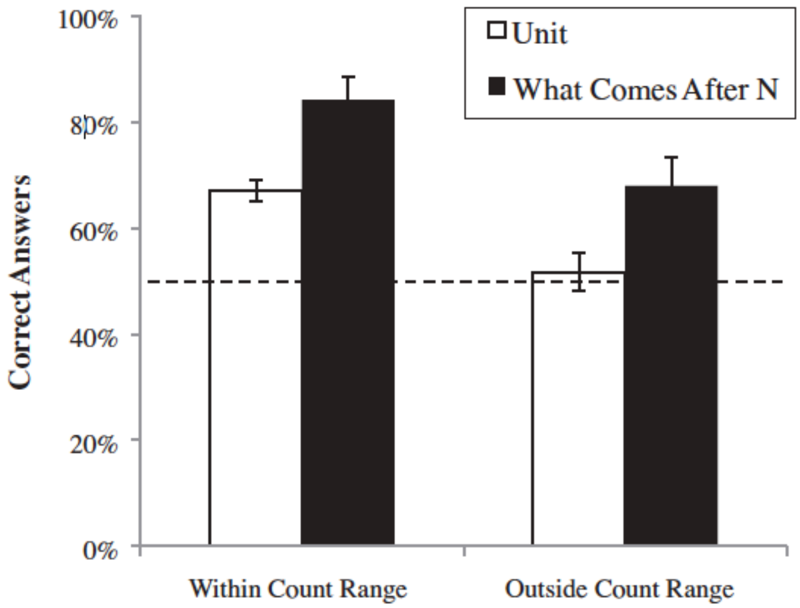


**Fig. 3**. Performance on the Unit and What Comes After *N* tasks for numbers that were within and outside children's counting ranges.

in the Unit task). Such a result would not remedy the current lack of evidence for a semantic induction, but would suggest that the Unit task cannot be used to show that an induction has not occurred. If, on the other hand, children are able to report the successor for a given numeral, then their poor performance on the Unit task would be best explained by a lack of semantic knowledge about what the ordering of numbers in the count list means for the quantities they represent (that one higher number in the count list represents one more in quantity).

Fig. 3 shows a comparison of the Unit task and the What Comes After *N* task for numbers that were either within or outside the counting range of individual children. To compare performance on these tasks, we conducted a weighted means ANOVA with Number Type (Within Counting Range vs. Outside Counting Range) as a within subjects factor and Task (Unit vs. What Comes After *N*) as a between subjects factor. There was a significant main effect of Number Type ($F(1,156) = 14.91$, $p < 0.0001$), indicating that, overall, children performed better for numbers within their counting range on both tasks. More important, there was also a highly significant main effect of Task ($F(1,286) = 16.73$, $p < 0.0001$), reflecting children's better performance on the What Comes After *N* task. On the What Comes After *N* task, children provided correct responses 84% of the time (*SD* = 23%) for numbers within their counting range, compared to only 67% (*SD* = 27%) on the Unit task. For numbers beyond their counting range, children also performed relatively well on the What Comes After *N* task, and responded correctly 68% of the time (*SD* = 28%) compared to only 52% of the time (*SD* = 29%) on the Unit task.[7] Critical to interpreting past reports, which tested only small numbers, we found no mean difference in performance for small numbers on the What Comes After *N* task across the low (90%), medium (95%), and high counters (92%), all $ps > 0.05$.[8] All were near ceiling in identifying the successors of small numbers. This is an important point, given that only high counters exhibit knowledge of the *semantic* relation between successors as measured by the Unit task. We conclude that CP-knowers' poor performance on the Unit task cannot be explained by a difficulty in identifying a numeral's successor from two alternatives. Instead, children's difficulty appears to be restricted to identifying how this successor relation relates to the semantics of number words.

*3.4. More task*

A second critical piece of knowledge that should accompany a semantic induction is the

---

[7] This relatively good performance outside children's counting range likely reflects that fact that counting is limited by decade transitions (e.g., what comes after 69?), whereas the What Comes After *N* task (and the Unit task) could be solved without this knowledge, by using the structure of the base 10 system to infer successors for even relatively large numbers (since decade transitions were not tested). For a related finding, see Barth, Starr, & Sullivan (2009).

[8] Consistent with this, a weighted means analysis of variance found a significant main effect of Task ($F(1,286) = 19.11$, $p < 0.001$) and Number Size ($F(2,156) = 6.52$, $p < 0.01$), but not Count Group ($F(2,156) = 2.16$, $p > 0.1$).

understanding that numbers later in the count list denote larger quantities. Not only does this later-greater principle flow naturally from the successor principle, but it is consistent with a more global recognition of the semantic ordering of numbers in the count list.

Overall, CP-knowers performed well on this task, and provided correct responses 75% of the time on average (where chance was 50%; see Fig. 4). A repeated measures analysis of variance with Trial Type (small vs. large numbers) as a within subjects factor and Count Group (low vs. medium vs. high counters) as a between subjects factor found a significant main effect of Trial Type ($F(1,156) = 6.86$, $p < 0.01$) and a significant main effect of Count Group ($F(2,156) = 7.84$, $p < 0.001$), but no interaction. As expected, low and medium counters did not differ from chance for large numbers, which were outside their count list, though performance approached significance for medium counters ($t(16) = 1.83$, $p = 0.09$ and $t(27) = 1.98$, $p = 0.06$, respectively). High counters performed significantly better than chance for both small ($t(35) = 12.12$, $p < 0.001$) and large numbers ($t(35) = 9.16$, $p < 0.001$), and performance for small and large numbers did not differ significantly ($t(36) = 1.07$, $p > 0.1$). Still, a somewhat surprising number of CP-knowers had difficulty judging which of two numbers was more, even when these numbers were very low in the count list (e.g., 5 vs. 6). For comparisons of small numbers – i.e., 5 vs. 6; 5 vs. 7, 6 vs. 9, and 8 vs. 9 – low counters made correct judgments only 72% of the time, medium counters did so 76% of the time, and high counters were slightly better at 85%. Overall, only 38 out of 84 total CP-knowers exhibited perfect performance for comparisons of small numbers. Although perfect performance is a high bar for developmental studies, the failures that did occur were for very small, frequent, and familiar number words, well within the counting range of all children tested. Critically, on the view that these children have performed a semantic induction over their count list and have used the ordering relation to make this induction, this is an

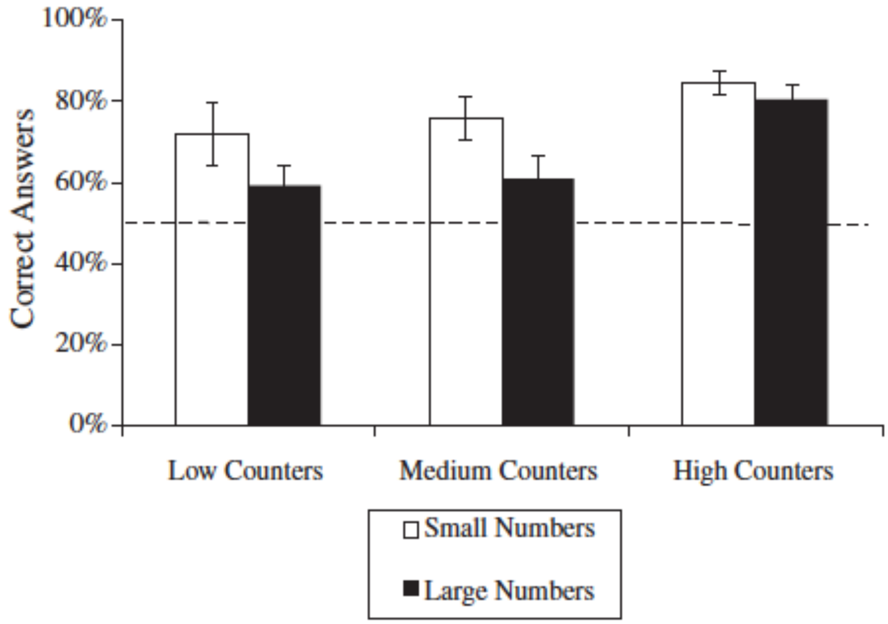


**Fig. 4.** Average percentage of correct answers by count group on the More task with small number trials and large number trials.

important failure. If children really knew the meanings of words like 5 and 7 on the basis of the successor principle, then we would expect them to *always* know that 7 is greater than 5.

*3.5. Fast Dots task*

Thus far we have probed children's logical knowledge of counting by examining their understanding of the later-greater and successor principles. We found substantial heterogeneity among CP-knowers and among set sizes as a function of counting ability. In the following analysis, we assessed additional knowledge that all children eventually acquire – a mapping of their count list to representations in the approximate number system (ANS), which tests a very different kind of number knowledge than the other tasks. Representations in the ANS are approximate, such that error in estimates increases as a function of the size of the quantity estimated (see Dehaene, 1997, for review). Previous reports have found that children's knowledge of how numerals are mapped to representations in the ANS, like their understanding of the logical principles described here, emerges gradually after children become CP-knowers (Barth et al., 2009; Le Corre & Carey, 2007; Lipton & Spelke, 2005). On the basis of such evidence, Le Corre and Carey (2007) argued that mappings to the ANS could not possibly drive a CP induction, since mappings were acquired after children become CP-knowers. Our data, however, draw this conclusion into question. If a semantic induction occurs sometime after children become CP-knowers (by Wynn's criteria) then children's understanding of counting could, in theory, rely on a mapping to the ANS. More generally, children's estimation ability offers an alternative to Highest Count as a proxy for number word experience. Thus, it also allowed us to explore whether performance on the Unit task was systematically correlated to other forms of learning that are thought to take place gradually after the CP transition.

Previous studies have documented that when adults are shown rapidly presented arrays of objects (e.g., dots), their mean numerical estimate of the set size is linearly related to the set's actual size (for review, see Dehaene, 1997). However, children do not initially exhibit such a linear mapping of number words to object arrays when estimating, and many recent CP-knowers do not exhibit linear estimates for even relatively small numbers like 6 through 10 (e.g., Le Corre & Carey, 2007). Following Le Corre and Carey (2007), we classified children individually as either ‘‘mappers’’ or ‘‘non-mappers’’ according to the estimates they produced for set sizes on the Fast Dots task. Children were categorized as ‘‘mappers’’ (indicating an ability to make generally higher estimates for larger numbers) if the slope of their responses was greater than 0.3 (Le Corre & Carey, 2007). All other children were categorized as ‘‘non-mappers’’. We calculated the slope for each child based on a linear regression of their estimates for each of the ten Fast Dot set sizes outside of the subitizing range (5, 6, 8, 10, 12, 16, 20, 25, 30, or 51). To do so, we used the formula $[\sum(x - x^1)(y - y^1)]/\sum(x - x^1)^2$, where *x* was the number of dots presented on the screen, *y* was the child's response, and $x^1$ and $y^1$ were the means of each sample. In our study, mappers (*N* = 61) had slopes ranging from 0.32 to 2.61, with a mean of 1.33 (*SD* = 0.70). The non-mappers (*N* = 23) had slopes ranged from – 0.26 to 0.26, with a mean of 0.07 (*SD* = 0.11). Thus, although the set sizes presented to children were perceptually quite distinct from each other, many CP-knowers failed to estimate larger quantities with higher number words.[9]

Estimation ability was related to children's logical understanding of number words as measured by both the Unit and More tasks (Fig. 5). First, mappers' performance on the Unit task was significantly better than nonmappers' ($t(82) = 2.03$, $p < 0.05$). Similarly, on the More task the performance of mappers was significantly better than performance of non-mappers ($t(82) = 3.20$, $p < 0.01$).

Although both counting and estimation abilities were related to performance on the Unit and More tasks, they were not significantly related to one another. Pairwise correlations for highest count, mapping, and the Unit and More tasks indicate that both highest count and mapper-hood were significantly correlated with performance on the Unit ($r(81) = 0.37$, $p < 0.001$ and $r(84) = 0.22$, $p < 0.05$, respectively) and More tasks ($r(81) = 0.32$, $p < 0.01$ and $r(84) = 0.33$, $p < 0.01$, respectively), but there was no significant correlation between mapping and highest count ($r(81) = 0.09$, $p > 0.1$). Performance on the Unit task and the More task were also highly correlated ($r(84) = 0.44$, $p < 0.0001$). Additionally, a linear regression predicting performance on the Unit task using age, highest count, and mapping ability, found that age was not a significant

[9] We also calculated children's slope for only 'small numbers' (5–10), comparable to the numbers 6–10 originally tested in Le Corre and Carey (2007). For this restricted set, 72/84 children were mappers (slope > 0.3), far more than the 40/71 found by Le Corre and Carey. This is perhaps because our children were slightly older (*M* = 4;5 compared to 3;11 in Le Corre and Carey), or because providing a wider range of numerosities changes the distribution of responses for the small number range.

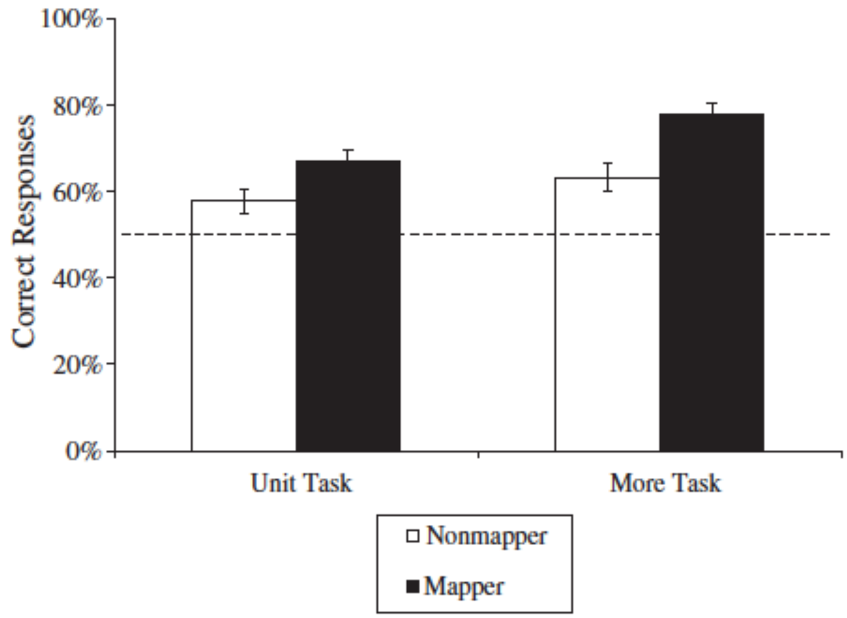


**Fig. 5.** Performance by mappers and nonmappers on the More task and the Unit task.

predictor ($\beta$ = 0.17, $p$ = 0.11) when highest count ($\beta$ = 0.31, $p < 0.01$) and mapping ability ($\beta$ = 0.27, $p < 0.05$) were also considered. Similar results were found for the More task. Again, highest count ($\beta$ = 0.27, $p < 0.05$) and mapping ability ($\beta$ = 0.28, $p < 0.01$) were significant predictors, whereas age was not ($\beta$ = 0.11, $p > 0.1$). Thus, when age, counting ability, and estimation ability were all considered, both counting and mapping were significant predictors of performance on the More task or the Unit task, whereas age was not. We take this to suggest that the development of the logical underpinnings of higher number knowledge – the successor principle and the later-greater principle – is advanced by increases in children's overall experience with numbers, in using the count list and in mapping to the ANS.

## General discussion

We began this investigation by asking how learning to count is related to acquiring the meanings of large number words. Previous studies have suggested that children proceed through stages in which they learn the meanings of *one, two, three,* and *four* in sequence, and then discover how counting represents number, at which time they become ‘‘cardinal principle knowers’’. Not only can these children correctly count sets and answer ‘‘how many’’, but they can generate and give a requested amount using the counting procedure. Our question was what these CP-knowers understand about number, and specifically, whether becoming a CP-knower by Wynn’s (1990, 1992) criteria involves making a semantic induction over all known numbers.

To address this question, we tested a large group of children who were all classified as CP-knowers by the Give-a-Number task. We asked whether children understood the logical relationship between their number words by testing their knowledge of the successor principle (via the Unit task) and of the later-greater principle (via the More task). We found that many CP-knowers appear to lack knowledge of these principles. The majority of CP-knowers showed no evidence of understanding the successor principle even for the very smallest numbers. This is despite the fact that young CP-knowers readily identify the successors of numbers in their count list, and even outside their counting range (as measured by the What Comes After *N* task). The only group in which a majority of children performed better than expected by chance was the high counters. These high counters, despite doing well for small numbers, performed poorly for larger numbers within their count list. Thus, not only was there variability among CP-knowers, but children exhibited very different behaviors for different numbers within their counting range. An assessment of the later-greater principle also revealed that many children were inconsistent in judging which of two small numbers (e.g., 5 vs. 6) denoted a larger quantity. In fact, fewer than half of the children in our study made such judgments correctly 100% of the time for the smallest, most familiar numbers. Although imperfect performance is not uncommon in developmental studies, these data are nonetheless surprising on the view that such knowledge is a prerequisite for becoming a CP-knower.

Critically, these results do not represent a failure to replicate previous studies. Our overall results for the Unit task are similar to those reported by Sarnecka and Carey (2008), and other recent studies have found that many CP-knowers have a poor understanding of ordinality, despite being classified as CP-knowers (e.g., Le Corre, in preparation; Sullivan & Barner, in preparation). However, as we have shown here, it is critical to consider differences between children, as well as differences between numbers of different sizes. By dividing children according to counting and estimation ability, we discovered important differences between children as a function of their overall familiarity with counting and using numbers. The least skilled counters knew little to nothing about the successor principle, for example, even for very low numbers like 5 and 6. Had we averaged the results of children with different counting abilities, we would have found, like previous studies, that as a group CP-knowers show better than chance knowledge of the successor principle (unlike subset knowers). However, by dividing children according to their counting ability we see that this success rate is driven mainly by very skilled counters. Also, whereas previous studies that argue for an induction tested only small numbers, like 5, we tested larger numbers too, and found important differences. Even very skilled counters failed to exhibit knowledge for larger numbers, despite the fact that these numbers were well below their highest count. Thus, the success of CP-knowers as a group was not only driven by skilled counters, but was driven mainly by their judgments for very small numbers.

In light of these data, our conclusion is that there is little evidence to suggest that becoming a CP-knower involves a semantic induction. Although previous studies report that older children (5-year-olds) do have an excellent command of the logic of numbers even for large numbers (Lipton & Spelke, 2005), we find that there is little reason to believe that this knowledge is acquired at the moment that children become CP-knowers. Assuming that children's understanding of the positive integers arises from an induction over early number word meanings (although see Rips, Asmuth, & Bloomfield, 2006; Rips, Bloomfield, & Asmuth, 2008, for discussion), it would appear that such an inference must occur sometime *after* children learn to correctly label and generate sets using the counting procedure, and that not just one inference, but several, are involved.

Upon becoming CP-knowers, children may know little more than a collection of procedures for how to use counting to label and generate sets. As noted in the Introduction, prior to becoming a CP-knower some children are able to provide a correct answer to the question, ‘‘How many?’’ despite being unable to give a correct amount when asked for a specific quantity (Frye et al., 1989; Fuson, 1988; see Sarnecka & Carey, 2008, for discussion). Accordingly, some researchers believe that this knowledge does not involve a semantic induction, but instead reflects a semantically inert procedure – e.g., repeat the last word in a count when asked ‘‘How many?’’ Similarly, we see no reason to believe that a semantic induction is required to explain children's behavior on the Give-a-Number task. Children may learn yet another way in which counting results in a correct response: when asked to ‘‘give *n*’’, the child learns to give whichever individuals are implicated in a count of *n*. Although such a routine provides content to the counting procedure to the onlooker (by placing number words in one-to-one correspondence with sets of the appropriate size), it in no way requires the type of semantic knowledge that researchers typically ascribe to CP-knowers (i.e., the successor relation).

If this suggestion is right, then the learning mechanism that causes children to become CP-knowers, by Wynn's (1990, 1992) definition, is not the same mechanism that leads them to induce logical knowledge like the successor principle. If this is correct, then theories that equate this knowledge of counting with a semantic induction cannot be right. For example, in their study of CP-knowers, Sarnecka and Carey (2008) suggested that ‘‘putting together the puzzle of how counting implements the successor function is indeed what turns a subset-knower into a cardinal-principle knower’’. (p. 673). Similarly, Piantadosi et al. (in press) presented a Bayesian computational model designed to explain the inductive process, which explicitly equates children’s implementation of the successor principle with becoming a CP-knower. However, if the semantic induction comes after children become CP-knowers, then such models cannot be right not only in their proposal for *when* such an induction takes place, but also with respect to the mechanism of change.

What our study leaves open is how the induction might work if it does occur after becoming a CP-knower. One possibility is that a semantic induction occurs later in development as a result of mapping number words to approximate magnitudes. In our study, we found that children’s estimation skills were significantly related to their understanding of both the successor principle and the later-greater principle. This point is potentially important, since previous studies (Le Corre & Carey, 2007) have argued that a mapping between the count list and the approximate number system could not support a semantic induction for the whole count list (e.g., of the successor principle). The basis for this argument was that children became CP-knowers *before* they learned to map number words to approximate magnitudes. Since it was assumed that the semantic induction had already occurred for these CP-knowers, a mapping to approximate magnitudes could not possibly explain the induction. Our results, however, suggest that the semantic induction may take place later. Therefore, at least on these grounds, it remains possible that the semantic induction is, in fact, driven by a mapping between the count list and the approximate number system (see Dehaene, 1997; Gallistel & Gelman, 1992, 2000; Wynn, 1998).

The problem with this hypothesis, we believe, is that it does not explain how such an induction might work, since the approximate number system lacks content like ‘‘successor’’ and cannot represent precise differences between sets. As noted by Carey (2009), ‘‘Although a child might use her analog magnitude system to discover that numerosity increases monotonically as one proceeds through the count list, nothing in the analog magnitude system would appear to inform the child that each count increases numerosity by exactly 1.’’ (p. 313). If Carey is right, then the approximate number (i.e., analog magnitude) system could not support children’s induction of the successor principle, at least not directly. However, Carey’s point does not rule out a less direct role for the approximate number system. For example, if children used mappings to learn the later-greater principle, they might be in a position to generalize this knowledge to all words in their count list, an induction that would bring children one step closer to the successor principle. By this account, after becoming CP-knowers by Wynn’s criteria, children could begin forming associations between the approximate magnitudes of counted sets and their labels, as generated by counting. In doing so, they could learn that words later in the count list generally denote larger quantities, without understanding how the logic of counting guarantees this. Having learned these associative mappings between numbers and approximate magnitudes, children would then be in a position to generalize the later-greater principle to all words in their count list, via an induction. The only missing knowledge, on this view, would be the realization that the difference between any two successive integers is exactly 1.

Although this account is possible, there are reasons to doubt it, and to suspect that estimation abilities depend on knowledge of the later-greater principle, rather than being responsible for it. One reason to doubt that associative mappings drive learning about later-greater is that even adults appear to have relatively weak associations between large number words and approximate magnitudes. For example, when shown two sets that differ by a ratio of 2:1 (e.g., 75 vs. 150), adults discriminate these sets with nearly 100% accuracy, but have difficulty mapping a provided number word (e.g., 150) to the correct set (Sullivan & Barner, 2010). Also, when adults are provided with misleading feedback about the largest set that they will see in an estimation experiment, they adjust their mean estimates for almost all quantities that they see, but the linearity of their responses remains unaffected (with the exception of very small numbers, which appear to be resilient to miscalibration; see Izard & Dehaene, 2008; Sullivan & Barner, 2010). These data suggest that accurate estimation may rely upon a malleable structure mapping between the count list and states of the approximate number system, and not on associative mappings between the two. To deploy this mapping during an estimation task, adults likely rely on the later-greater principle, plus their knowledge of how they mapped number words to sets on previous trials. For example, upon making an estimate of *fifty* for a set, a mature estimator may infer that any larger set should receive a larger estimate (e.g., for a quantity that is 50% bigger they should generate a number word that is 50% larger). Mappings between number words and approximate magnitudes may thus rely on inferences based on knowledge of the ordering of number words.

If performing estimates relies on a structure mapping, then becoming a mapper may depend on first learning that, at least for some numbers, words that are later in the count list denote greater quantities. Learning this fact about a subset of small numbers (e.g., 3 and 4) might eventually support an induction whereby the later-greater principle is applied to all numbers. This new knowledge would not only facilitate estimation, but also lay the groundwork for a final induction regarding the precise difference in quantity denoted by successive number words (i.e., exactly 1). The data presented here indicate that many children who are mappers and who understand the later-greater principle still do not exhibit knowledge of the successor principle. This suggests that this knowledge – that successive numbers differ by a cardinality of 1 – may be the last piece to emerge in a sequence of learning events.

Clearly our study does not address how such inductive inferences might work. However, the study does provide evidence that, when they learn the successor principle, children may have much more knowledge to draw on than previously suspected, including an ability to generate sets using the counting procedure – a substantial source of evidence for learning about relations between numbers. More importantly, we find little evidence to support the idea that children make a sweeping semantic induction over their count list when they become CP-knowers by Wynn's (1990, 1992) criteria. Future studies should explore the relationship between later acquired knowledge – like the later-greater principle, mappings to approximate magnitudes, and the successor function – to determine how these pieces of knowledge might be related and whether multiple learning events, rather than a single inductive leap, might drive their understanding of how counting represents number.

**Acknowledgements**

Thank you to Noah Goodman, Steve Piantadosi, Mike Frank, Mathieu Le Corre, Mark Appelbaum, Jessica Sullivan and three anonymous reviewers for their comments on this manuscript, and to the staff and families at the Ruben H. Fleet Science Center for their collaboration. This project was supported in part by a James S. McDonnell Foundation Scholar Award to D.B.

**References**

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

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

Condry, K., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General, 137*(1), 22–38.

Dehaene, S. (1997). *The number sense*. New York, NY: Oxford University Press.

Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development, 60*(5), 1158–1171.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43–74.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*(2), 59–65.

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

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*, 1221–1247.

Klahr, D., Chase, W., & Lovelace, E. (1983). Structure and process in alphabetic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 462–477.

Le Corre, M. (in preparation). Verbal counting becomes a tool for linking number words to cardinalities before it becomes a tool for numerical comparisons.

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. A., Brannon, E., & Carey, S. (2006). Revisiting the performance/competence debate in the acquisition of counting as a representation of the positive integers. *Cognitive Psychology, 52*(2), 130–169.

Lipton, J., & Spelke, E. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76*, 978–988.

Piantadosi, S., Tenenbaum, J., & Goodman, N. (in press). Bootstrapping in a language of thought: A formal model of numerical concept learning.

Rips, L., Asmuth, J., & Bloomfield, A. (2006). Giving the boot to the bootstrap: How not to learn the natural numbers. *Cognition, 101*, B51–B60.

Rips, L., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences, 31*, 632–642.

Sarnecka, B., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*, 662–674.

Sarnecka, B., & Gelman, S. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Coginition, 92*, 329–352.

Scharroo, K., Leeuwenberg, E., Stalmeier, P., & Vos, P. (1994). Alphabetic search: Comment on Klahr, Chase, and Lovelace (1983). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 236–244.

Sullivan, J., & Barner, D. (2010). Mapping number words to approximate magnitudes: Item-based associative learning or structure mapping? *In Proceedings of the 32nd annual conference of the cognitive science society.* Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

Sullivan, J., & Barner, D. (in preparation). Children's use of structure mapping in numerical estimation.

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36,* 155–193.

Wynn, K. (1992). Children's acquisition of number words and the counting system. *Cognitive Psychology, 24*, 220–251.

Wynn, K. (1998). Psychological foundations of number: Numerical competence in human infants. *Trends in Cognitive Sciences, 2,* 296–303.