Available online at www.sciencedirect.com
SCIENCE DIRECT®

Cognitive Psychology 52 (2006) 130–169

Cognitive Psychology

www.elsevier.com/locate/cogpsych

Re-visiting the competence/performance debate in the acquisition of the counting principles

Mathieu Le Corre [a,*], Gretchen Van de Walle [b]
Elizabeth M. Brannon [c], Susan Carey [a]

[a] *Harvard University, USA*
[b] *Rutgers University, USA*
[c] *Duke University, USA*

Accepted 28 July 2005
Available online 20 December 2005

Abstract

Advocates of the "continuity hypothesis" have argued that innate non-verbal counting principles guide the acquisition of the verbal count list (Gelman & Gallistel, 1978). Some studies have supported this hypothesis, but others have suggested that the counting principles must be constructed anew by each child. Defenders of the continuity hypothesis have argued that the studies that failed to support it obscured children's understanding of counting by making excessive demands on their fragile counting skills. We evaluated this claim by testing two-, three-, and four-year-olds both on "easy" tasks that have supported continuity and "hard" tasks that have argued against it. A few noteworthy exceptions notwithstanding, children who failed to show that they understood counting on the hard tasks also failed on the easy tasks. Therefore, our results are consistent with a growing body of evidence that shows that the count list as a representation of the positive integers transcends pre-verbal representations of number.

Keywords: Counting; Number words; Number; Cognitive development; Performance/competence

* Corresponding author. Present address: Laboratory for Developmental Studies, Shannon Hall, 25 Francis Av., Cambridge, MA 02138, USA.
E-mail address: lecorre@wjh.harvard.edu (M. Le Corre).

0010-0285/$ - see front matter
doi:10.1016/j.cogpsych.2005.07.002

1. Introduction

When Leopold Kronecker said "The integers were created by God; all else is man-made" (Weyl, 1949, p. 33) he was making a *metaphysical* claim. Yet, the remark inspires a natural position concerning the *cognitive foundations* of arithmetical thought. If we replace "God" with "evolution," the position would be that evolution provided us with the capacity to represent the positive integers, the *natural* numbers, and that the capacity to represent the rest of the arithmetical concepts, including the rest of the number concepts (rational, negative, 0, real, imaginary, etc.) was culturally constructed by human beings. On this interpretation of Kronecker's remark, he espoused the "continuity hypothesis" with respect to integer representations; i.e., he believed that these representations are available throughout human development, historically and ontogenetically.

In a series of important publications, Gelman and colleagues (Cordes & Gelman, 2005; Gallistel & Gelman, 1992; Gelman, 1993; Gelman & Gallistel, 1978) have also put forward the continuity hypothesis, arguing that representations of the positive integers are part of our innate cognitive endowment. In their view, there is an innate system of non-verbal symbols whose deployment conforms to three integer-constitutive counting principles: the stable order principle, the one-to-one correspondence principle, and the cardinal principle. The stable order principle captures the fact that the symbols are applied in a consistent order across counting episodes. One-to-one correspondence means that every individual is tagged with one and only one symbol and each symbol is applied to one and only one individual. Finally, the cardinal principle entails that the last symbol of a count represents the number of individuals enumerated during the count.

Gelman and Gallistel (1978) further suggest that these non-verbal counting principles guide children's acquisition of the verbal count list. *"One could argue that skill in reciting count-word sequences precedes and forms a basis for the induction of counting principles. We, however, advance the opposite thesis: A knowledge of counting principles forms the basis for the acquisition of counting skills" (p. 204).* To support this proposal, they observed that children are able to generate counting strategies that differ from the familiar left-to-right sequence that parents typically employ in counting activities, suggesting that children are not simply executing a meaningless routine. Further, some young children use an idiosyncratic count list following a stable order (e.g., "one, three, four, six..." always in this order), and others occasionally, if rarely, use the wrong type of ordered list, usually the alphabet, as their count list. Gelman and Gallistel liken these behaviors to children's over-generalization of linguistic rules (e.g., "I goed to the store"). That is, children are not merely passively reproducing the patterns produced by their caregivers, but are actually *interpreting* the input in terms of rich innate mental structures. When children establish the numerical relevance of a memorized list that generally has a stable order, they do so because they have interpreted their count list as the verbal instantiation of the non-verbal counting principles.

A substantial body of research is at odds with the continuity hypothesis. Schaeffer, Eggleston, and Scott (1974), for example, demonstrated that many children who could successfully count arrays of between 5 and 7 objects nonetheless failed to provide a cardinal response when asked how many objects there were. These children occasionally recognized and constructed small arrays of 1–4 objects successfully, but did so without counting. Similarly, Fuson and colleagues (Fuson, 1988; Fuson & Hall, 1983; Fuson, Lyons, Pergament, & Hall, 1988) found that even when children provide the last number of a count in response to a cardinal query, they often seem to be following a "last word rule" rather than demonstrating true understanding of the cardinal principle. For example, children at times provide the last word following a count that has grossly violated stable order and/or one-to-one correspondence. They also at times relate the last word to the last individual counted rather than to the entire set.

Wynn (1990, 1992) also found that for a period of about a year after they had first memorized a short count list, children (1) only knew the exact meaning of a *subset* of the number words in their count list; (2) rarely used counting to solve tasks that required them to determine the cardinality of a set; and (3) often blatantly violated the counting principles when they counted. Measures from two different tasks supported these generalizations. In Wynn's "Give a Number" task (GN), children were given a bowl full of objects and were asked to give the experimenter various numbers of them. In the Point-to-X task, children were shown two cards depicting objects (e.g., sheep), one with N objects and the other with $N + 1$. Children were asked, "Can you show me the N sheep?" Both tasks showed that children first learned the meaning of "one", then that of "two", then that of "three". For example, "one"-knowers[1] could only reliably give the correct number of objects when asked for "one"; they gave more than one when larger numbers were requested, but the numbers they gave were not systematically related to those requested (Wynn, 1990). These very children only succeeded on the Point-to-X task if one of the two cards showed one object. If both showed more than one, they chose at chance. Similarly, "three"-knowers reliably gave only "one", "two", and "three" and failed on the Point-to-X task if both cards showed more than three objects. However, when children understood "four", their learning of the number words appeared to change radically: Wynn found no children who knew "four" who did not also know the exact meaning of the other number words in their count list, suggesting that children learned how their count list represents number when they learned the meaning of "four". We will henceforth refer to children who understand how counting represents number as "CP-knowers" (where "CP" stands for "cardinal principle") and will refer to "one"-, "two"-, and "three"-knowers as "subset-knowers" because they know the exact cardinal meanings of only a subset of the numerals in their count list.

Wynn identified this sequence in both cross-sectional (Wynn, 1990) and longitudinal (Wynn, 1992) studies. Further, both data sets suggest that this developmental sequence occurs over a highly extended timeframe. Wynn's cross-sectional data showed that children are subset-knowers until about age $3\frac{1}{2}$ (range 2;11–4;0) and her longitudinal data showed that, on average, 4–5 months elapse between each "stage" such that about a year elapses between the time at which children are "one"-knowers and the time at which they become CP-knowers.

Several measures suggested subset-knowers had a qualitatively different understanding of counting from CP-knowers. First, subset-knowers were much less likely to use counting in tasks requiring the exact determination of the cardinality of a set. Second, when asked, "How many X's are there?" (the "How Many?" task) after counting a set of objects, subset-knowers reported the last word of their count only about 20% of the time, whereas CP-knowers did so about 70% of the time. CP-knowers were also three times more likely to

[1] Henceforth, we will use the expression "N-knower" to refer to a child who only knows a subset of the number words (e.g., a "two"-knower is a child who only knows the adult meaning of "one" and "two"), based on Wynn's GN task.

repeat the last word of their count following a correct count than an incorrect count (84% vs. 28%). Subset-knowers appeared insensitive to the accuracy of their counting, providing a last word response equally often following correct and incorrect counts (Wynn, 1990). Finally, subset-knowers sometimes violated the cardinal and stable order principles. When these children did use counting to assemble a set, their count ended at a number other than the requested number about 50% of the time (Wynn, 1992). In the GN task, these children at times simply declared their response to be correct even when their own count had just shown the number was incorrect. They also at times "fixed" the set by altering the count sequence so that the last word was the number requested (e.g., "1, 2, 5" or "3, 3, 3, 3"). These behaviors were almost never observed in CP-knowers.

In sum, Wynn's data suggest that children memorize a part of their community's count list and establish the numerical relevance of the numerals in the list (e.g., learn the meaning of "one") almost a year before they learn how to use the list to determine the cardinality of sets. Thus, Wynn's data suggest that the acquisition of the verbal count list may involve the construction of a system of representation that is not innately available. Many have espoused this interpretation, although the exact nature of the process whereby the new representational system is constructed is under much debate (e.g., Bialystok & Codd, 1997; Briars & Siegler, 1984; Carey, 2004; Cooper, 1984; Fuson, 1988; Karmiloff-Smith, 1992; Klahr & Wallace, 1976; Mix, Huttenlocher, & Levine, 2002; Schaeffer et al., 1974; Spelke & Tsivkin, 2001; Starkey & Cooper, 1995; Strauss & Curtis, 1984). Because these proposals hold that children's representational resources undergo a drastic, qualitative change when they acquire the counting principles, we will refer to them (collectively) as the "discontinuity hypothesis."

It may be, however, that children's failures on tasks assessing their understanding of counting do *not* indicate a developmental discontinuity (Cordes & Gelman, 2005; Gelman, 1993; Gelman & Greeno, 1989; Greeno, Riley, & Gelman, 1984). Rather, children may possess an innate understanding of the counting principles but nonetheless perform poorly on tasks testing this understanding because of their limited abilities to understand what they are being asked to do (what Greeno et al., 1984 call "utilization skills") and/or to plan and execute counting procedures that will successfully meet the requirements of the task (what Greeno et al., 1984 call "procedural skills").

Indeed, many of the tasks used to assess children's understanding of counting arguably make excessive performance demands. For instance, Gelman (1993) argues that whereas the How Many? task presents a familiar means of eliciting the count list, it is a confusing means of testing knowledge of the cardinal principle. To demonstrate cardinal knowledge on this task, children must use the last word of their count to describe the set they just counted. While some children do so spontaneously, many count without producing a cardinal answer (e.g., they say "1, 2, 3, 4" instead of "1, 2, 3, 4. Four duckies!"). To elicit a cardinal answer, the latter are asked again, "So how many Xs are there?" Conversational pragmatics may lead children to infer that they had somehow erred in their first response (or why would the experimenter ask again for the same information just provided?) and should therefore repeat the counting procedure (see e.g., Freeman, Antonucci, & Lewis, 2000; Fuson, 1988; Gelman, 1993; Wynn, 1992). Children who have poorer utilization and procedural skills may be more uncertain of their count. Thus, they may be more likely to interpret this question as a challenge to the correctness of their count and so respond by recounting. These same children may be less aware of whether their count was in fact accurate, leading them to provide last word responses as often for correct counts as for incorrect ones (Gelman & Meck, 1986; Gelman, Meck, & Merkin, 1986).

The other two tasks designed to test the cardinal principle—Point-to-X and GN—are difficult because they require children to use counting to find or construct a set with some previously specified cardinal value. For example, in GN, children are given a target (e.g., "Can you give me five dinosaurs?") and then must remember this target while they use counting to assemble an array (see Frye, Braisby, Lowe, Maroudas, and Nicholls, 1989 for an argument that this poses problems for young children). This may be more difficult than deducing a cardinal value from a previous count. Also, each of these tasks requires sophisticated counting strategies that may exceed the utilization and procedural skills of younger children. To succeed in Point-to-X, children need to count each set, compare the obtained cardinal values to that requested, and then select the card depicting the requested cardinal value. In GN, children must coordinate counting with set construction, a process that may be too taxing for poorer counters (Cordes & Gelman, 2005; Fuson, 1988). Moreover, younger children may be unable to fix incorrect sets not because they are unaware of their error but rather because they are unable to implement some appropriate addition or subtraction strategy. Thus, these children may opt to extricate themselves from situations they cannot resolve by simply asserting that their response was correct or by adopting a strategy they know is incorrect but that results in the correct last word (e.g., having given three objects when five were requested, they may count the set as "5-5-5" or "1-2-5").

Finally, in their "videotape counting study", Gelman and Gallistel (1978) noted that children with poor counting skills were less likely to engage in counting large sets than children with better counting skills. Assuming that subset-knowers are really CP-knowers with poor counting skills, this could explain why one of the main differences between subset-knowers and CP-knowers is that the former are much less likely to count on the GN and Point-to-X tasks. It could also explain why subset-knowers only seem to know the meaning of "one," "two," and "three." While counting is the only way of determining the exact cardinality of large sets, that of small sets can be determined without counting by subitizing. Therefore, given that even 2-year-olds can subitize (Starkey & Cooper, 1995), children who are reluctant to count could succeed on the small number trials of the GN and Point-to-X tasks by relying on subitizing, but would fail on larger number trials because they require counting.

If younger children have failed to show their understanding of counting on tasks such as GN and Point-to-X because of performance limitations, it should be possible to reveal their understanding of counting by testing them on tasks that reduce demands on their procedural and utilization skills. Such evidence is in fact available. For example, when children are simply asked to assess the correctness of a puppet's counting and the resulting cardinal responses, they often perform quite well (Gelman & Meck, 1983; Gelman et al., 1986; but see Briars & Siegler, 1984; Frye et al., 1989). Gelman, Meck, and Merkin demonstrated, for example, that children as young as 3 years of age appreciated that a puppet who simply repeated the last word of a count in which it had skipped or double-counted an object had incorrectly answered a "how many" question. Likewise, they appreciated that if a puppet had correctly counted a set, its cardinal answer was only correct if it matched the last word of the count. The same children further inferred that a puppet who counted the same set twice but arrived at different values each time must have made a mistake in at least one of its counts. Finally, children were better able to generate correct, non-canonical counting strategies (e.g., beginning in the middle of a row) when they were allowed either to work first on small sets or to work on the task for multiple trials (with no feedback).

Further evidence in support of the performance account comes from Gelman's (1993) "What's on This Card?" (WOC) task. In this task, children were presented with sets of cards depicting from one to seven stickers, and were simply asked, "What's on this card?" On the first card of each set, the experimenter modeled the desired response: e.g., "That's right! It's *one* bee." On all subsequent trials, children were probed to elicit both counting and production of a cardinal response. For example, if a child counted a set of four stickers without producing a cardinal answer (e.g., "one, two, three, four!"), the experimenter asked, "So, what's on this card?" or "How Many bees is that?" to elicit a cardinal answer. If a child responded with a cardinal value alone (e.g., "That's four bees!"), the experimenter elicited counting with a prompt such as, "Can you show me?" Also, whenever it was deemed necessary, the experimenter provided children with counting assistance by either pointing to each object and/or by saying "one" and pointing to the first object to initiate the count, and then pointing to the other objects one at a time.

According to Gelman, the WOC task minimizes performance factors in at least two ways. First, the question, "What's on this card?" capitalizes on children's affinity for and familiarity with kind labeling situations (as opposed to situations focused on cardinal values). Second, counting assistance reduces the procedural demands of counting by helping children keep track of counted and uncounted objects. This may also help children overcome a lack of confidence in their counting ability, and it may increase their awareness that counting is relevant to the task.

To determine whether WOC would reveal earlier numerical competence than Wynn had shown, Gelman only tested children who were younger than the average age of Wynn's CP-knowers (i.e., 3;6). She determined the number of children who both produced a correct count (allowing for one counting error) and a cardinal answer matching the last word of the count on at least half of the trials. She found that between 70 and 90% of her young 3-year-olds met this criterion for all numbers tested. Moreover, 60% of her 2.5-year-olds met the criterion up to 5, but only 30% met it for larger numbers. These results led Gelman to conclude that the WOC task provided a more sensitive measure of children's early counting competence in that it showed that even children as young as 2.5 years of age were CP-knowers.

Taken together, Gelman et al.'s "Counting Puppet" studies and the WOC study support the view that young children failed Wynn's tasks because of performance limitations, supporting the continuity hypothesis. However, for several reasons, these studies do not yet settle the continuity/discontinuity debate. First, none analyzed children's performance as a function of their Wynn stage. This is important because the age at which individual children achieve different Wynn stages is quite variable; e.g., while the average age at which children become CP-knowers is 3;6, some reach this stage at 2;11 (Wynn, 1990). Thus, even though Gelman's children were all younger than Wynn's average CP-knower, each age group may nonetheless have included some children who would have been CP-knowers on Wynn's tasks. Likewise, many of the 3-year-olds in the Counting Puppet studies could have been CP-knowers. Therefore, these studies may have overestimated the extent to which they conflict with results that support discontinuity. The studies presented here will address this problem directly. Experiment 1 will examine children's performance on WOC as a function of their GN stage, and Experiment 2 will examine children's performance on a Counting Puppet study as a function of GN stage. According to the performance account, children should succeed on WOC and the Counting Puppet task even if they are classified as subset-knowers on GN. On the other hand, according to the discontinuity hypothesis, only those children classified as CP-knowers on GN should show that they understand counting on the WOC and Counting Puppet tasks.

Second, while showing some understanding of the counting principles in old 2-year-olds and young 3-year-olds is impressive, it is not sufficient to show that children's use of verbal counting is guided by innate counting principles. Gelman and Gallistel's continuity hypothesis predicts that the counting principles will motivate children to actively seek an ordered list of linguistic symbols and to interpret this list in terms of the counting principles as soon as they find it and have established its relevance to number representations. Previous studies have shown that some children have memorized a count list and have established its numerical relevance (e.g., they have learned "one") by the time they are 24 months old (e.g., Le Corre & Van de Walle, 2001). Therefore, because their samples did not include children that were young enough, the WOC and Counting Puppet studies have not yet provided the critical piece of evidence: that children are CP-knowers as soon as they have memorized a count list and have established its numerical relevance. We will address this issue in Experiment 1 by including children who have just memorized their count list in our sample.

The last two problems with the evidence for the continuity hypothesis concern the WOC task itself. First, Gelman does not report the *mistakes* made by children who produced both cardinal responses and counts on at least 50% of trials. An analysis of mistakes children make could reveal whether children who meet the 50% criterion really did understand the counting principles. For example, one can imagine a child who counted on every trial but whose cardinal answers matched her counts only 50% of the time, other times saying things like "One, two, three. That's two bears." Such a child would have met Gelman's criterion, but clearly should not be considered a CP-knower. Experiment 1 will address this issue by determining whether children ever produce such violations of the cardinal principle, and whether their frequency is a function of their GN stage. Finding that children who are subset-knowers on GN commit these errors would support the discontinuity hypothesis.

Finally, Gelman (1993) argues that the fact that the large majority of her youngest children met her criterion for small numbers (i.e., 2 and 3) is enough to show that these children are CP-knowers. While these results are impressive, they do not necessarily imply knowledge of the counting principles. Indeed, Wynn's subset-knowers could give or identify small sets without counting (e.g., when "two"-knowers correctly created or identified sets of two objects, they did so without counting). This suggests that subset-knowers learn the meaning of "one," "two," and "three" by mapping these number words onto the outputs of a subitizing process that determines small cardinal values without counting (Carey, 2001, 2004; Klahr & Wallace, 1976; Siegler, 1991, 1998; Starkey & Cooper, 1995). Moreover, we have seen that subset-knowers do sometimes count correctly, particularly if they are counting small sets. Thus, "two"- and "three"-knowers could correctly count small sets and then produce cardinal answers matching their count (i.e., "two" or "three"), not because they understand the cardinal principle, but rather because these are the number words that map onto the cardinal value given by subitizing. To examine this possibility, all of the analyses of counting in Experiment 1 will treat small (1–3) and large (4–8) numbers separately. According to the continuity hypothesis, children should respect the cardinal principle on both small and large numbers. However, if young children's respect of the cardinal principle on small numbers reflects an accidental match of correct but numerically meaningless counting with subitizing, then they should commit many errors on large numbers.

In sum, developmental science has yet to determine whether the count-based representation of the natural numbers is the work of evolution or that of human culture. Each position has its advocates, and each position is supported by rich but ambiguous evidence. The present studies are the first to seek to advance this debate by analyzing within-child consistency on tasks that have been taken to support each hypothesis. As both positions predict that children who are classified as CP-knowers on GN will succeed on the easier tasks, the critical question concerns the performance of children identified on GN as subset-knowers. On the continuity account, even the least advanced subset-knowers (e.g., "one"-knowers) should show evidence of understanding counting on the easier tasks. In contrast, if the subset-knower stages are real, "one"-, "two"-, and "three"-knowers should still prove to be subset-knowers, and not CP-knowers, on the easier tasks.

2. Experiment 1

2.1. Method

2.1.1. Participants

Fifty 2- to 4-year-old children (mean: 3;1, range: 2;0 to 4;0) participated. All were fluent English speakers recruited in the New York City area. Participants were tested either at a university child development laboratory or at local day care centers or nursery schools. Participants were initially recruited by letter and phone calls through commercially available lists or by letters sent home by the day care centers. Parents who came to the laboratory received reimbursement for their travel expenses and a token gift for their child. The majority of the children were from middle-class backgrounds, and most were Caucasian although a small number of Asian, African American, and Hispanic children participated. An additional six children's data were discarded, five because they did not know the count list to "six" (mean age 2;7, range: 2;3–3;8) and one because he could not be classified into a Wynn stage on GN.

2.1.2. Stimuli

2.1.2.1. How Many?. Stimuli consisted of small toy animals (e.g., frogs, puppies, and whales) presented in a single row. All the toys within each trial were identical. A Big Bird puppet was used to introduce the task.

2.1.2.2. What's on this card? (WOC). Stimuli consisted of four sets of eight cards with sets of 1–8 stickers placed on them in one or two rows. Each set had a distinct color and sticker type (apples, bears, cows, and turkeys). For each set, the 1-card was always presented first, followed by the 2- and 3-cards, followed by the 4- and 5-cards, and finally by the 6-, 7-, and 8-cards. The order within each subset varied across the four sets of cards. The apple set was always presented first; the order of the remaining three sets was varied across children.

2.1.2.3. Give-a-number (GN). Three sets of 15 small plastic toys were used: fish, horses and dinosaurs. A Kermit puppet was used to request the toys.

2.1.3. Procedure

The order of the WOC and How Many? tasks was counterbalanced. The GN task was always conducted last. Testing for each child was completed in two sessions each of which typically lasted 20–30 min. All sessions were videotaped and later transcribed.

2.1.3.1. How Many?. The experimenter (E) first introduced each child to a Big Bird puppet, saying, "Big Bird has a problem. He forgot how to count! Could you help Big Bird count his toys?" The E then showed the child a tray of two, three, five, or six identical toys on a plastic tray, and the child was asked to count the toys for Big Bird. The E presented each set size twice in a pseudo-random order, and on one of the trials for each set size the child was probed after the count for a cardinal response with the phrase, "OK, so can you tell Big Bird how many toys he has here?" No feedback was given.

2.1.3.2. What's on This Card? (WOC). The method was modeled after Gelman (1993). For each set of 8 cards, the child was first shown the card with one object (e.g., one apple) and asked, "What's on this card?" The expected response was "an apple" or "apple." Regardless of the child's response the E responded, "That's right, that's *one* apple." The remainder of the procedure departed from Gelman's in the following ways. First, our children were only probed to produce both a count and a cardinal response three times per card set rather than on every trial. On probe trials, if the child had given a cardinal response (e.g., "two cows") the E asked, "Can you show me?" in an effort to elicit a count response. If the child had spontaneously counted without providing a cardinal response, the E asked, "So, what's on this card?" which was meant to elicit a cardinal response. Second, since we wanted to know whether children would independently infer that counting was relevant to the task we never employed "How Many?" as a count probe. Children often learn to produce a count sequence in response to "How Many?" as part of a social routine without having any idea of what the sequence means or why they should produce it (akin to children's early production of the ABCs; Durkin, 1993). For the same reason, if a child was reluctant to respond, we did not initiate counting by pointing to the first object and saying, "one...." Finally, if a child gave non-numerical responses (e.g., "apples" or "I do not know") to three consecutive cards, that set was terminated, and the E continued with the next set of cards.

2.1.3.3. Give-a-Number (GN). To begin, the E introduced the child to Kermit the Frog and said, "Kermit wants to play with a particular number of toys, and he's going to ask you for the number he wants. You see if you can help him!" The E placed a collection of toys in front of the child and made Kermit ask, "Could you give me *one* dinosaur? Put it right here, just *one* dinosaur!" The child was coaxed with phrases like, "he only wants X dinosaur(s)," and "can you pick out X toy(s) for Kermit?" When the child provided a set of any number, Kermit said things like, "Yay, I like to play with my dinosaurs!"

After the initial demonstration, the E proceeded to ask for up to six toys. A titration method modeled after Wynn (1992) was used whereby if the child succeeded at giving X dinosaurs, Kermit requested $X+1$ on the next trial. If the child then failed to give $X+1$ dinosaurs, X was requested on the subsequent trial. Children were tested only up to the number that they could give correctly at least two out of three times.

When children did not count when producing a set, they were asked, "*Can you count and make sure you gave Kermit X toys?*" regardless of the number they had given. If children counted and the last number of their count did not match the number of objects requested, the E then probed with, *"But Kermit wanted X dinosaurs—can you fix it so that there are X?"* If children did not count in fixing the set, they were asked to verify the numerosity given by counting once more, and they were asked to fix it if their count failed to match the number requested. Note that because children were allowed a single counting error, they could be credited with having given the correct number even when they had actually given $X \pm 1$.

2.2. Results

2.2.1. How Many?

Evidence that a given child does not know what number words mean, or how the count list represents number, would be of no interest if he or she does not know the words and the counting routine. We thus analyzed children's performance on How Many? to ensure that they had learned a count list. To be included in the sample, each child had to have memorized a stable count list containing at least six number word types. Because children often counted during WOC and GN, and because both of these tasks involved sets greater than six, the highest correct count each child produced often exceeded six. The averages are reported in Table 1. All children used the standard English count list. As Gelman and Gallistel (1978) documented, even our young 2-year-olds had learned a small count list.

2.2.2. Give a Number (GN)

Before we report our results, a few comments on terminology. Children who can give all numbers requested will be referred to as "CP-knowers", while children who only succeed at giving a subset of the numbers requested will be referred to as "subset-knowers" or as "*N*-knowers" (where *N* is the highest number on which they succeeded). We are aware that, insofar as the expressions "subset-knower" and "*N*-knower" imply that some children do not understand all number words in their count lists and thus that they do not understand counting, our terminology presupposes what we set out to determine. Despite this shortcoming, we decided to use these expressions because we could not find others that were theoretically neutral but nonetheless allowed for the easy integration of our results with extant literature. The reader should bear in mind that, despite this terminology, we will not come to a conclusion on the actual nature of subset-knowers' understanding of counting until we have assessed their performance on the WOC task and on the new Counting Puppet task presented in Experiment 2.

Table 1
Knower-level groups from Give a Number (GN)

Levels	*N*	Age[a]		Count list length[b]	
		Mean	Range	Mean	Range
0-knowers	7	2;7	2;1–3;2	8	6–10
"One"-knowers	7	2;6	2;1–3;11	8	6–12
"Two"-knowers	10	3;4	2;10–4;1	10.1	8–11
"Three"-/"four"-knowers[c]	11	3;1	2;7–3;11	8.8	6–11
CP-knowers	15	3;8	2;9–4;0	8.7	8–12

[a] Ages are in years and months (years;months).

[b] The count list length is the number of distinct number words in each child's count list. All children used lists that followed the standard adult order.

[c] This group comprised eight "three"-knowers (mean age = 3;2) and three "four"-knowers (mean age = 3;0).

We first established that we could place our young participants in the same stages Wynn found. Next, we asked whether the children who succeeded for all the numbers requested (designated CP-knowers) were indeed the only children who provided unambiguous evidence of understanding counting.

2.2.2.1. Highest number given. We adopted Wynn's (1990, 1992) criteria for determining the highest number of objects each child could give. To be considered to know the exact meaning of a number word "*N*", children had to:

(1) Give *N* objects at least 67% of the time when asked for that number.[2]
(2) Give *N* objects no more than half as often when asked for a different number.
(3) Satisfy conditions 1 and 2 for all numbers less than *N*.[3]

For example, if a child always gave two objects when asked for "two" but also gave two objects on most trials on which she was asked for other numbers, she would not be considered to know "two". Also, because of criterion (3), if a child met our criterion for "one" and "three" but not for "two", she would be deemed to succeed only on "one". Children who succeeded with all numbers requested (i.e., children who could give at least up to "six") were classified as CP-knowers.

Most of our children (47/50) fell into one of the five groups described by Wynn (0-, "one"-, "two"-, "three"-, and CP-knowers; see Table 1). The three children who were classified as "four"-knowers were the only ones who did not belong in any of the Wynn groups. Because there were so few of them, "four"-knowers will be combined with "three"-knowers in most of the analyses below. Table 1 shows that all of the children in Experiment 1, even the 0-knowers, could count at least to "six." Since "six" was the largest set ever requested, subset-knowers' failure to give the correct number of objects asked for cannot be due to their lack of knowledge of the lexical items themselves or by their counting range.

2.2.2.2. Spontaneous use of counting to produce sets. We first computed the percent of trials on which each child counted spontaneously, and then averaged these percentages for each knower-level. Because most children will count when prompted, responses to E's request to count and check were excluded. Trials in which small numbers (2 or 3) were requested and those in which large numbers (4 or more) were requested were analyzed separately. Fig. 1 illustrates means for small and large numbers for all knower-level groups except 0-knowers; 0-knowers were not included because they did not have any large number trials. Consider first small numbers. Parametric analyses revealed a main effect of

[2] In some cases, children were considered to have given the correct number even if the actual number of objects they had given was wrong. First, if they counted to give a set of objects or gave a set of objects and then counted them, they were allowed one counting mistake. Thus, a child could give 5 when asked for "four" and be considered correct if he had counted the set as 4. Second, on trials where they were given the opportunity to fix an incorrect number, children's fix was coded relative to their count. For example, if a child who had been asked for "five" gave 6 and counted them as 7, he was considered to have fixed the number correctly if he removed 2 objects.

[3] Only 4 of the 50 children were affected by this criterion. Two of them succeeded on "one" and "three" and were classified as "one"-knowers, one succeeded on "two" and "four" and was classified as a "two"-knower, and the last succeeded on "three" and "five" and was classified as a "three"-knower.

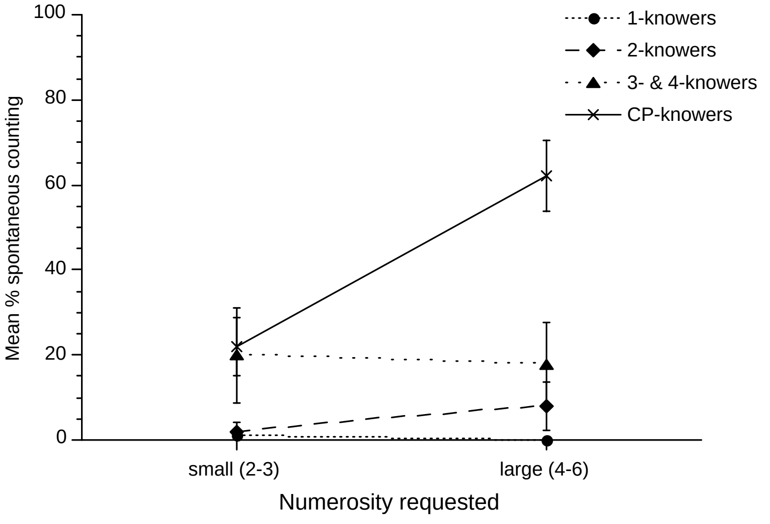

Fig. 1. Mean percentage of trials in which children spontaneously used counting to construct sets in the GN task as a function of set size requested and knower-level. Counting was deemed "spontaneous" if it occurred prior to probing by the experimenter on a particular trial; e.g., counting that occurred in response to the experimenter's request to "check and make sure this is *N*" was not included.

group, $F(4,45) = 2.54$, $p = 0.05$. This main effect stemmed from differences between the CP-knowers and "three"-/"four"-knowers who counted about 20% of the time when small numbers were requested, and 0-, "one"-, and "two"-knowers who only counted about 1% of the time. For large numbers,[4] we also found a main effect of group, $F(3,45) = 10.54$, $p < .001$. Post hoc Tukey HSD tests showed that CP-knowers spontaneously counted more frequently than all other groups ("one"-knowers, $p < .01$; "two"-knowers, $p < .001$; "three"-/"four"-knowers, $p < .005$). No other differences were significant.

An additional analysis investigated whether children in each group used counting more often to construct large sets than small ones. Zero- and "one"-knowers' data were excluded from this analysis since they never counted to construct large sets. A difference score (percentage of large sets constructed by counting minus percentage of small sets constructed by counting) was computed for each of the children in the remaining groups. CP-knowers were the only ones who were more likely to count on large number trials ("two"-knowers, mean difference = 5%, $t(9) = 1.47$, $p = .18$; "three"/-"four"-knowers, mean difference = −4%, $t(9) = .54$, $p = .60$; CP-knowers, mean difference = 40%, $t(14) = 5.98$, $p < .001$).

These data replicate Wynn's findings that subset-knowers do not spontaneously use counting to construct sets, not even for large ones. Of course, this analysis does not locate the source of the failure—perhaps, as Gelman (1993) suggests, the difference between the two groups is one of utilization or procedural skills. Perhaps the subset-knowers actually understand counting but rarely spontaneously use it to construct sets in this task because

[4] Zero-knowers were excluded from this analysis because only one was tested on large numbers. He never counted.

they are not aware that counting is a possible means of set construction or perhaps because the co-ordination of counting and set construction places too many processing demands on their fragile understanding of counting.

If the subset-knowers really do understand counting, their counts, however rare, should at least end at the target number as frequently as those of the CP-knowers. To determine whether this was the case, we computed the percentage of spontaneous counts ending at the target for each child, and then computed the average percentage for each knower-level. Because so few subset-knowers ever counted, we grouped all subset-knowers together. The group was mostly composed of "three"/"four"-knowers (5 out of 8). Almost all of the CP-knowers were included (13/15). The remaining two CP-knowers were not included because they never counted spontaneously; rather, they always put objects on the table without counting and then correctly fixed the sets when they were asked to do so (see section on fixing below).

CP-knowers were marginally more likely to end their spontaneous counts at the target number than subset-knowers (94% vs. 58%; $t(8) = 2.01$, $p = .08$). Inspection of the performance of individual subset-knowers suggests that the group's positive performance was exclusively due to the "three"-/"four"-knowers. The few "one"- and "two"-knowers who were included in this analysis *never* ended their counts at the target number, but the "three"-/"four"-knowers did so 93% of the time.

In sum, although the CP-knowers used counting to construct sets of all sizes, they were much more likely to count when they were asked for large numbers; when asked for small numbers, they often grabbed the correct number of toys without counting them. When counting, CP-knowers almost always stopped at the target number. In contrast, all of the subset-knowers predominantly used a non-count based strategy regardless of the size of the requested sets; in fact most of them *never* counted. Of the subset-knowers who did sometimes count, only those who were "three"-/"four"-knowers ended their counts at the target number; the few "one"- and "two"-knowers who sometimes counted never ended their counts at the target.

2.2.2.3. Fixing sets with incorrect numbers of objects. If children understand how counting represents number, then when their counting reveals that the set they constructed is not comprised of the requested number of objects, they should at least attempt to fix the set by adjusting it in the right direction. To test this prediction, we selected trials for which children initially gave the wrong number of objects (e.g., gave six objects when asked for four) but then correctly counted the set they had given. Then we determined whether children understood that their count revealed that they had given the wrong number by analyzing what they did when they were asked to fix their answer. In particular, we determined the probability that they would either fix the set in the wrong direction or leave it unchanged. Children who understand how counting represents number should rarely leave the number unchanged or change it in the wrong direction; in other words, they should *at least* change their answer in the right direction. Trials on which children gave an incorrect number of objects but did not count them correctly were not included because, under these conditions, children could have left incorrect sets unchanged because they doubted the accuracy of their counts.

Three "one"-knowers, 8 "two"-knowers, 8 "three"-/"four"-knowers, and 10 CP-knowers were included in this analysis. The excluded CP-knowers did not have any fix trials because they always gave the correct number on their own. In contrast, the excluded subset-knowers did not have fix trials either because they did not count in response to the E's request to check their answer or because they did not count correctly when they were asked to check. The mean percentages of incorrect sets left unchanged or corrected in the wrong direction are reported in Fig. 2. A one-way ANOVA showed a main effect of group on percentage of incorrect sets left unchanged or changed in wrong direction, $F(3, 25) = 5.45$, $p < .01$. Tukey post hoc tests showed that CP-knowers were significantly less likely to fail to fix correctly than "two"-knowers ($p < .01$) and "three"-/"four"-knowers ($p < .05$). That CP-knowers and "one"-knowers did not differ is likely due to the small number of "one"-knowers who contributed to this analysis, because the "one"-knowers did not differ from the other subset-knowers. The difference between CP-knowers and the subset-knowers reflects the fact that 70% of the CP-knowers (7/10) *always* fixed correctly, whereas about 60% of the subset-knowers (11/19) *never* fixed correctly. Subset-knowers' most common response was to leave the sets unchanged and state that they had given the right number.

2.2.3. Interim conclusions from GN analyses

Our initial sample of 56 children included younger participants than those of Wynn (1990, 1992) or Gelman (1993). Of those, five were not analyzed because they could not count up to "six", and seven were 0-knowers, suggesting that our sample indeed encom-

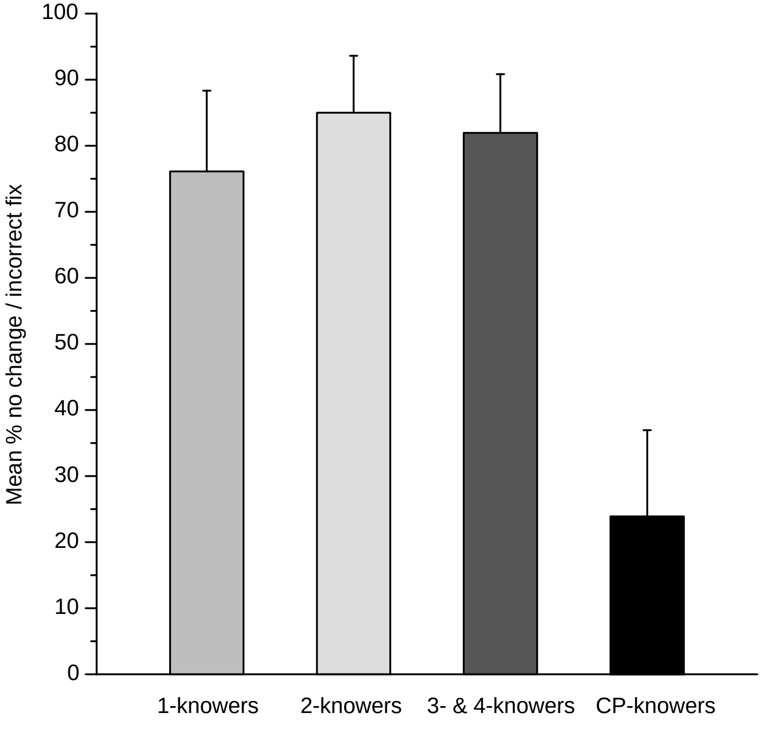

Fig. 2. Mean percentage of trials in which children left their answer unchanged or changed it in the wrong direction (e.g., adding objects when there already were too many) when they had given the wrong number of objects but correctly counted them so that the experimenter asked "But I wanted *N*! Can you fix it and make it *N*?"

passed the earliest steps of acquiring an integer list representation of number. The substantial number of 0-knowers among our sample is consistent with Fuson's (1988) claim that children learn the count list as part of a meaningless routine before they have assigned any numerical meaning to any of the number words. Our results were consistent with those of Wynn (1990, 1992). Like hers, the rest of our sample consisted of subset-knowers and CP-knowers. Subset-knowers could all recite the count list at least up to "six" but could not reliably construct sets of more than one, two, three, or four objects. CP-knowers could construct sets of all sizes requested, suggesting that, unlike subset-knowers, they could have constructed sets of any size represented in their count list. Subset-knowers almost always used non-count based strategies to construct sets (e.g., randomly grabbing a set), and their choice of strategy for producing sets was insensitive to the size of the requested set. When they did count, they did not use the information provided by their count to fix incorrect sets. In contrast, CP-knowers often used counting, particularly when they were asked for numbers that lay outside of the subitizing range, and they consistently used information provided by their counts to fix incorrect sets. One result is at odds with Wynn's. The five "three"- and "four"-knowers who used counting to construct sets were as likely as CP-knowers to stop counting when they had reached the target set size. Thus, GN may have underestimated understanding of the numerical meaning of counting in some of the more advanced subset-knowers.

The difference in the fixing abilities of subset-knowers and CP-knowers is particularly interesting because it demonstrates that the difference between the two groups was not just a matter of accuracy. Rather, CP-knowers were the only ones who always fixed sets in the right direction; subset-knowers not only left incorrect sets unchanged but also sometimes fixed sets in the *wrong* direction (e.g., added more objects to a set that already had too many). This result supports the discontinuity hypothesis, for it suggests that subset-knowers truly did not understand counting. Adherents of the continuity hypothesis might reply that these children may have known that the number they had given was wrong but failed to fix the set properly because they could not implement successful set transformation strategies.

If subset-knowers' knowledge of counting was indeed masked by procedural obstacles, the WOC task, a "count to cardinality" task that makes fewer processing demands on children, should reveal their knowledge. On the other hand, if the subset-knowers' failures were caused by the fact that they do not yet understand how counting represents number, then children should perform consistently on both tasks, despite the fact that WOC is an easier task.

2.2.4. What's on This Card? (WOC)

This task elicited a wide variety of response patterns. The three major response types were (1) noun phrases without number words (e.g., "apple" or "apples"); (2) cardinal answers (e.g., "three"; "three apples", or "one, two, three. That's three apples!"); and (3) counts without cardinal answers (e.g., "one, two, three, four" or "one, two, three, four apples"). Most children produced all three response types, though the dominant type was quite variable. Because this task is less familiar in the literature, Appendix A provides excerpts of interviews of three children, each one illustrating a particular performance type. Since our interest is in whether WOC will reveal that children's failures on GN reflect deficits in procedural or utilization skills, all of our analyses of WOC will be presented as a function of GN knower level. As in our analyses of GN, "three"- and "four"-knowers will be analyzed together.

2.2.4.1. Spontaneous counting. In both the GN task and the Point-to-X task CP-knowers switch from non-count based strategies to counting when they are presented with large set sizes, whereas subset-knowers rarely count, even when they are presented with large set sizes (Wynn, 1990, 1992; Experiment 1). As argued above, the lack of counting by subset-knowers on GN may have been due to excessive task demands. WOC requires only that children count a single, already-constructed set. Thus, perhaps even subset-knowers would be more likely to spontaneously count large than small sets on this less demanding task. We determined the number of times each child counted spontaneously (i.e., prior to probing) out of the total number of small (2 and 3) and large (4 to 8) number trials.

Fig. 3 depicts the mean percent of trials with spontaneous counting for each GN level. A 2×5 repeated measures ANOVA analyzed the effects of set size (small, large) and GN level (0-, "one"-, "two"-, "three"-/"four"-, CP-knower) on the percentage of trials in which children spontaneously counted on WOC. The ANOVA revealed a main effect of set size $F(1,45) = 33.6$, $p < .001$; overall, children were more likely to count spontaneously on large sets ($M = 51\%$) than small sets ($M = 20\%$). A marginal main effect of GN level, $F(4,45) = 2.53$, $p = .054$, revealed that the knower levels differed in their frequency of spontaneous counting (0-knowers = 14%, $SE = 9.1$; "one"-knowers = 34%, $SE = 9.1$; "two"-knowers = 34%, $SE = 7.1$; "three"-/"four"-knowers = 32%, $SE = 7.2$; and CP-knowers = 48%, $SE = 6.2$). Most importantly, there was a significant set size by knower-level interaction, $F(4,45) = 10.57$, $p < .001$. Simple effect analyses revealed that counting was more frequent on large sets than on small sets for CP-knowers and "three"-/"four"-knowers, $F(1,14) = 157.35$, $p < .001$ and $F(1,10) = 11.45$, $p < .01$, respectively. "One"- and "two"-knowers also showed this pattern, although weakly, $F(1,6) = 8.33$, $p < .05$, and $F(1,9) = 3.61$, $p < .10$, respectively. Only the 0-knowers failed to count more on the larger sets, $F(1,6) = 1.78$, $p = .23$.

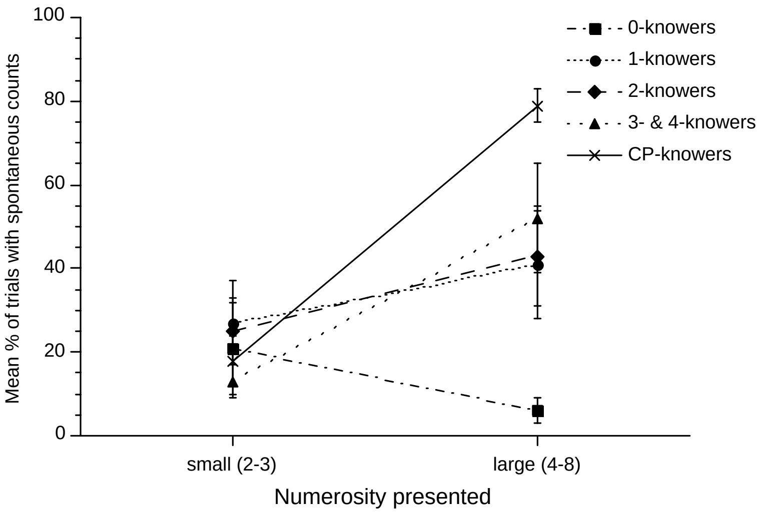

Fig. 3. Mean percentage of trials in which children spontaneously counted the sets presented to them in the WOC task as a function of the size of the presented set (small: 2–3; large: 4–8) and of GN knower-level. Counting was deemed "spontaneous" if it occurred prior to probing on a particular trial; e.g., counting that occurred after the experimenter asked "Can you show me?" was not included.

To assess whether the size of the difference in the frequency of counting on small and on large numbers was a function of GN knower-levels, we calculated the difference between the percent of counting on small and large trials for children in each GN level. An ANOVA revealed that this difference score varied as a function of GN level, $F(4,45) = 10.57$, $p < .001$. Post hoc comparisons revealed that CP-knowers (mean difference score (D) = 61%, $SE = 4.9$) were more sensitive to set size than "one"-knowers ($D = 15\%$, $SE = 5.1$, $p < .01$) and "two"-knowers ($D = 18\%$, $SE = 9.7$, $p < .01$). CP-knowers did not differ from "three"-/"four"-knowers ($D = 46\%$, $SE = 11.8$, $p = .31$). Finally, both CP- and "three"-/"four"-knowers differed from 0-knowers ($D = -15\%$, $SE = 11.2$, both p's $< .01$).

These results indicate that CP-knowers and subset-knowers produce the cardinal value of small sets without counting. Most importantly, although CP-knowers are most sensitive to set size in their spontaneous use of counting, both subset and CP-knowers counted more often on larger sets than smaller ones; only the 0-knowers were completely unaffected by set size. Unlike the Point-to-X and GN tasks, then, WOC suggests that subset-knowers indeed have some appreciation that while counting is not necessary to determine the cardinal value of small sets, one must count to determine the value of large sets. WOC may thus be a more sensitive measure of children's numerical knowledge than GN.

2.2.4.2. Respect of the cardinal principle. While the previous results show that WOC is a more sensitive measure of subset-knowers' propensity to *use* counting, they do not show whether subset-knowers *understand* what counting *means*. To determine whether children truly understand counting, one must examine their counting for evidence that they understand the cardinal principle. To do so, we selected all trials in which children produced a correct count (modulo at most one error in 1–1 correspondence) followed by a cardinal response (e.g., "one, two, three, four, five." So what's on this card? "Five apples," or "five"). Cardinal answers following incorrect counts were not included because, in such cases, children could fail to produce a cardinal answer that matches the last word of their count because they were not sure that their count was correct rather than because they did not understand the cardinal principle.

These data were analyzed in two steps. We first assessed the frequency of this structure (i.e., correct count followed by a cardinal response). For children who produced this structure at all, we then asked whether the cardinal answer matched the last word of the count. For each child, we calculated a "match score": the proportion of cardinal responses that matched the previous count out of all instances in which the child produced counts followed by cardinal answers. Because children could have large match scores on small numbers without understanding counting (see Section 1), we computed two match scores for each child, one for small numbers and one for large numbers.

The first result of note is that despite our use of probes aimed at getting children to complete their counts with cardinal responses, correct counting followed by a cardinal answer was relatively rare, particularly in subset-knowers. For *small* numbers, this structure was produced by only 3 of the 7 "one"-knowers, 2 of the 10 "two"-knowers, 3 of the 11 "three"-/"four"-knowers, and 5 of the 15 CP-knowers. For *large* numbers, no 0-knowers, three "one"-knowers, four "two"-knowers, and five "three"-/"four"-knowers produced at least one correct count followed by a cardinal response. In contrast, all fifteen CP-knowers did so. The subset-knowers who produced this structure did so an average of 1.4 times on small numbers and 1.3 times on large numbers. CP-knowers produced the structure much more frequently: 2.4 times on small numbers and 6.7 times on large numbers. The low frequency was due to the fact that many subset-knowers and some CP-knowers rarely produced both parts of the structure on the same trial; they either counted without completing their count with a cardinal answer or volunteered cardinal answers without having counted. Even on trials probing them to complete their counts with a cardinal answer (e.g., "So, what's on this card?"), many children (mostly subset-knowers) did not produce cardinal answers but responded by recounting or producing bare plurals (e.g., "apples").

Of course, the important question is whether children produced cardinal responses that matched the last word of their counts, and whether CP-knowers were any more likely to do so than subset-knowers. Fig. 4 shows that they were, particularly as compared to "one"- and "two"-knowers. Whereas CP-knowers nearly always gave a cardinal answer that matched the last word of their count, "one"- and "two"-knowers did so only on small numbers. Their errors were of many types; sometimes they were only off by one, but they were also often off by more than one (see Appendix A for some examples of errors committed by a typical "one"-knower). With an average match score of 89% for small numbers and 68% for large numbers, the "three"-/"four"-knowers were right between the "one"- and "two"-knowers and the CP-knowers.

Because of the small sample sizes, these results were analyzed with non-parametric tests. The effect of knower-level was first analyzed with the Kruskal–Wallis test. For small numbers, this effect was not significant, $\chi^2(3) = 4.49$, $p = .21$. For large numbers, the effect of group was significant, $\chi^2(3) = 20.51$, $p < .001$. Pair-wise comparisons using the Mann–Whitney test show that, for large numbers, the average match score for CP-knowers ($M = 99\%$, $SE = 1.0$) was greater than that of every other group ("one"-knowers: $M = 0\%$, $SE = 0$, $Z = 3.67$, $p < .001$; "two"-knowers: $M = 6\%$, $SE = 6.0$, $Z = 3.88$, $p < .001$; "three"-/"four"-knowers: $M = 68\%$, $SE = 18$, $Z = 2.68$, $p < .01$). The

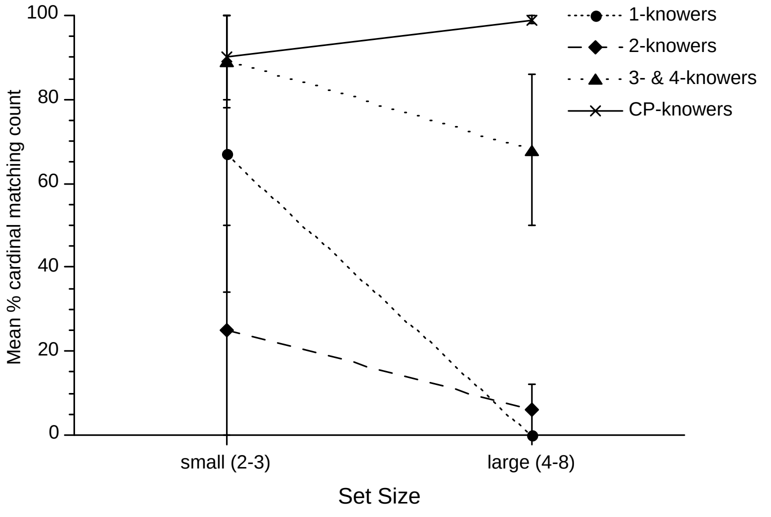

Fig. 4. Mean match scores as a function of set size and GN knower-level. Each child's match score was equal to the number of cardinal answers matching the last word of the count preceding it out of the total number of cardinal answers preceded by a correct count in the WOC task.

“three”-/“four”-knowers were intermediate between CP-knowers and the other subset-knowers (“one”-knowers, $Z = 1.92$, $p < .05$; “two”-knowers, $Z = 1.93$, $p < .05$). Clearly, “two”-knowers did not differ from “one”-knowers, $Z = 0.87$, $p > .10$.

Fig. 4 displays another important result. The subset-knowers’ match scores were worse for large sets than for small sets. This suggests that their ability to produce cardinal answers matching the last word of their count was not a product of knowledge of the relation between counting and cardinality. Rather, it may have reflected an accidental correspondence between procedurally correct but numerically meaningless counting of small sets and subitizing. CP-knowers, in contrast, apparently understood how counting represents numerosity, as reflected in their production of matches irrespective of set size.

In sum, the analyses discussed thus far show that while both subset-knowers and CP-knowers count large sets more often than small sets, subset-knowers differ radically from CP-knowers in ways that suggest that they do not understand the cardinal principle. That is, the critical difference between subset-knowers and CP-knowers lies not in their procedural and utilization skills but in their understanding of counting as a representation of number. There is, however, one notable exception to this generalization: the “three”-/“four”-knowers. About a third of these children produced cardinal answers that matched the last word of their count most of the time, for both small and large sets, raising the possibility that at least some of these children may in fact have understood how counting represents number.

2.2.4.3. Knower-levels from WOC. With the exception of a few “three”-/“four”-knowers, the above analyses suggest that GN and WOC provide consistent pictures of children’s understanding of counting. However, they do not address whether children’s knower-levels were the same across the two tasks. Wynn (1990, 1992) had found that children’s GN knower-levels were the same as their knower-levels as assessed by the Point-to-X task. “One”-knowers on GN could only correctly identify a target set in the Point-to-X task if one of the choices was a set with only one object, “two”-knowers on GN could only do so if one of the choices was a set of two objects, and so on. Such consistency supports the discontinuity hypothesis; it suggests that “one”-knowers truly only know the numerical meaning of “one,” “two”-knowers truly only know the numerical meaning of “one” and “two,” and so on, across tasks with distinct processing demands. Thus, given that WOC makes fewer procedural demands than Point-to-X, finding that children’s GN knower-level is the same as their knower-level on WOC would provide strong evidence in favor of the discontinuity hypothesis.

To address this question, we examined children’s cardinal answers on WOC. Cardinal answers consisted of: number words occurring alone or as noun phrase quantifiers, either produced after a count (e.g., “One, two, three. Three!” or “one, two, three. That’s three bears”) or outside the context of counting (e.g., “three” or “three bears”). To be granted knowledge of the exact meaning of a number word, children had to:

(1) Say “*N*” at least 67% of the time when presented with *N* stickers.[5]
(2) Say “*N*” no more than half as often when presented with different numbers.

[5] If children incorrectly counted a set of *N* stickers and repeated the last word of the count, they were considered to have said “*N*” only if their count contained no more than one error. Trials in which the counts contained more than one error were excluded from this analysis.

Table 2
Knower-level groups from What's on This Card (WOC)

Levels	N	Age[a]		Count list length	
		Mean	Range	Mean	Range
0-/"one"-knowers[b]	14	2;5	2;1–3;3	7.7	6–10
"Two"-knowers	8	3;2	2;6–3;11	9.5	6–12
"Three"-knowers	9	3;2	2;9–4;0	10.5	9–12
CP-knowers	19	3;7	2;9–4;0	8.8	7–12

[a] Ages are in years and months (years; months).
[b] Because children hardly ever produced number words to describe sets containing only one sticker, we could not distinguish 0-knowers from "one"-knowers.

(3) Satisfy conditions 1 and 2 for all numbers less than *N*.[6]

Children were considered CP-knowers either if they met criteria (1), (2), and (3) for all numbers tested *or* if the cardinal answers they produced on *large numbers* matched their counts at least 50% of the time. In other words, children could be considered CP-knowers if their match scores for large numbers were greater than 50%. A high match score for small numbers was not taken as evidence of knowledge of the cardinal principle for the reasons outlined above.

Because children rarely produced number words to describe sets containing only one individual, we could not determine whether children who failed on all numbers but one were actually 0-knowers or "one"-knowers. These children were grouped together and will be referred to as 0-/"one"-knowers"(see Appendix A for an example of a typical 0-/"one"-knower).

Based on these criteria, we found that children could be divided in four groups: 0-/"one"-knowers, "two"-knowers, "three"-knowers, and CP-knowers. The size, average age, and age range of each group are reported in Table 2. The 0-/"one"-knowers included one child who did not produce any cardinal responses. The rest did, but, on average, they produced the correct number word only 12% of the time for arrays of 2–8 stickers. The "two"-knowers were almost always correct on sets of two (94% correct on average), but they hardly ever produced the right number words for arrays of 3–8 (16% correct on average). The "three"-knowers were correct 97% of the time on sets of 2 and 3 but only 14% of the time for arrays of 4 through 8. In contrast, the CP-knowers were almost always correct (89% of all trials). The marked difference between the way in which subset-knowers used the number words for small and large sets shows that these stages are not an artifact of our criterion; rather, they are stable, readily discriminable stages.

Of the 19 CP-knowers, 7 met criteria (1), (2), and (3) for all numbers tested. The other 12 failed to meet these criteria on at least one number, but they were nonetheless categorized as CP-knowers because their match score for large numbers was greater than 50%. The average match score was 96% for children categorized as CP-knowers and 3% for subset-knowers.

[6] Only three children failed to demonstrate that they understood the exact meaning of all number words below the highest number word which satisfied our criterion. Two were classified as "two"-knowers (one passed the criterion for "one," "two," and "seven"; the other for "one," "two," and "six"). The other was a "three"-knower who passed the criterion for all number words up to "three" and for "five".

The existence of subset-knowers on this task provides support for the discontinuity hypothesis. Since all children could recite the count list up to “six”, and most of them could recite it up to “eight” (40/50), it need not have been the case that some of them would not be able to provide correct cardinal responses for any set greater than one or only for sets of two, or only for sets of two and three. Yet, number word usage in 31 of the 50 children followed one of these patterns. This suggests that these children only knew the exact meanings of “one,” or “one” and “two,” or “one,” “two,” and “three,” and had not yet worked out how counting represents number.

2.2.4.4. WOC level as a function of GN level. As can be seen from Table 3, the two tasks were highly consistent. Assuming that children who were 0- or “one”-knowers on GN and were 0-/“one”-knowers on WOC had the same level on both tasks, 33 of the 50 children were classified in the same levels on the two tasks. However, some children had a higher knower-level on WOC than on GN. While most of these were only differences of one level, and almost none of them consisted of GN subset-knowers being classified as CP-knowers on WOC, this suggests that children may have been categorized on higher levels on WOC than on GN. To determine whether this difference was systematic, within-child knower-level differences were analyzed with a Wilcoxon Signed Ranks test. Since GN CP-knowers could not possibly have had a higher knower-level on WOC than on GN, they were not included in this analysis. Difference scores were computed by assigning numerical codes to each knower-level. WOC 0-/“one”-knowers and GN 0- and “one”-knowers were all coded as “0.5”; all others were coded with the number corresponding to their knower-level (e.g., “two”-knowers were coded as “2”). While more children had higher knower-levels on WOC than GN ($n = 12$) than the other way around ($n = 5$), this was not significant, $Z = 0.79$, $p = 0.4$. Thus, there is no evidence that children’s knower-levels were systematically higher on WOC than on GN.

As mentioned above, a few GN subset-knowers were classified as CP-knowers on WOC (4/30); all of them were “three”- or “four”-knowers. To further investigate whether these “three”-/“four”-knowers used counting more like CP-knowers than like subset-knowers, we re-analyzed spontaneous counting on WOC with the “three”-/“four”-knowers separated in two groups: *true* “three”-/“four”-knowers and *CP* “three”-/“four”-knowers (see Appendix B for an example of a true “three”-/“four”-

Table 3
Assessment of the consistency between Give a Number (GN) knower-levels and What’s on This Card (WOC) knower-levels

GN knower-levels	WOC knower-levels			
	0-/”one”	“Two”	“Three”	CP
0-knowers	6	1	0	0
“One”-knowers	5	2	0	0
“Two”-knowers	1	4	5	0
“Three”-knowers	2	0	3	3
“Four”-knowers	0	1	1	1
CP-knowers	0	0	0	15

Note. Each number at the intersection of a row and a column represents the number of children at each of the possible combinations of GN and WOC knower-levels. For example, all children who were CP-knowers on GN were also CP-knowers on WOC.

Table 4
Re-analysis of children's spontaneous use of counting on What's on this Card? (WOC) with give a Number (GN) "three"-/"four"- knowers divided into true "three"-/"four"-knowers and CP "three"-/"four"-knowers

Re-analyzed GN knower-levels	% of trials with spontaneous counting	
	Small number (2–3) trials	Large number trials (4–8)
0-knowers	21	6
"One"-knowers	27	41
"Two"-knowers	25	43
True "three"-/"four"-knowers	9	42
CP "three"-/"four"-knowers	19	71
CP-knowers	18	79

Note. *CP* "three"-/"four"-knowers are those GN "three"-/"four"-knowers who were CP-knowers on WOC; *true* "three"-/"four"-knowers are those who were classified as subset-knowers on WOC.

knower and Appendix C for a CP "three"-/"four"-knower). If "three"-/"four"-knowers constitute a genuine stage, and if some children classified as "three"-/"four"-knowers on GN have indeed been misclassified, then the percent of WOC trials on which the true "three"-/"four"-knowers spontaneously count should resemble those of the other subset-knowers, whereas that of the CP "three"-/"four"-knowers should cluster with those of the CP-knowers. Table 4 shows that this is exactly what happens. The pattern is particularly striking for spontaneous counting on large numbers where the score for the true "three"-/"four"-knowers is equal to that of the other subset-knowers, whereas that of the CP "three"-/"four"-knowers is essentially equal to that of the children who were classified as CP-knowers across both tasks.

2.3. Discussion

Like Wynn, we found that children can be reliably classified as subset-knowers or CP-knowers, and only the latter provide evidence of understanding how counting represents number. On GN, only CP-knowers regularly count to create sets of given numerosities, especially large ones, and only CP-knowers correctly adjust their sets if a count reveals they have made a mistake. Although all children in our sample could count at least to "six," several did not know the numerical meaning of any number words. Among subset knowers, some children knew only the meaning of "one," some knew only "one" and "two," some knew "one, two," and "three." A minor exception to our replication of Wynn's levels was the few subset-knowers who knew up to "four."

Our comparison of WOC and GN supports Gelman's contention that the greater task demands of GN affect some aspects of children's performance. WOC provides a better context for young children to spontaneously use counting than does GN. That is, while our subset-knowers rarely counted on GN, even when large numbers were requested, they counted large sets more often than small ones when tested on WOC. In addition, we found that WOC may reveal earlier *understanding* of how counting represents number. Indeed, a third of the children who were classified as "three"-/"four"-knowers on GN showed they understood the cardinal principle on WOC.

Still, the important result from Experiment 1 is that, except for these few "three"-/"four"-knowers, the two tasks provided overwhelmingly consistent pictures of what children understand about how counting represents number. In neither task did 0-, "one"-, "two"- and true "three"-/"four"-knowers demonstrate knowledge of the fundamental purpose of counting, namely determining cardinality. When they produced cardinal responses following a count on WOC, it was often the wrong one. Similarly, when they were provided with an opportunity to fix incorrect sets in GN, they often left them unchanged or even changed them in the wrong direction. Further, despite the vastly different procedural and utilization demands these two tasks placed on children, for the most part, their subset-knower level was the same on both tasks. Those who did not have the same knower-levels on both tasks usually varied by only one level in either direction (e.g., some GN "two"-knowers were "three"-knowers on WOC, and one was a "one"-knower on WOC). This finding extends the within-child knower level consistency from GN and Point-to-X (Wynn, 1990) to GN and WOC. Thus, as a whole, these results provide strong support for Wynn's characterization of the stages of number word learning, and fail to support the continuity hypothesis.

While our results strongly replicate Wynn's, we have not entirely replicated Gelman's WOC data. Table 5 shows the proportion of children who were CP-knowers on WOC in six different age groups, with the groups between age 2;6 and 3;5 corresponding to Gelman's (1993) three age groups. Gelman found that 30% of her two-year-olds (2;6 to 2;10), 80% of her young three-year-olds (2;11 to 3;2), and 80% of her old three-year-olds (3;3 to 3;5) both counted and produced the last word of their count on at least 50% of trials with large set sizes. Insofar as we probably would have classified these children as CP-knowers, Table 5 suggests that Gelman found more CP-knowers than we did, particularly amongst her young 3-year-olds. Thus, it is possible that our version of WOC underestimated children's competence. Our procedure differed from Gelman's in some respects that might be crucial. By not probing on each trial and by not providing counting assistance, perhaps we failed to overcome utilization and procedural limitations in our subset-knowers. Had we done so, we may have found that GN does massively underestimate children's understanding of the cardinal principle.

To address this possibility (and may we dare hope, to settle this debate once and for all), we created a new task that puts as few procedural and utilization demands on the child as we could imagine. We drew on the procedure introduced by Briars and Siegler (1984) and Gelman and Meck (1986) in which children judge whether a puppet has counted correctly. All previous studies using this procedure have analyzed children's performance as a function of age; here, as in Experiment 1, we will analyze their performance as a function of their knower-level on GN.

Table 5
Percentage of children classified as CP-knowers on What's on this Card? (WOC) as a function of age

Age range (years;months)	*N*	% of children in age range classified as CP-knowers on WOC
2;0 to 2;5	8	0
2;6 to 2;10	9	10
2;11 to 3;2	10	20
3;3 to 3;5	7	57
3;6 to 3;9	6	100
3;10 to 4;0	10	60

Note. Age ranges between 2;6 and 3;5 correspond to the three age ranges used in Gelman (1993) original What's on this Card? experiment.

3. Experiment 2

Briars and Siegler (1984) argued that the possession of abstract counting principles could only be revealed by one's ability to distinguish counts that violate the principles from counts that violate conventional usage but do not violate the principles (e.g., counting the right half of a line of objects before counting the left half). This argument led to a series of studies in which children watched as a puppet produced non-conventional correct counts, conventional counts, and incorrect counts and were asked to evaluate the acceptability of the puppet's counts (Frye et al., 1989; Gelman & Meck, 1983, 1986; Gelman et al., 1986). The results of these studies have been rather contradictory.

Gelman and Meck (1983) demonstrated that children as young as 3 years accepted non-conventional counting sequences but correctly rejected double-counting or skipping an object, using an incorrect sequence of number words, and reporting an incorrect cardinal value following a correct count for set sizes up to 20. To be sure, 3-year-olds' ability to detect violations of the cardinal principle could have been explained by their having learned to pay special attention to the last word of a count for non-numerical reasons (e.g., because the last number word in a count is often stressed more than the others, as in "one, two, three, FOUR!"); i.e., their behavior could have been explained by their having learned a numerically meaningless "last word rule." To control for this, Gelman and Meck also presented children with "trick trials." On such trials, the puppet first correctly counted a row of tokens and reported the correct cardinal value, but then counted in the other direction (i.e., from right to left), made a hidden mistake, and consequently reported a different cardinal value. Three-year-olds correctly inferred that the puppet must have made a mistake during its second count, suggesting that they were not following a numerically meaningless last word rule (Gelman & Meck, 1983; Gelman et al., 1986). Insofar as these results suggest that 3-year-olds follow abstract counting principles in evaluating novel counting situations, they are consistent with the continuity hypothesis.

However, other variations of the puppet procedure did not replicate these results. Briars and Siegler (1984), for example, found that although children as young as 3 years of age accepted unconventional counts more often than incorrect counts (e.g., double-counting or skipping an object), they often failed to reject incorrect counts. Conversely, 4- and to some extent 5-year-olds almost always rejected incorrect counts but they also rejected correct but unconventional counts. In sum, 3-year-olds, 4-year-olds, and even some 5-year-olds failed to distinguish conventional from unconventional correct counts. Moreover, Briars and Siegler reported that many children could count correctly on their own before they could successfully distinguish unconventional correct counts from incorrect counts in the error detection tasks, suggesting that counting skill comes before knowledge of the counting principles.

Gelman and her colleagues have appealed to differences in method to account for the discrepancy. For example, Briars and Siegler's children were asked to count sets on their own before they did the judgment task, whereas Gelman's children only did the judgment task. Gelman and Meck (1986) provided some evidence that having children count sets on their own before evaluating the puppet's performance might have induced children to incorrectly reject more unconventional counts. However, the reason why Briars and Siegler's 3-year-olds failed to distinguish non-conventional from incorrect counts is that they tended to *accept* both of them. Thus, no simple conclusion can be drawn from these studies.

The implications of these results for theories of the ontogeny of the counting principles are muddled further by two issues. First, asking children to judge whether a count is acceptable probably tests their metaconception of the counting principles rather than their ability to *use* these principles. In particular, distinguishing violations of conventionality from counting errors requires *explicit* knowledge of the counting principles, something children may not have until much later (Cowan, Dowker, Christakis, & Bailey, 1996). Thus, the counting puppet studies may have *under*-estimated children's knowledge by making excessive demands on utilization skills. Second, for reasons discussed in the introduction, these studies cannot be easily compared to Wynn's (1990, 1992) because they analyzed children's performance as a function of their age rather than as a function of their Wynn stage.

Experiment 2 attempts to settle this debate by testing children on GN and on a version of the puppet task that does not test children's metaconception of the counting principles. In our version of the puppet task, an experimenter first told a puppet how many objects to put into a container, and the puppet correctly counted out N objects or $N \pm 1$ objects, counting aloud as it placed each toy into the container. The experimenter then simply asked the child whether the puppet had indeed put N objects in the container. This version is likely the easiest employed to date. It retains the experimenter's doing the counting for the child, minimizing demands on procedural and utilization skills, and it removes the metacognitive aspect of past incarnations by asking children to evaluate the *number* of objects the puppet produced rather than the acceptability of the counting procedure. Moreover, insofar as our Counting Puppet task is even less demanding than WOC, Experiment 2 will also address concern that our version of WOC underestimated children's numerical competence.

Note that children could pass this task by simply deploying a numerically meaningless last word rule. Thus, this task could overestimate children's understanding of the cardinal principle. However, it is very hard to see how it could underestimate it. Therefore, if the GN CP-knowers, and only these children, pass the task, we will have very strong evidence for a conceptual discontinuity between subset-knowers, on the one hand, and CP-knowers, on the other.

3.1. Method

3.1.1. Participants

Thirty-seven 3- and 4-year-olds (mean = 3;9, range: 3;0 to 4;7) participated in a two-session study. The first session (not reported here) tested children's ability to estimate cardinality without counting, and to make non-verbal ordinal judgments. The second session consisted of GN, our Counting Puppet task, and a control count-list elicitation task, in that order. All of our children were monolingual speakers of English. They were recruited from birth records the Greater Boston area, and from the Kiddie Lodge (Framingham, Massachusetts). Some were tested at a university child development laboratory, and some were tested at the Kiddie Lodge. A caregiver accompanied all children tested at the laboratory. Children received a small gift for their participation.

3.1.2. Count list elicitation control task

3.1.2.1. Procedure and results. The sole purpose of this task was to make sure that children had learned a sufficient count list. Children were simply asked to count a row of ten identical toy animals (e.g., gorillas). If they counted without paying attention, the experimenter would slowly point to each animal and ask children to count again with him/her. Every child could recite the count list at least to "ten" in the conventional order. Insofar as our tasks never involved number words beyond "nine", this effectively rules out the possibility that children failed on any of the following tasks for lack of the relevant number words.

3.1.3. Give a Number (GN)

3.1.3.1. Procedure and results. Using the same procedure as in Experiment 1, we divided children into three knower-level groups, namely "one"-/"two"-knowers, "three"-/"four"-knowers, and CP-knowers. To create these groups, we determined the highest number of objects each child could reliably correctly give using the same criteria as in Experiment 1. Table 6 reports the number of children in each group, the mean age of children in each group, and their age range. The "one"-/"two"- and "three"-/"four"-knowers in this experiment were older than the ones in Experiment 1 (compare Tables 1 and 6). This is because many of the "one"-/"two"- and "three"-/"four"-knowers in Experiment 1 were 2 years old, but all of participants in Experiment 2 were at least 3 years old. The greater age of the subset-knowers in Experiment 2 thus provides a strong test of the discontinuity hypothesis.

These groups were chosen because they captured the different performance patterns on GN and WOC in Experiment 1. While both tasks suggest that 0-, "one"- and "two"-knowers do not understand how their count list represents number and that CP-knowers clearly do, the status of "three"-/"four"-knowers on GN is ambiguous. As Experiment 1 showed, GN clearly underestimates *some* of these children's understanding of how counting represents number; this task may reveal that it underestimates all of them. Hence we set out to determine whether "three"-/"four"-knowers would pattern more like "one"-/"two"-knowers or more like CP-knowers on the Counting Puppet task.

3.1.4. Counting Puppet task

3.1.4.1. Stimuli and procedure. This task involved a puppet, 15 small plastic elephants, and a small opaque trash can covered with a rotating lid. The experimenter (E) introduced children to the task by saying, "I am going to ask Kermit to put some elephants in the trash can, and you tell me if he does it right." Then, the E told Kermit how many elephants to put in the trash can. The E always began by asking Kermit to put six elephants in the trash can, then seven, and then eight. Some children were tested on a second block

Table 6
Knower-levels of children in Experiment 2 as determined with give a number (GN)

Knower-levels[b]	*N*	Age[a]	
		Mean	Range
"One"-/"two"-knowers	12	3;8	3;0–4;4
"Three"-/"four"-knowers	12	3;6	3;0–4;0
CP-knowers	13	3;9	3;3–4;7

[a] Ages are in years and months (years;months).

[b] Knower-levels were assessed with GN. There were six "one"-knowers, six "two"-knowers, nine "three"-knowers, and three "four"-knowers.

of trials ($n = 15$). The second block was comprised of the same numbers requested in the reverse order.

We chose six, seven, and eight as targets to be sure that children would not be able to succeed by using some analog of subitizing for visual events. With these numbers, counting should be the only way of determining exactly how many objects are in a set. Thus, children should only succeed if they understand counting, at least to the level of the last word rule.

In response to the E's request, the puppet put toys in the trash can by slowly counting them one at a time. The puppet always used number words in the conventional English order, and always respected one-to-one correspondence. However, it only counted correctly when it was asked for "seven". When it was asked for six, the puppet counted to "five" and placed five objects in the trash can, and when it was asked for eight, it counted to "nine" and placed nine objects in the trash can. On each trial, when the puppet was done counting, the E asked, "Is that *N*?" (where *N* was the number the puppet had been asked for).

3.1.4.2. Results. The three GN groups' average percent correct are presented in Fig. 5. A one-way ANOVA revealed that the three groups' performance differed, $F(2, 36) = 8.01$, $p < .001$. Tukey's post hoc tests showed that CP-knowers ($M = 88\%$ correct, $SE = 6.9$) performed significantly better than the "one"-/"two"-knowers ($M = 53\%$ correct, $SE = 7.5$, $p < .01$) and the "three"-/"four"-knowers ($M = 52\%$ correct, $SE = 9.8$, $p < .01$). We also tested each group's performance against chance (50%). Only the CP-knowers differed from chance ("one"-/"two"-knowers: $t(11) = .44$, *ns*; "three"-/"four"-knowers: $t(11) = .17$, *ns*; CP-knowers: $t(14) = 6.26$, $p < .001$).[7] The "one"-/"two"-knowers and the "three"-/"four"-knowers did not differ from each other. To ensure that the fact the some children were tested on two blocks of trials did not influence the results, we also looked only at first block responses. While this slightly increased the means ("one"-/"two"-knowers: 58% correct, "three"-/"four"-knowers: 56% correct, and CP-knowers: 92% correct), the results of the parametric analyses were the same.

An inspection of the number of children who always answered correctly confirmed the group results. We examined children's answers on the first block of trials and tested our data against the binomial distribution (with $p(\text{always correct}) = 1/8$ and $q(\text{other}) = 7/8$). Ten of the 13 CP-knowers were perfect; this was significantly different from chance ($p < .001$). In contrast, always saying "yes" was the dominant pattern in both subset-knower groups (five "one"-/"two"-knowers, and four "three"-/"four"-knowers). However, six subset-knowers always answered correctly. Three of them were "two"-knowers, two were "three"-knowers, and the other was a "four"-knower. If all subset-knowers were taken together, the number of children who always answered correctly (6/24) was marginally greater than expected by chance ($p = .07$). However, one of these six subset-knowers was tested on two blocks of trials, and while he performed perfectly on the first three trials, he

[7] Given that most of the children only had three trials, the number of possible percent correct scores for each child was very small. Thus, because we were worried that our data did not approximate continuous data, we also analyzed them non-parametrically. The ANOVA was done as a Kruskal–Wallis test, and the post hoc tests as Wilcoxon rank-sum tests. We also were able to test each group's performance against chance with Pearson's Chi-Square by using only the first three trials of the children who were tested twice in each condition. Our results were exactly the same when they were analyzed with these tests.

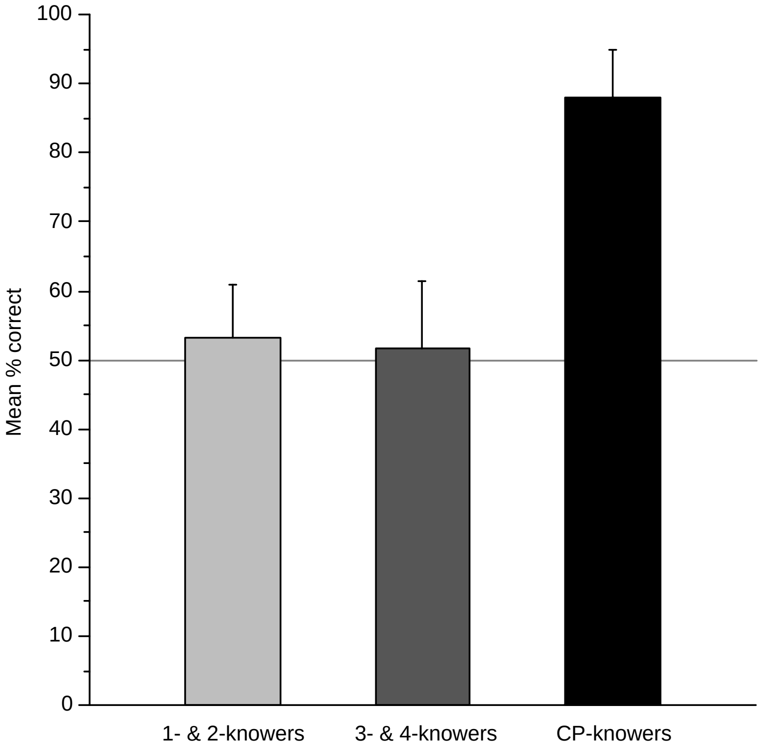

Fig. 5. Mean percentage of trials in which children correctly determined whether the puppet in the Counting Puppet task had counted out the requested number of objects as a function of GN knower-level.

made two mistakes on the next block of trials. None of the CP-knowers who were tested on two blocks did worse on the second than on the first block.

Thus, the data from the Counting Puppet task suggest that GN slightly underestimates children's understanding of counting. At best, it shows that *some* children categorized as subset-knowers on GN may actually be CP-knowers (or at least have a "last word rule"). The important result, of course, is the consistency between GN level and success on this task. Subset-knowers overall failed and CP-knowers robustly succeeded.

It could be that the difference between CP-knowers and subset-knowers was that subset-knowers used an "at least" interpretation of number words, whereas CP-knowers used an exact interpretation. On the "at least" interpretation, the puppet made a mistake when it put in five in response to a request for six, but not when it put in nine toys in response to a request for eight, as it *did* put in eight. Only two subset-knowers (a "one"-knower and a "three"-knower) produced a pattern of answers consistent with this interpretation. This is not different from what would be expected by chance (binomial test, $p = 0.8$); the difference between subset-knowers and CP-knowers thus cannot be attributed to a change from an "at least" interpretation to an exact one. Moreover, since only one CP-knower answered this way, these data show that our children understood us to be requesting *exact* numbers (see Papafragou and Musolino, 2003 for evidence that young children prefer the "exactly *N*" interpretation of number words).

We also analyzed children's performance as a function of their age to compare our results to studies that used similar tasks to assess children's comprehension of the cardinal principle but did so as a function of age (Gelman & Meck, 1983; Gelman et al., 1986). Recall that in two studies, Gelman's group found that 3-year-olds almost always correctly determined whether a puppet had correctly answered a "How Many?" question after it had counted a set. Moreover, using "trick trials" (see introduction to Experiment 2), Gelman et al. (1986) showed that 70% of their 3-year-olds knew that different counts of the same set (e.g., counting from left to right and counting from right to left) should always yield the same cardinal value. Such knowledge could not have followed from a numerically meaningless last word rule. Therefore, they argued, these children must have known how counting represents number.

In our study, the average percent correct for the 3-year-olds in the age range used in the Gelman studies (3;2 to 3;10) was 63%. While this is much lower than Gelman and her colleagues found, it is significantly greater than chance, $t(22) = 2.15$, $p < .05$. Moreover, whereas at least 70% of the 3-year-olds tested by Gelman and colleagues truly understood how counting represents number, on our task only 8/23 (35%) children in this age range could be considered to understand counting in that they answered all three questions correctly. Though this proportion is only half of that found by Gelman et al., it is significantly greater than would have been expected by chance, binomial test, $p < .001$. Thus, while our results fail to replicate the size of the effects found by the Gelman research team, they are consistent with their claim that 3-year-olds understand how counting represents number. What is new in our data is that this generalization holds only for those children revealed to be CP-knowers on Wynn's GN task.

3.2. Discussion

Our version of the Counting Puppet task minimized demands on children's procedural and utilizational skills more than any other test of children's understanding of counting. Because it required children to evaluate a state of affairs in the world (i.e., the number of objects given by the puppet) rather than the acceptability of a procedure yielding a representation, our task eliminated the metacognitive demands of previous Counting Puppet tasks. Moreover, this task was clearly easier than GN because children did not have to figure out that counting was relevant to this task nor did they have to keep track of a target while trying to co-ordinate counting and set construction. Finally, because children simply observed counting produced by a puppet instead of having to count by themselves, this task was a more sensitive measure of children's understanding of counting than WOC or GN.

Therefore, if children categorized as subset-knowers on GN were really CP-knowers with fragile procedural and utilization skills, their performance on the Counting Puppet task should have been as good as that of children categorized as CP-knowers. Likewise, if children's GN performance predicted their performance on WOC because our version of WOC was too difficult, we should not have found a difference between subset-knowers and CP-knowers on our Counting Puppet task. On the other hand, if children perform like "one"-, "two"-, "three"-, or "four"-knowers on GN or WOC because they don't understand how counting works and only know the meaning of a subset of the number words in their count list, then subset-knowers should nonetheless fail the much easier Counting Puppet task.

We found some evidence that GN underestimates children's numerical competence. Five of the 24 children classified as subset-knowers with GN always answered correctly, and one always answered correctly on the first block of trials but made mistakes on the second one. This is consistent with our finding that some children classified as "three"-/"four"-knowers with GN were categorized as CP-knowers when tested on WOC.

These few subset-knowers aside, the dominant pattern in our results was that children's performance on GN was a strong predictor of their performance on the Counting Puppet task. Whereas the large majority of children classified as CP-knowers on GN were always able to determine whether the puppet had counted out the right number of objects, the large majority of subset-knowers failed to do so, even though all of them could recite the count list up to "ten" and the puppet never counted further than "nine". Therefore, these data show that the consistency between GN and WOC cannot solely be explained by the use of a difficult version of WOC. Moreover, they show that, while *some* children who are classified as subset-knowers on GN may actually be CP-knowers, performance on GN is generally a very good index of children's knowledge.

Our results also underscore the potential pitfalls of grouping children as a function of their age, and of attributing the average or modal performance to all children within an age range. Indeed, despite the strong relation between GN and the Counting Puppet, we found that the average performance of our 3-year-olds was better than chance. Thus, had we only analyzed children's performance on the Counting Puppet task as a function of their age, we could have overlooked the fact that 3-year-olds are actually a mixed group of subset-knowers and CP-knowers.

In sum, the results of this experiment provide striking confirmation of a qualitative difference between subset-knowers' and CP-knowers' understanding of the cardinal principle, supporting the discontinuity position. When children become CP-knowers, they have figured out something about counting that subset-knowers simply do not know. Most likely, they have created, for the first time, an integer list representation of the positive integers.

4. General discussion

Two- to 4-year-old children were tested on tasks that assessed their understanding of counting, but that varied greatly both in the type and the extent of their performance and utilization demands. In some cases, variations in task demands did affect children's performance. First, we found that one of the major differences between subset-knowers and CP-knowers—namely, the fact that CP-knowers use counting more often to determine the cardinality of a set if it is large, whereas subset-knowers rarely use counting even for large sets—is attenuated when children are tested on WOC. Subset-knowers were more likely to count when presented with large sets than with small sets in WOC, although this difference was much more pronounced amongst CP-knowers. Insofar as subset-knowers arguably understood that the goal of the WOC task was to name the cardinality of each set, this suggests that they are beginning to understand that counting is somehow related to cardinality, even though they may not yet have worked out exactly how. Second, both tasks provided some evidence that some children categorized as subset-knowers when tested on GN might actually have been CP-knowers. In Experiment 1, we found that four of nine children who had been categorized as "three"- or "four"-knowers on GN were identified as CP-knowers when they were tested on WOC. In Experiment 2, one-quarter of the children who had been categorized as "two"-, "three"- or "four"-knowers on GN performed just like CP-knowers on the Counting Puppet task; they always correctly determined whether the puppet had counted out the correct number of objects.

Although they show that procedural and utilization demands of the tasks used to assess children's understanding of counting can affect children's performance, these data fall far short of confirming the continuity hypothesis that competence with counting is present from the earliest moments. This is because the most prevalent (and perhaps surprising) pattern of performance was one of consistency across tasks. This consistency held at two levels of description. First, children who failed to show that they understood how counting determines the cardinality of sets on one of these tasks (i.e., children who were subset-knowers on GN or WOC), also failed to show this on all the other tasks. Children who failed to assemble sets within their counting range in GN also failed to link counting to cardinality in the arguably less demanding WOC task. Even more dramatically, they failed to understand the relation between counting and cardinality on the Counting Puppet task, even though here they had only to appreciate that the production of a given numerosity requires that a correct count must terminate at the numerosity requested. On the other hand, children who showed that they understood how to use counting to represent number in one task did so in all the tasks. The large majority of the children who could give any set requested on GN, and who were able to use counting both to create and fix sets, always respected the cardinal principle when tested on WOC, and were always able to determine whether the puppet had counted out the correct number of objects in the Counting Puppet task.

Second, children could be classified as 0-, "one"-, "two"-, or "three"-knowers in both tasks (some "four"-knowers were also found in GN), and most children had the same knower-level on both tasks. For example, children who were "one"-knowers in one task usually were "one"-knowers on the other. When knower-levels were discrepant, they were generally off by only one level in either direction. This is the first demonstration that knower-levels also obtain in language production tasks, and that knower-levels are typically the same regardless of whether they are assessed via language comprehension (e.g., with GN or Point-to-X) or production (e.g., with WOC).

Our analyses of knower-levels are consistent with previous cross-sectional and longitudinal descriptions of the sequence in which children move from one knower-level to the next. Experiment 1 showed that when children first memorize a short count list, they do not know the meaning of any of the number words in their list, not even that of "one;" these are the children who were dubbed "0-knowers." Moreover, both experiments (GN and WOC in Experiment 1, and GN in Experiment 2) showed that children then go through a period during which they can count quite high (often at least up to "eight"), but only understand the numerical meanings of the small number words in their count list. That is, during this period, children learn the numerical meaning of "one," then "two," then "three," and, rarely, "four." Finally, these studies confirm Wynn's findings that the period during which children only know the meaning of a subset of the number words in their count list lasts at least a year. Children need a year or more to come to understand how their count list represents the positive integers.

In sum, despite some inconsistencies between tasks, the current results show that subset-knowers are not CP-knowers with fragile procedural and utilization skills; rather, they are at a stage where their representation of number differs qualitatively from that of CP-knowers. Therefore, the results strongly support the proposal that the acquisition of the representational resources embodied in the count list involves a qualitative shift in children's representation of number.

This conclusion is supported by four other types of evidence: the relations between stages of acquisition and non-verbal representations of number, cross-cultural patterns of acquisition of the count list, anthropological and linguistic investigations of the cultural history of count-based representations of the integers, and studies of the representations of number in human infants.

First, subset-knowers differ from CP-knowers not only in their interpretation of numerical language but also in their non-verbal numerical abilities. Mix and her colleagues found that children classified as having "minimal counting proficiency" (a close equivalent of our CP-knowers) with the GN task and the How Many? Task can match the cardinality of visually presented sets of dots both with that of other sets of dots and with that of sequences of sounds, but children classified as "less than minimally proficient" (a close equivalent of our subset-knowers) only find numerical matches if the sets were both presented visually (Mix, Huttenlocher, & Levine, 1996; Mix, Levine, & Huttenlocher, 1999). Similarly, children who were not yet "two"-knowers on GN were unable to identify numerical equivalence across visually dissimilar sets ranging in size from two to four (Mix, 1999). Brannon and Van de Walle (2001; see also Rousselle et al., 2004) have also found that non-verbal numerical abilities are different across subset-knower levels. Only children who were at least "two"-knowers on WOC could determine which of two area-controlled sets was greater in number. In sum, the stages of mastery of the counting principles as measured by GN and WOC are related to performance across a wide range of non-verbal numerical tasks, suggesting once more that the distinction between these groups reflects a difference in underlying competence rather than merely differences in counting skills.

Second, recent studies have used GN and WOC to study number word acquisition in Russian (Sarnecka, 2004), Chinese (Li, LeCorre, Shui, Jia, & Carey, 2003), Korean (Le Corre, Li, & Lee, 2004), and Japanese (Sarnecka, 2004) children. These studies found that cross-linguistic variations in the structure of numerical morphology (e.g., the presence or absence of singular/plural morphological markers) affected the rate at which children progressed through the knower-level stages. These data support the speculation that the meaning of number words is initially shaped by the way in which other linguistic structures encode numerical distinctions (Bloom & Wynn, 1997) rather than by their position in the count list.

Despite these cross-linguistic differences, the sequence of acquisition of exact number word meanings was the same in all of these cultures. That is, children in each of these cultures could be divided into the now familiar "subset-knower" and "CP-knower" groups, and the "subset-knowers" further subdivided into "one"-, "two"-, "three"-, and "four"-knowers. This suggests that knower-levels are so robust a phenomenon as to occur in very similar forms in very different languages and cultures. Thus, though the count list itself is a cultural construction, the process of construction in childhood (and perhaps in human history) must be guided by innate cognitive universals. We speculate that these universals include those that support the acquisition of non-integer quantifiers (e.g., "a" and "some") and numerical morphology (e.g., singular/plural or singular/dual/plural).

Third, if the initial state of the representation of the naturals is not formulated in terms of counting, there should have been a time in human history where humans did not use counting as a representation of the naturals. Bodies of evidence from archeology, anthropology, and linguistics suggest that this was so. The archeological record suggests that while counting was a relatively early invention, the earliest representations of number were not count-based. The oldest recorded form of number representation is a bone plaque dated from the Upper Paleolithic (ca. 10,000 BC; Marshack, 1991). The plaque is covered with groups of 29 small incisions, possibly representing the lunar cycle. This suggests that the very first forms of representation of exact number were based on the creation of *model collections* rather than on an ordered list of symbols. While model collections have the same content as symbols in an ordered list, their format is radically different. That is, whereas each of the components of an ordered list (e.g., number words) can stand for a cardinal value, no single component of a model collection does so; the model collection stands for the cardinal value only when taken as a whole. Therefore, insofar as it suggests that the very first spontaneous representations of number were not based on counting principles, the archeological record is consistent with discontinuity.

Moreover, while rare, some cultures do not have count-based representations of number. The Pirahã, a small Amazonian tribe, represent the best-documented case. Their language only has number words for "one", "two", and "many" and has no counting system (Gordon, 2004). Gordon tested Pirahã adults on simple numerical tasks, most of which involved reproducing sets of up to ten objects. The Pirahã never used any strategy that remotely resembled counting to solve these tasks. They did not even use strategies based on one-to-one correspondence and performed rather poorly, particularly when trying to reproduce larger numbers. This is strong evidence for discontinuity insofar as it is an existence proof that exact representations of large numbers are not universal. Moreover, the fact that a culture without counting nonetheless has number words for "one" and "two" is consistent with the conclusion that in cultural history, the meanings of number words were not initially derived from a count list but may have been derived from quantificational resources of natural language, just as they are in language acquisition.

Analyses of morphological and syntactic features of contemporary languages suggest that the early forms of human languages resembled Pirahã in that they only included a small number of words denoting particular cardinal values. First, in a review of some 250 languages, Corbett (1996) argues that one, two, and three are the only cardinal values that are marked morphologically. That is, some languages are like English in that they only make a singular/plural distinction, but others make a tri-partite distinction between singular, dual, and plural (e.g., Upper Sorbian, a West Slavonic language), and yet others have separate markers for singular, dual, trial, and plural (e.g., Larike, an Indonesian language). However, no languages have separate markers for numbers beyond three. According to Corbett, number words are likely to be the origin of these morphological markers because the typical pattern of grammaticalization is for independent words to progressively turn into pieces of bound morphology. Therefore, it could be that the upper limit on number morphology was determined by the fact that early languages did not have any number words for numbers greater than 3. On the other hand, it could be that early languages had many more number words, but that only the most frequent were grammaticalized. Since number word frequency is inversely proportional to the number referred to (Dehaene & Mehler, 1992), this could be why only the first three numbers were grammaticalized.

A review of other lines of evidence supports the former view—i.e., that the first three number words existed prior to counting (Hurford, 1987). First, in many inflecting languages (e.g., Modern Hebrew, Russian, Welsh, and Ancient Greek), the first three or four number words agree in gender and/or case with the nouns they quantify; other number words are invariant. Hurford argues that this suggests that the number words beyond "three" were initially not used as means of quantifying noun phrases, but were only used as symbols in a rote-memorized sequence. Second, in many languages, the ordinal forms of the first two or three numbers cannot be predicted from their cardinal forms (e.g., compare "one" and "first", or "two" and "second" with "three" and "third"). This suggests that cardinal and ordinal meanings were initially unrelated. Insofar the count list unites the ordinal and cardinal meanings of number words, this further suggests that the first two or three number words existed before counting was developed. Finally, in some languages such as Russian and Mandarin, some of the words in the count list are not the same as those used to quantify noun phrases. Russian has two words for "one"—"raz" for counting, and "odin" for quantification—and Mandarin has two words for "two"—"erh" for counting and "liang" for quantification. This further supports the contention that the small number words and the count list had initially independent linguistic histories.

Studies of number representations in preverbal infants provide yet another line of evidence that the acquisition of counting as a representation of number requires the construction of a new representational resource. To be sure, this research does not show that infants have *no* representations of number. Quite the contrary, it has shown that infants are endowed with two distinct systems of representation with numerical content (see Feigenson, Dehaene, and Spelke, 2004 for a review). One system represents small sets by tracking each individual separately, establishing one symbol for each individual in the set. For example, infants would represent three crackers as "object_1, object_2, and object_3" or as "cracker_1, cracker_2, and cracker_3." Many have argued that this system is the same as the "object-file" system described in research on adult attentional resources (e.g., Kahneman, Treisman, & Gibbs, 1992). The other core system—best known as the "analog magnitude system"—represents number by generating signals the magnitudes of which are a linear or logarithmic function of cardinal values (see Dehaene, 1997 for a review of research on analog magnitudes).

This research is consistent with evidence that counting is constructed because it shows that the core number systems differ from counting in their format and expressive power. Unlike counting, object-files do not represent sets with a single symbol. In the above example with the three crackers, three different symbols were used to represent the set of crackers. Thus, the format of object-files is much more similar to that of the model collections witnessed on the Upper Paleolithic bone plaque than to that of counting. Moreover, when compared to counting, the expressive power of object-files is dramatically limited: this system cannot represent sets comprised of more than three individuals in infants (Feigenson & Carey, 2003; Feigenson, Carey, & Hauser, 2002) or four individuals in adults (Trick & Pylyshyn, 1994) and perhaps even in young toddlers (Le Corre & Carey, 2005).

Gallistel and Gelman (1992) have argued that analog magnitudes, unlike object-files, have the same format as counting. That is, they have argued that the mechanism whereby magnitudes are generated works just like counting. While this is possible in principle, much evidence suggests that whereas counting is iterative—i.e., it involves the serial repetition of the process of assigning numerals to objects—the mechanism whereby magnitudes are generated is not iterative (Barth, Kanwisher, & Spelke, 2003; Nieder & Miller, 2004). For example, in adults, the time required to compare two sets of objects using analog magnitudes is a function of the ratio of the sets, not of their absolute size (Barth et al., 2003). Thus, deciding whether a set of 30 dots is larger than a set of 20 takes as much time as deciding which of a set of 90 or 60 dots is larger. Since an iterative process would require more time to generate representations of larger sets, this result shows that the process that generates magnitudes does not have the same format as counting—i.e., it suggests that, unlike counting, this process is not iterative (see Church and Broadbent, 1990; Dehaene and Changeux, 1993; Verguts and Fias, 2004 for non-iterative models of analog magnitudes). Recently, Wood and Spelke (2005) have extended this finding to young infants, showing that the amount of time infants need to discriminate two sets using analog magnitudes is a function of the ratio of the sets, but not of their size—e.g., the amount of time infants needed to discriminate 4 from 8 (about 2s) was the same as they needed to discriminate 8 from 16.

Moreover, analog magnitudes and counting differ in terms of expressive power. Because numbers represented as magnitudes can only be distinguished given favorable ratios (e.g., at least 2:3 for 9-month-olds; Lipton & Spelke, 2003), magnitudes cannot distinguish all pairs of successors (e.g., the 9-month-old's system cannot distinguish 4 from 5). Therefore, magnitudes cannot represent the successor function, and cannot be the basis of the induction that "next" in the count list encodes "add 1."

Thus, studies of the structure of the core representations of number provide further evidence that the acquisition of the counting principles requires the construction of a new representational format and increases the expressive power of children's conceptual system. Of course, the core systems are almost certainly drawn upon in this construction process. Many have proposed that the construction process involves mapping the numerals in the count list onto representations in the core systems. Some proposals implicate only object-files (Carey, 2004; Hurford, 1987), others only implicate analog magnitudes (Dehaene, 1997; Verguts & Fias, 2004), and yet others implicate both core systems (Spelke & Tsivkin, 2001). To foreshadow, analyses reported by Le Corre and Carey (2005) show that children only need to map "one", "two", "three", and sometimes "four" onto core systems to construct the counting principles; number words beyond "four" are only mapped onto analog magnitudes *after* children have constructed the principles. Since the number words that are learned prior to the induction of the counting principles correspond to the numbers that fall within the range of the adult object-file system, this finding suggests that the construction process involves this system—either alone or together with analog magnitude representations of small sets—but does not involve analog magnitude representations of sets of 5 or more.

In sum, archeological, anthropological, and linguistic investigations of numerical representations across time and cultures, and analyses of the content and format of infant numerical representations all converge on the same interpretation: the integers were not created by God or natural selection but were slowly created by homo sapiens, over historical time. Moreover, our work and that of many others suggests that they must be re-created by each child who is exposed to the artifact that embodies the product of this long construction process, namely the verbal count list.

Acknowledgments

We thank Alice Dong, Geffen Goddard, Tom Noerper, and Shruti Ravikumar for their help with the collection and analysis of data for Experiment 1 and Julie Goldman and

Yoon Lee for their help with the collection and analysis of data for Experiment 2. Finally, we thank the many parents, children, and day care personnel who generously contributed to this project. This research was supported by FCAR Masters Research Scholarship 10378 to ML, NICHD Grant #R03 MH060757 to GAV, NIMH Grant #RO1 MH066154-01 to EMB, and NIH Grant #R01 HD038338 and NSF Grant #REC-0196471 to S.C.

Appendix A

What's on this Card? performance of a child classified as a "one"-knower with GN and as a 0-/"one"-knower with WOC.

Experimenter (E): What's on this card? (1 teddy bear sticker on card)
Child (C): A teddy bear.
E: That's right! One teddy bear.
E: What's on this card? (3 bears on card)
C: Two teddy bears. One, two, three (points to each bear while counting).
E: What's on this card? (2 bears)
C: Two teddy bears.
E: Can you show me?
C: One, two (points to each bear while counting).
E: What's on this card? (4 bears)
C: A lotta teddy bears. One, two, three, four (points to each bear while counting).
E: What's on this card? (5 bears)
C: Lotta teddy bears. One, two, three, four, five (points to each bear while counting).
E: So, what's on this card?
C: Two teddy bears.
E: What's on this card? (7 bears)
C: Two teddy bears.
E: What's on this card? (6 bears)
C: Two teddy bears.
E: What's on this card? (8 bears).
C: Two teddy bears. One, two, three, four, five, six, eleven, twelve, thirteen (counts one of the bears twice; points to all others once while counting).

Appendix B

Performance of a child classified as a "three"-/"four"-knower on both GN and WOC (referred to as a true "three"-/"four"-knower).

Experimenter (E): What's on this card? (1 bear sticker on card) Child (C): One teddy bear.
E: That's right! One bear.
E: What's on this card? (3 bears on card)
C: Three.
E: What's on this card? (2 bears)

C: Two.
E: Can you show me?
C: One, two (points to each bear while counting).
E: What's on this card? (4 bears)
C: Three, four, five, six (points to each bear while counting).
E: What's on this card? (5 bears)
C: One, two, three, four, five (points to each bear while counting).
E: So then what's on this card?
C: Four.
E: What's on this card? (7 bears)
C: Five teddy bears.
E: What's on this card? (6 bears)
C: Four bears.
E: Can you show me?
C: One, two, three, four, five, six.
E: So what's on this card?
C: Seven bears.
E: What's on this card? (8 bears).
C: One, two, three, four, five, six, seven, eight, nine, ten (points to each bear while counting up to "eight" and then says "nine, ten" without pointing to any bears).

Appendix C

What's on this Card? performance of a child classified as a "three"-/"four"-knower with GN and as a CP-knower with WOC (referred to as a CP "three"-/"four"-knower).

Experimenter (E): What's on this card? (1 apple sticker on card)
Child (C): A apple.
E: That's right! It's one apple.
C: An apple.
E: What's on this card? (2 apples on card)
C: Two apples.
E: What's on this card? (3 apples)
C: Three apples (shows three fingers).
E: Can you show me?
C: There is one, two, three (counts her fingers one by one).
E: What's on this card? (4 apples)
C: (shows four fingers).
E: Can you tell me?
C: Five.
E: What's on this card? (5 apples)
C: This (shows five fingers and counts them one by one silently). Mother: Can you count them on the card?
C: One, two, three, four, five (points to each apple while counting). There is five (says that after card has been taken away).
E: What's on this card? (7 apples)

C: One, two, three, four, five, six, seven (points to each apple while counting). There is seven.
E: What's on this card? (8 apples)
C: One, two, three, four (points to four of the eight, but can't reach the others).
E: (moves child closer to table).
C: One, two, three, four, five, six, seven, eight (points to each apple while counting).
E: What's on this card? (6 apples).
C: One, two, thee, four, five, six (points to each apple while counting).
E: So then what's on this card?
C: One, two, thee, four, five, six (points to each apple while counting). Six.

References

Barth, H. C., Kanwisher, N., & Spelke, E. S. (2003). The construction of large number representation in adults. *Cognition, 86*, 201–221.

Bialystok, E., & Codd, J. (1997). Cardinal limits: Evidence from language awareness and bilingualism for developing concepts of number. *Cognitive Development, 12*, 85–106.

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language, 24*, 511–533.

Brannon, E. M., & Van de Walle, G. A. (2001). The development of ordinal numerical competence in young children. *Cognitive Psychology, 43*, 53–81.

Briars, D., & Siegler, R. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology, 20*, 607–618.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind and Language, 16*, 37–55.

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus, 133*, 59–68.

Church, R., & Broadbent, H. (1990). Alternative representations of time, number, and rate. *Cognition, 37*, 55–81.

Corbett, G. (1996). *Number*. Cambridge University Press.

Cordes, S., & Gelman, R. (2005). The young numerical mind: When does it count. In J. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 127–142). New York: Psychology Press.

Cooper, R. G. (1984). Early number development: Discovering number space with addition and subtraction. In C. Sophian (Ed.), *Origins of Cognitive Skills* (pp. 157–192). Hillsdale, NJ: Erlbaum.

Cowan, R., Dowker, A., Christakis, A., & Bailey, S. (1996). Even more precisely assessing children's understanding of the order-irrelevance principle. *Journal of Experimental Child Psychology, 62*, 84–101.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5*, 390–407.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43*, 1–29.

Durkin, K. (1993). The representation of number in infancy and early childhood. In C. Pratt & A. F. Garton (Eds.), *Systems of representation in children: Development and use* (pp. 133–166). NewYork: Wiley.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object files: Evidence from infants' manual search. *Developmental Science, 6*, 568–584.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science, 13*, 150–156.

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314.

Freeman, N. H., Antonucci, C., & Lewis, C. (2000). Representation of the cardinality principle: Early conception of error in a counterfactual test. *Cognition, 74*, 71–89.

Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). 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., & Hall, J. W. (1983). The acquisition of early number word meanings: A conceptual analysis and review. In H. P. Ginsburg (Ed.), *The Development of Mathematical Thinking*. New York: Academic Press.

Fuson, K. C., Lyons, B., Pergament, G., & Hall, J. W. (1988). Effects of collection terms on class-inclusion and on number tasks. *Cognitive Psychology, 20*, 96–120.

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

Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. Advances in research theory. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation* (pp. 61–96). San Diego: Academic Press.

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

Gelman, R., & Greeno, J. G. (1989). On the nature of competence: Principles for understanding in a domain. In L. B. Resnick (Ed.), *Knowing and Learning: Issues for a Cognitive Science of Instruction* (pp. 125–186). Hillsdale, NJ: Earlbaum.

Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition, 13*, 343–359.

Gelman, R., & Meck, E. (1986). The notion of principle: The case of counting. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: The Case of Mathematics* (pp. 29–57). Hillsdale, NJ: Erlbaum.

Gelman, R., Meck, E., & Merkin, S. (1986). Young children's numerical competence. *Cognitive Development, 1*, 1–29.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*, 496–499.

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual change and children's counting. *Cognitive Psychology, 16*, 94–143.

Hurford, J. R. (1987). *Language and number: The emergence of a cognitive system*. New York: Blackwell.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Klahr, D., & Wallace, J. G. (1976). *Cognitive development: An information-processing view*. Hillsdale, NJ: Erlbaum.

Le Corre, M., & Van de Walle, G. A. (2001, April). *On the Role of the Singular/Plural Distinction in Early Number Word Learning*. Poster presented at the 64th biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

Le Corre, M., & Carey, S. (2005). On the role of core representations of number in the construction of the counting principles. Manuscript submitted for publication.

Le Corre, M., Li, P., & Lee, Y. (2004). [Numerical bilingualism: Count list acquisition in Korean]. *Unpublished raw data*.

Li, P., LeCorre, M., Shui, R., Jia, G., & Carey, S. (2003, October). *Effects of plural syntax on number word learning: A cross-linguistic study*. Paper presented at the 28th Boston University Conference on Language Development, Boston, MA.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large number discrimination in human infants. *Psychological Science, 14*, 396–401.

Marshack, A. (1991). The Tai plaque and calendrical notation in the upper Paleolithic. *Cambridge Archeological Journal, 1*, 25–61.

Mix, K. S. (1999). Preschoolers' recognition of numerical equivalence: Sequential sets. *Journal of Experimental Child Psychology, 74*, 309–332.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (1996). Do preschool children recognize auditory–visual numerical correspondences?. *Child Development 67*, 1592–1608.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. New York: Oxford University Press.

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology, 35*, 164–174.

Nieder, A., & Miller, E. K. (2004). Analog numerical representations in rhesus monkeys: Evidence for parallel processing. *Journal of Cognitive Neuroscience, 16*, 889–901.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition, 86*, 253–282.

Rousselle, L., Palmers, E., & Noel, M.-P. (2004). Magnitude comparison in preschoolers: What counts. Influence of perceptual variables. *Journal of Experimental Child Psychology, 87*, 57–84.

Sarnecka, B. W. (2004). *Language as lens: Plurality marking and numerical learning in English, Japanese, and Russian*. Unpublished manuscript.
Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6*, 357–379.
Siegler, R. S. (1991). In young children's counting, procedures precede principles. *Educational Psychology Review, 3*, 127–135.
Siegler, R. S. (1998). *Children's thinking* (third ed.). Upper Saddle River, NJ: Prentice Hall.
Spelke, E. S., & Tsivkin, S. (2001). Initial knowledge and conceptual change: Space and number. In M. Bowerman & S. Levinson (Eds.), *Language Acquisition and Conceptual Development*. Cambridge, UK: Cambridge University Press.
Starkey, P., & Cooper, R. G. (1995). The development of subitizing in young children. *British Journal of Developmental Psychology, 13*, 399–420.
Strauss, M. S., & Curtis, L. E. (1984). Development of numerical concepts in infancy. In C. Sophian (Ed.), *Origins of Cognitive Skills* (pp. 131–155). Hillsdale, NJ: Erlbaum.
Trick, L., & Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently. A limited-capacity preattentive stage in vision. *Psychological Review, 101*, 80–102.
Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience, 16*, 1493–1504.
Weyl, H. (1949). *Philosophy of mathematics and natural science*. Princeton, NJ: Princeton University Press.
Wood, J. N., & Spelke, E. S. (2005). Chronometric studies of numerical cognition in 5-month-old infants. *Cognition, 97*(1), 23–29.
Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*, 155–193.
Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology, 24*, 220–251.