Cognition 168 (2017) 243–255

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Do analog number representations underlie the meanings of young children's verbal numerals?

Susan Carey [a,*], Anna Shusterman [b], Paul Haward [a], Rebecca Distefano [a]

[a] Harvard University, United States
[b] Wesleyan University, United States

ARTICLE INFO

Article history:
Received 16 August 2015
Revised 2 June 2017
Accepted 19 June 2017
Available online 18 July 2017

Keywords:
Conceptual development
Number representations
Analog number system

ABSTRACT

Children learn to count, and even learn the cardinal meanings of the first three or four verbal numerals ("one" through "three" or "four"), *before* they master the numerical significance of counting. If so, it follows that the cardinal meanings of those first few numerals cannot be derived, initially, from their place in the count list and the counting routine. What non-verbal representations, then, support the cardinal meanings of verbal numerals before children have mastered how counting does so? Four experiments addressed the commonly adopted assumption that in the earliest period of learning the meanings of number words, children map verbal numerals to regions of the analog number system (ANS), a system of representation with numerical content that is widely attested in animals and in human infants. Experiment 1 confirmed that children who know what "three" means, but who do not yet know what "four" means, and do not yet know how counting represents number, can be easily taught the meaning of "four," if they are trained to indicate sets of four when they are paired with a series of sets that contrast numerically with four. If children learn "four" by mapping the word to an ANS representation of sets of four, and if such ANS value-to-word mappings underlie the meanings of other known numerals early in development, then analogous teaching should enable young children to establish a ANS value-to-word mapping for between "ten" and sets of 10 as specified by the ANS. Furthermore, the ease of learning should be a function of the ratio of the number of individuals in the comparison set to 10. Three further experiments tested these hypotheses by attempting to teach young Cardinal Principle-knowers the meaning of the word "ten," under the same training conditions "three-"knowers are easily taught the meaning of "four". The children learned which *picture* in each training pair had "ten." However, test trials with novel animals and spatial configurations showed that they had failed to learn what *set sizes* should be labeled "ten", even when, after training, they were asked to indicate a set of 10 vs. a set of 20 or 30 (well within the ratio sensitivity of the ANS even early in infancy). Furthermore, there was no effect of ratio on success during test trials. These data provide new evidence that ANS value-to-word mappings do not underlie the meanings of number words early in development. We discuss what other non-verbal representations might do so, and discuss other ways the ANS may support learning how counting represents number.

1. Introduction

Mathematics does not come for free by virtue of being born a human being. Historically, the cultural construction of mathematics began with arithmetic (Dantzig, 1967; Ifrah, 1985). As the foundational concepts in arithmetic are the positive integers, a good place to start in understanding the *ontogenesis* of mathematics is to account for the ontogenetic origin of representations of the positive integers. In the first systematic attempt at such an account, Piaget (1952) argued that integer representations must await the logical developments of concrete operational thought. He offered non-conservation of number by preoperational children as evidence that concepts of integers do not become available until age 5 or 6.

In the first major re-evaluation of Piaget's position, Gelman and Gallistel (1978) countered that verbal counting, when deployed in accordance with the counting principles of stable order, 1–1 correspondence and the cardinality principle, constitutes a representation of at least a finite subset of the positive integers. Gelman

* Corresponding author.
E-mail address: scarey@wjh.harvard.edu (S. Carey).

http://dx.doi.org/10.1016/j.cognition.2017.06.022
0010-0277/© 2017 Published by Elsevier B.V.

and Gallistel provided evidence for mastery of these three counting principles by young 2-year-olds, and proposed that learning to count is supported by an innate "numeron" list, used in accord with the counting principles to represent cardinal number. Learning to count in a natural language, according to this hypothesis, requires only that the child identify what ordered list of words should be mapped to the innate numeron list. Thus, Gelman and Gallistel, in contrast to Piaget, argued that the positive integers are innate.

Subsequent work has undermined the empirical support for an innate count list. Two year olds indeed know the count routine, and deploy it in stable order and in 1–1 correspondence to the individuals counted, but much evidence suggests they do not know that the last word reached in a count represents the cardinal value of the set (the cardinal principle) until months or even years later (Fuson, 1988; Le Corre, Van de Walle, Brannon, & Carey, 2006; Mix, Huttenlocher, & Levine, 2002; Sarnecka & Lee, 2009; Siegler, 1991; Wynn, 1990; Wynn, 1992; see Carey, 2009 for review). Rather, they assign numerical meaning to the verbal numerals in a piecemeal way, first learning what "one" means (i.e., become "one"-knowers), then some 6 months later become "two"-knowers, then "three"-knowers, and then "four"-knowers. Children who know only the meanings of some of the numerals between "one" and "four" are designated "subset-knowers," for they know the cardinal meanings of only a subset of the numerals they can recite. Middle-class, English learning, children become cardinal principle-knowers (CP-knowers) around age 3 ½ to 4 ½, and can then use the count routine to assign a cardinal value to any words in their known and practiced count list (Gunderson, Spaepen, & Levine 2015; Le Corre & Carey, 2007; Sarnecka & Lee, 2009).

If we accept that children in the subset-knower period do not know the significance of counting, it follows the cardinal content of the words "one" through "four" in the subset-knower cannot be provided by their role in a counting procedure constrained by the counting principles (e.g., the numeral "four" cannot receive its meaning by virtue of being the fourth word in the count list). This conclusion raises an important question: if the meaning of the first verbal numerals is not provided by their role in counting, how *do* they get their numerical content?

One likely source of number word meanings is antecedently available *non-verbal* representations of number. It is very difficult to see how meanings for number words might be constructed entirely from representations with no numerical content. Indeed, non-human animals, human infants, children, and adults share two quite different evolutionarily ancient systems of non-verbal representations with numerical content: (1) the analog, or approximate, number system (ANS); and (2) parallel individuation (PI), a structure in which working memory models of small sets of individuals are constructed with one symbol in the model for each individual in the set. These are the only two systems of nonverbal representations with numerical content for which there is evidence in non-human animals and very young human infants (see Carey, 2009; Feigenson, Dehaene, & Spelke, 2004, for review).

The ANS consists of analog representations that are a linear or logarithmic function of the cardinal values of the set represented. These representations express cardinal values only approximately. One signature of the ANS is that magnitudes are discriminated one from another on the basis of their ratios; thus, discriminability accords with Weber's law and exhibits scalar variability (the standard deviation of multiple estimates of a the number of items in a set is a linear function of that set's cardinal value). ANS representations support many different arithmetical computations, including numerical comparison, addition, subtraction, multiplication, and division (see Carey, 2009; Dehaene, 2011; Gallistel, 1990, for review).

PI, a second preverbal system with numerical content, consists of working memory representations of small sets of individuals. The symbols in this system represent individuals (e.g., a set of three crackers is represented CRACKER, CRACKER, CRACKER, probably iconically for each cracker). Unlike the ANS, the PI working memory system is not a dedicated number representation system, *nor are there any symbols* that represent cardinal values in these models; there are only symbols for individuals, held in working memory. The numerical content in PI is implicit, carried by the computations that ensure that the symbols in a working memory model stand in one-to-one correspondence with the individuals in the sets modeled, and the computations that allow models to be compared on the basis of one-to-one correspondence to determine numerical equivalence. There is a strict upper limit to capacity of working memory, a function of the number of encoded individuals, the complexity of the representations of individuals, and the complexity of the computations to which the models serve as input (Brady & Alvarez, 2015; Xu & Chun, 2009; Zosh & Feigenson, 2009). For twelve-month-olds, this limit on working memory representations of single sets is three distinct, perceptually simple individuals (Feigenson & Carey, 2003; Ross-Sheehy, Oakes, & Luck, 2003); with development this capacity expands a bit, reaching a limit of four or five in older preschoolers (Starkey & Cooper, 1995).

Although researchers for the most part have abandoned the hypothesis of an innate numeron list and counting routine, almost all agree with Gelman and Gallistel's crucial insight that the count list, deployed in accord with the counting principles, constitutes a representation of at least a subset of the positive integers. Furthermore, neither preverbal representational system, on its own, is capable of expressing integers: the ANS because it only approximates cardinal values and does not naturally implement the successor function, and PI because it contains no symbols for cardinal values and has a capacity limit on the size of sets it can represent. Thus, much work in the field concerns the process through which children learn the cardinal principle, as the counting principles ensure that verbal numerals *do* represent integers. All theories posit innate numerical resources *in addition to PI and the ANS*; examples include an innate successor function (Leslie, Gelman, & Gallistel, 2007); an innate tally system based on the iteration of 1 (Leslie, Gelman, & Gallistel, 2008); and quantification in natural language morpho-syntax and semantics (Almoammer et al., 2013; Barner, Libenson, Cheung, & Takasaki, 2009; Bloom & Wynn, 1997; Le Corre, Li, Huang, Jia, & Carey, 2016; Sarnecka, Kamenskaya, Yamana, Ogura, & Yudovina, 2007). All suggest that some process of combining or aligning antecedently independent representational systems is likely involved (e.g., Carey, 2009; Leslie et al., 2007, 2008; Spelke, 2003). In order to evaluate these different proposals, we need to know what *non-verbal* representations underlie children's meanings of the first number words, for those meanings will play a central role in the construction of explicit representations of positive integers.

Here we assume that the two well attested systems of representation are the only *preverbal* representations with numerical content available to underlie the meanings of the words "one" through "four" in the subset-knower stage. This is because there is no *evidence* for an innate successor function or an innate tally system based on the iteration of 1. Our question is whether one system, or both, do so, and how. Several writers presuppose (Bugden & Ansari, 2011; Odic, Le Corre, & Halberda, 2015; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013) and/or explicitly argue (Dehaene, 2009; Gallistel & Gelman, 2000; Piazza, 2010; Starr, Libertus, & Brannon, 2013; Verguts & Fias, 2004; Wagner & Johnson, 2011) that the ANS provides such meanings, and it does so though the creation of mappings of each number word "one" through "four" with an ANS region (as ANS values can be specified at different levels of precision). We call this the "ANS value-to-word mapping" hypothesis. Fig. 1, taken from Gallistel and Gelman (2000) illustrates this assumption. The top part of the figure illustrates the operation of the non-verbal ANS; the bottom illustrates a bidirectional mapping between the first 8 verbal numerals and ANS values that is the starting point for Gallistel and Gelman's (2000) proposal for the transition to representations of integers.

Rats can learn rules about quantities formulated in terms of ANS values (e.g., press the left bar 25 times, and then switch to the right bar; Platt & Johnson, 1971). Furthermore, Platt and Johnson provide good evidence the ANS provides the numerical content in the rule, as the rats' number presses exhibit scalar variability. Thus, it is certainly possible that ANS values alone can support the meaning of the word "four," similarly providing the numerical content to generalizations such as "sets of 4 are called "four." Beyond possible, this proposal has considerable plausibility. Adults have mapped at least some verbal and Arabic numerals to specific ANS values (e.g., Dehaene, 2011; Izard & Dehaene, 2008; Moyer & Landauer, 1967; Sullivan & Barner, 2012; Whalen, Gallistel, & Gelman, 1999). The ANS is an evolved system of representation with numerical content; clearly, integrating culturally constructed explicit representations with ANS representations is one important way that verbal and written representations gain numerical meaning.

However, the above studies do not speak to *when* in development a mapping between some verbal numerals and ANS values is constructed, nor do they establish that such a mapping constitutes the *meaning* of the verbal numerals at *any* point in development. In spite of the plausibility of the hypotheses that mappings between the ANS and verbal numerals play an important role in supporting the meanings of number words from the beginning of learning, as well as in the induction of the counting principles, Le Corre and Carey (2007) argued on logical grounds that these hypotheses are likely to be wrong. First off, the meanings of the numerals "one" through "four" in the subset-knower period must be able to support the transition to CP-knower, during which children connect the counting routine with cardinal meanings for number words above 4. The ANS does not naturally support the induction of the counting principles because it does not contain a representation of exactly one. Furthermore, within the ANS, quantities are compared by their ratios rather than their absolute differences, and adjacent cardinal values beyond 8 or 9 cannot even be discriminated. Thus, the numerical representations of the ANS obscure rather than facilitate the use of counting to implement the successor function, and implementing the successor function is necessary for counting to support integer meanings for verbal numerals.

Le Corre and Carey (2007) also provided empirical evidence that undermined the hypothesis that ANS values form the basis of numeral meanings that scaffold the transition to CP-knower. Asked to estimate the number of dots in visual arrays without counting, all subset-knowers and many young CP-knowers showed no evidence for any such mappings to numerals *above four*. When these children's verbal estimates of cardinal values were plotted as a function of the presented set sizes in the range of 6–10, the slope was 0. Providing the same estimates for sets of 10 as for sets of 6 suggests that they had not mapped "ten" and "six" to ANS representations of approximately 10 and 6. Le Corre and Carey (2007) concluded that *these* mappings could play no role in the transition to CP-knower, since they observed a six-month age gap between CP-knowers who had made such mappings and younger CP-knowers who had not yet done so. Subsequent studies have shown that *some* subset-knowers and young CP-knowers do have positive slopes in their functions between set-sizes and numerical estimations over the whole range of 1 through 10 (Cheung, Slusser, & Shusterman, 2016; Gunderson et al., 2015; Odic et al., 2015); however, not all subset-knowers or even young CP-knowers do (Gunderson et al., 2015; Le Corre & Carey, 2007). Thus, whatever kind of mapping between the ANS and verbal numerals supports this function is not *necessary* for the transition to CP-knower. Furthermore, no study has found a *linear* mapping over the whole range of 1 – 10 in the *subset-knower* stage, such that a bidirectional ANS-to-number-word mapping up to 10 could possibly facilitate the CP transition.

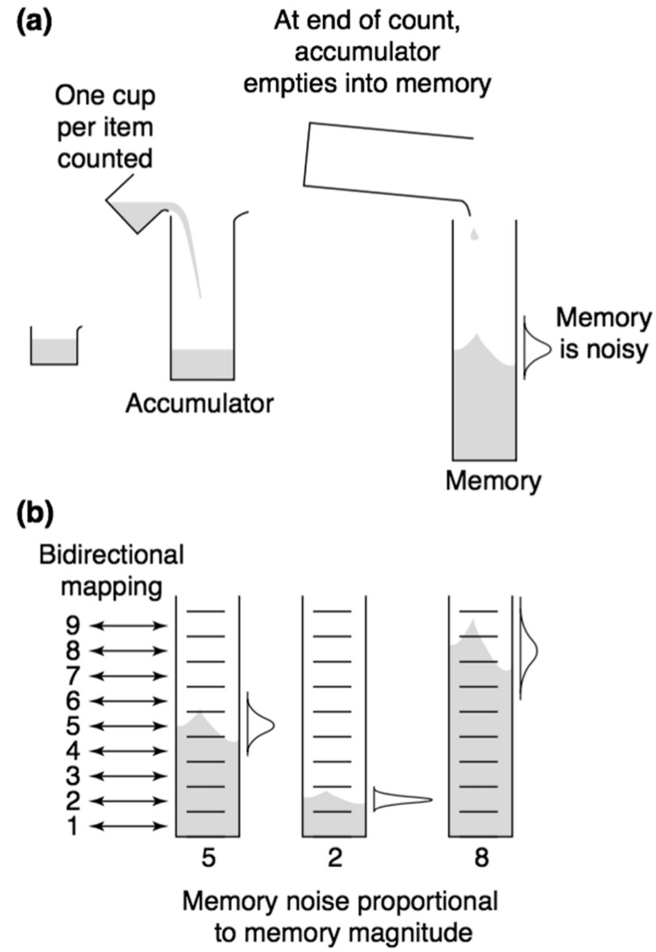

Fig. 1. Figure illustrating hypothesized bidirectional verbal numeral-ANS value mapping. From Gallistel and Gelman (2000).

Still, many subset-knowers and all young CP-knowers *do* have linear mapping between set sizes and verbal numerals in the range of "one" through "four." In principle this could reflect a mapping between words and ANS values for these words that are learned in the subset-knower phase. Indeed, what alternative *is* there? In the next few paragraphs, we lay out an alternative to the ANS for the initial meanings of the first number words—enriched parallel individuation. The arguments that follow are meant only to make the enriched PI proposal plausible, not necessarily provide convincing evidence for it. The studies in this paper then explicitly test predictions that should hold if the ANS underlies the initial meanings of numerals "one" to "four."

Parallel individuation operates over the small set sizes that are lexicalized in the subset-knower period, but does not contain symbols for cardinal values of those sets. Rather, as mentioned above, its numerical content is implicit. Even more importantly, the representations in the PI system are working memory models of immediately present (although possibly occluded) sets; they are not long term memory representations that could underlie the meanings of words. In response to these properties of the PI system, Le Corre and Carey (2007; also Carey, 2009) proposed that PI representations would need to be enriched to support the meanings of the first four numerals, and could be as follows: During the subset-knower stage, children could create *long term memory* representations of a particular set consisting of a single individuals (e.g., {me} or {*i*}), a particular pair of individuals (e.g., {Mommy, Daddy} or {*j*, *k*}), a particular tripleton ({these three fingers} or {*m,n,o*}), and a particular quadrupleton (e.g., {Mommy, Daddy, Bobby, me} or {*a, b,c,d*}). These long term memory models are mapped to the number words "one" through "four", respectively, and are deployed according to a procedure that determines the number word that applies to an attended set by finding which of those long term memory models stands in 1–1 correspondence to it. Such a representational system makes use of no numerical content or computations not attested in PI working memory models, but is enriched by long term memory models. Upon hearing this idea explained in a class, a colleague reported that for a couple of months, his 3-year-old daughter always commented on sets of three thus: "there are three kittens, mommy, daddy, me, three;" "Look, three cement mixers, mommy, daddy, me, three."

One suggestive source of evidence that enriched parallel individuation might actually underlie the meanings of the first few verbal numerals is linguistic. Kayne (2016) summarizes linguistic evidence that "one book" means "a single book," that "two books" means "two (book and book)", "three books" means "three (book and book and book)" and "four books" means "four (book and book and book and book)." Kayne presents evidence that this expansion of the plural in terms of PI representations no longer occurs with numerals higher than four. Relatedly, the morphology of the first numerals sometimes reflects the long term memory models of enriched parallel individuation (e.g., in Daw, "one" is lexicalized with the same word that means "unity", "two" is lexicalized with a compound word that means "eye-quantity" and "three" with a compound that means "rubber tree seed quantity;" Barner, 2017).

Le Corre and Carey (2007) commented that the enriched PI proposal predicts that the subset-knower period would be restricted to "one" through "four" alone, followed by a discontinuity above "four." However, many have pointed out that the observed discontinuity can be accommodated by the ANS value to number word mapping hypothesis as well, on the grounds that the ratio sensitivity of young children's ANS is 3:4, at best, making it difficult for the ANS to discriminate the values that differentiate the meanings of numerals above five (e.g., Dehaene, 2009). In addition, the frequency of small number words in cardinal contexts in input to toddlers is vastly greater for the numerals "one" through "four" than for "five" and higher (as recently verified by Le Corre et al., 2016, in a corpus analysis of parental speech to toddlers). Ramscar, Dye, Popick, and O'Donnell-McCarthy (2011) showed that these two factors (greater frequency of lower numerals, worse discrimination among higher numbers) together predict seeming discontinuity in the ease of learning the meanings of the smaller numerals, on the one hand, and those of the higher numerals, on the other, with a discontinuity at 4. Thus, there is a need for additional evidence in order to adjudicate between the hypotheses that the ANS or enriched PI (or some other alternative representational scheme) is more likely to support the meanings of verbal numerals, before the numerals' place in the count routine can do so.

The present studies are designed to straightforwardly test the ANS value-verbal numeral mapping hypothesis concerning the meanings of verbal numerals prior to the CP induction, and the hypothesis that such mappings play a necessary role in that induction. If so, then young CP-knowers should have an over-hypothesis that any verbal numeral expresses an ANS value. Over-hypotheses about word meanings are broad hypotheses or constraints about the properties of the meaning of a word that influence the induction of word meanings. Young children learn over-hypotheses concerning the meanings of new count nouns or new mass nouns, such that they can learn the meanings on just one or two encounters (Soja, Carey, & Spelke, 1991; Yu & Smith, 2007). Such an over-hypothesis should make it possible to teach children a new mapping between a relatively infrequent number word and an ANS value, so long as that value is contrasted with other ANS values differing by ratios that vastly exceed the resolution of the ANS.

In support of this prediction, Huang, Spelke, and Snedeker (2010) taught "three"-knowers the meaning of the word "four" from just a few pairings of "four" with sets of 4, when those sets of 4 were paired with sets clearly distinguished from 4 by the ANS. It follows that, if the transition from subset- to CP-knower involves mapping verbal numerals *higher than 4* to regions of the ANS, then it should be possible to teach a young CP-knower an association between the word "ten" and the relevant region of the ANS representation, given unambiguous input such as that in Huang et al. Specifically, young CP-knowers, in possession of the relevant over-hypothesis, should easily be able to learn what ANS value "ten" expresses, if the pairing of "ten" with sets of 10 is contrasted with other set sizes that differ from 10 by the same ratios that the contrast set sizes in Huang et al. differed from 4.

We present a set of experiments that test this prediction. By age 3 ½, the ANS system of children can discriminate sets that differ by a ratio of 3:4 (Halberda & Feigenson, 2008), so training sets of 10 contrasted with sets of 5, 7, 15, 20, and 30 all fall within the child's ANS discrimination capacity. Furthermore, if ANS values are the sole source of meaning for verbal numerals during the subset-knower phase, as well as the sole source of numeral meanings, along with the counting principles, immediately after the induction of the cardinal principle, then the ratio signature of the ANS should be apparent in children's learning from training about which labels apply to which sets. Sets labeled "ten" should be discriminable from other sets according to the ratio between that set and 10. That is, the contrast between 10 and 30 (1:3) should be the easiest to learn, followed by 10 vs. 5 = 10 vs. 20 (1:2), followed by 10 vs. 7 = 10 vs. 15 (1:7). The present experiments test these hypotheses.

2. Experiment 1

Experiment 1 is a replication of Huang et al. (2010), to ensure we would succeed, as they did, in teaching "three"-knowers the meaning of the word "four" from a few pairings of the word with sets of four when these sets contrasted with other set sizes. Unlike the earlier study, during the demonstration portion, when a card was introduced with four animals, the child was told it had "four, not one, two or three, not five, six, or seven, but four." This provided lexical contrasts to the word "four" and unambiguously conveyed that "four" refers to a specific set size, and during the training trials, the children were told the number on the comparison card as well as the target card ("four") – a second lexical contrast providing evidence that numerals refer to specific cardinalities. Finally, we included three sets of training trials instead of just one as in Huang et al. (2010), increasing children's opportunity to learn "four."

2.1. Method

2.1.1. Participants

The first 21 3-year-olds classified as "three"-knowers by Give-a-Number participated (10F; 36–44 months; M = 39 months). Two were excluded from analysis because they did not complete the Teach "Four" task.

Families were recruited from public birth records. Volunteers were largely middle class with a stay-at-home parent. Ethnicity for all four experiments was approximately 70%, non-Hispanic White and 9% Hispanic, with the remaining 21% of the sample comprising Native American, Asian, Native Hawaiian, and African-American participants. All children had English as their primary language, although some also spoke additional languages. Each child received a toy and each family received a five-dollar travel reimbursement.

2.1.2. Screening tasks

Children's knowledge of the count list and counting routine were assessed by a Counting task, and "three"-knowers were identified by the Give-A-Number task. In the counting task, ten toy fish were placed in a line and the child was asked to count them while the experimenter pointed to each fish. All children in this paper could count to 10 without error.

The Give-A-Number task used ten fish arranged in a jumbled pile. Beginning with one fish, children were asked to place a certain number of fish into a bowl. After each the trial, the experimenter placed the fish back in the pile. The experimenter then asked the child for two fish and so on, following a titration method in ascending order through the number six. If a child produced an incorrect response, the experimenter took the fish out and placed them in a line for the child to count. Once the child successfully counted the fish, the experimenter gave them a second attempt at placing the requested number in the bowl by saying "but I wanted X fish, can you put X fish in the bowl." The child's response was recorded as a second trial at this set size. If the child produced the correct amount of fish on the second try, the experimenter then placed the fish back in the pile and asked for the next number in the count list. If the child failed on both attempts, the experimenter placed the fish back in the pile and asked for the number below it. The task ended with the highest numeral the child produced correctly at least 2 of 3 times.

Each child's knower-level was determined as in Wynn (1992) and Le Corre et al. (2006): the highest numeral for which the child created the correct set size at least 2 of 3 times, provided the child did not produce this same set size for larger numerals on more than two-thirds of trials. Specifically, the child was credited with being a "three" knower, if they gave 3 fish at least two out of three times when asked for "three," and avoided giving 3 fish on at least two of three trials when asked for larger sets (e.g., "four" or "five").

2.1.3. Teach "Four"

The Teach "four" task had two parts: Training and Test.

2.1.3.1. Materials. Stimuli for the training component were two 11″ by 14″ demonstration cards, one with four identical pigs and one with four identical chickens. Stimuli also consisted of seven pairs of 11″ by 14″ training cards, each pair consisting of two sets of the same kind of animal, i.e., goats, pigs, horses, sheep, cows, chickens, or cats. One card in each pair had four animals, and the other had a contrasting number of animals (either sets of 1, 2 or 3 animals, i.e., sets labeled by number words known by our "three"-knower participants, or critical contrasts of 4 with sets of 5, 6, 10, or 16 animals, i.e., labels unknown by "three"-knowers). As in Huang et al. (2010), the pairs of sets on the training cards were not controlled for continuous extent; all of the items were roughly 1″ in diameter, such that the summed area covered by the animals was proportional to number, whereas in the test trials, number and spatial extent were deconfounded by equating total surface area across the two sets within a pair. Test trials used novel animals in novel spatial configurations. See Fig. 2 for an example of one training pair and one test pair.

2.1.3.2. Training procedure. Training unfolded in two steps: demonstration and training trials.

2.1.3.3. Demonstration. Children were presented with two cards, one at a time, each with four animals (pigs and chickens). The experimenter pointed to the card and stated the number of animals (e.g., "this picture has four pigs"), then provided lexical contrast ("not one, two, or three pigs; not five, six, or seven pigs"), and restated the number of animals on the card ("but four pigs"). The child was then asked how many animals were on the card. All children replied "four."

The demonstration then proceeded to the seven training pairs of cards, each pair consisting of one card with a set of 4 animals and one with a contrasting set (1, 2, 3, 5, 6, 10, or 16 animals). In this part, the experimenter emphasized to the child that the game is a guessing game and discouraged the child from counting. Each child was shown the cards in one of four randomized orders. A demonstration run through the pairs informed the child, for each pair, which card had four animals and the number of animals on the other card ("This card has four horses; this card has ten horses, but this card has four horses"). The child was then asked to point to the card with 4 horses. Errors during demonstration trials were infrequent and corrected.

2.1.3.4. Training runs with feedback. The training session then proceeded to three training runs through the same seven pairs of cards as in the demonstration, placed side by side. For each pair, the experimenter asked the child to indicate to the picture with 4 animals. For example, for cards depicting 4 horses and 10 horses, the child was asked, "Which card has four horses?" The order of the pairs was randomized across the three training runs, and the side with 4 animals was counterbalanced. If the child pointed to the correct card, positive feedback was provided ("Great! That card does have four horses."). If the child pointed to the incorrect card, the experimenter provided feedback ("Good try, but *this* card has four horses"). Each child was given all three blocks of training trials unless they correctly identified the card with 4 animals on six or seven of the seven training pairs on both of the first two training runs.

2.1.3.5. Test trials without feedback. There were 10 pairs of 11″ by 14″ test cards, depicting novel animals not used in the training set. On three "known-known" trials, both cards displayed set sizes for which both verbal labels were known by "three"-knowers (1v2, 1v3, and 2v3). The experimenter always asked for the larger number in the pair. These pairs simply ensured that children were still engaged in the task, and helped to balance roughly the number of trials where the correct set was the larger option with trials where the correct set was smaller. On three "trained-known" trials, children saw a card with 4 animals and a card with 1, 2, or 3 animals (set sizes with known labels). The four critical "trained-unknown" pairs contrasted a set of 4 with a set of 5, 6, 10, or 16 (set sizes with unknown labels). The child was always asked to indicate the set with "four." Unlike in the training and practice sessions, feedback was not provided. Instead, the experimenter provided a neutral, yet encouraging response ("Okay!").

2.1.3.6. Analysis. All analyses in this paper fit logistic regression models to the binary responses of each child on each trial (correct/incorrect) using the lme4 package (Douglas, Maechler, Bolker, & Walker, 2015) in the R statistical language (http://www.r-project.org/). For tests against chance, we analyzed the significance of a model with a single intercept: model = glmer (Response $\sim$ 1 + (1|ID)), or model = glm(Response $\sim$ 1) if each child received only one trial for that analysis. When testing for the significance of a particular variable, we ran a similar model with the addition of a fixed effect: e.g., for the variable Ratio:

Fig. 2. Examples of training pairs and test pairs from Experiments 1–3.

model = glmer(Response ~ Ratio + (1|ID)). For all analyses, we included subject as a random effect to account for the fact that each child received multiple trials. These nonparametric analyses yield the same results as parametric analysis conducted on each child's % correct across the trials they received. For simplicity, we include % correct statistics in addition to the coefficients generated from the non-parametric analysis.

2.2. Results

2.2.1. Training trials

Using a 'passing' criterion of 6 or 7 correct out of 7 trials, only one child passed both the first and second training runs, and thus was not presented with the third training run. Children performed well above chance on trained-known trials on their last two training runs (two runs of 4 vs. 1/2/3, $M = 79\%$ correct, $b = 1.4$, $SE = 0.3$, $z = 4.8$, $p < 0.001$, 1-tailed[1]). This is as expected for "three"-knowers, since they know the meanings of "one," "two," and "three" and take all other numerals to contrast with the sets labeled by these words. Thus, success on these trials does not show that children had learned anything about the meaning of the word "four."

Performance was also above chance on the critical trained vs. unknown trials (two runs of 4 vs. 5/6/10/16: $M = 69\%$ correct, $b = 0.9$, $SE = 0.3$, $z = 3.3$, $p < 0.001$). Replicating Huang et al. (2010), children quickly learned to apply "four" to the particular sets that were demonstrated on the practice trials, but performance was well below ceiling. Huang et al. observed this learning on the first set of training trials, whereas for the most part, our participants required three.

2.2.2. Test trials

Children could succeed on the practice trials by simply remembering, for each particular pair, which set to pick. The test trials involved new animals, in novel configurations, and deconfounded total surface area and number. These trials thus provide a strong test of the hypothesis that children learned what set size is labeled by "four." Children performed at ceiling on known-known trials (one run of 1vs.2/2vs.3/1vs.3: $M = 98\%$ correct, $b = 4.0$, $SE = 1.0$, $z = 3.99$, $p < 0.001$), and well above chance on

[1] All tests that compare performance to chance are 1-tailed, as there is no reason to expect that a child taught that "four" labels sets of 4 (or that "ten" labels sets of 10 in Experiments 2 and 3) objects would systematically indicate the *other set size* in the pair when asked to indicate which had "four" or which had "ten."

trained-known trials, (one run of 4 vs. 1/2/3: $M = 86\%$ correct, $b = 1.8$, $SE = 0.4$, $z = 4.8$, $p < 0.001$), as expected of "three"-knowers (who know, without teaching, that "four" cannot apply to sets of 1, 2 or 3). Performance on the critical trained-unknown trials (one run of 4 vs. 5/6/10/16), displayed on Fig. 3, was above chance ($M = 61\%$ correct, $b = 0.4$, $SE = 0.2$, $z = 1.8$, $p = 0.04$), replicating Huang et al. (2010).

2.2.3. Effects of ratio

Unlike Huang et al., we did not find better performance on high-ratio trials than low-ratio trials (see Fig. 3). A mixed model logistic regression with ratio as a fixed effect, subject as a random effect, and performance (correct/incorrect) as the dependent variable showed no effect of ratio in the four critical trial pairs (4 vs. 5/6/10/16) on performance on the test trials; this model did not differ significantly from the same logistic regression model without the inclusion of ratio as a fixed effect ($F(3,72) = 0.79$, $p = 0.50$).

Huang et al. generously made their raw data available to us. A mixed model logistic regression examined the fixed effects of experiment (Huang et al., Experiment 1) and trial pair (4v5, 4v6, 4v10, and 4v16), with subject as a random effect, on performance on the critical test trials (correct/incorrect). There was no effect of experiment ($F(1,88) = 1.55$, $p = 0.22$) and no interaction between experiment and ratio difficulty ($F(1,88) = 1.39$, $p = 0.24$). There was a marginal main effect of ratio difficulty ($F(3,88) = 2.65$, $p = 0.054$); post hoc comparisons showed this effect to be due to the difference between performance on the 4v5 pair (46% correct) and the 4v16 pair (77% correct). We conclude that "three"-knowers can robustly and easily be trained to identify sets that should be labeled "four" under the conditions of training common to Experiment 1 and to Huang et al. (2010).

What numerical representations underlie the meaning of the newly learned verbal numeral "four?" Huang et al. argue that the effect of ratio suggests that children have associated "four" with the analog magnitude representation of approximately 4 individuals. However, this distance effect would also be predicted if non-verbal counting or enriched parallel individuation underlie the meaning of "four," because errors are more likely for close comparisons than far ones in these representational systems as well. Another form of evidence is therefore needed to determine the nature of the non-verbal representations that subset-knowers and young CP-knowers map to verbal numerals.

If, under the brief training in this study, children map the verbal numeral "four" to an analog magnitude symbol for approximately 4, then they must more generally have the hypothesis available to them that number words correspond to cardinal values specified by analog magnitudes. If so, young CP-knowers should be able to map other verbal numerals, for example "ten," to analog magnitude representations, given comparable training regimes (e.g., pairings of "ten" with sets of 10, contrasted with other sets by large ratios), and if ANS/accumulator representations are the sole source of meaning for verbal numerals, performance at identifying which set has 10 should be a strict function of ratio between the comparison set's number of elements and 10 (the target set's number).

3. Experiment 2

In Experiment 2, we attempted to teach young cardinal principle knowers (CP-knowers) the meaning of the word "ten" using the same training and testing procedure from Experiment 1, in which we successfully taught "four" to "three"-knowers. As in Experiment 1, number was confounded with spatial extent during training and de-confounded during test (see Fig. 1).

3.1. Method

3.1.1. Participants

The first twenty 3-year-olds children classified as CP-knowers by the Give- A-Number task participated in the Teach "Ten" task ($M = 43.6$ months, range: 35–47 m; 10F, 10 M). CP-knowers successfully give 1, 2, 3, 4, 5, and 6 fish on the Give-a-number task.

3.1.2. Training procedure

Two demonstration cards, each with 10 animals (pigs or fish) approximately 1″ in diameter, were presented to the child one at a time. The child was told, for example, "This card has ten pigs, not one, two, or three pigs, not seven, eight, or nine pigs, but ten," and then asked how many pigs were on the card. All answered "ten." This was followed by a demonstration of the six sets of training pairs (10 vs. 3, 5, 7, 15, 20, and 30). The child was shown each pair (e.g., 10 cats and 20 cats), and told, "This card has ten cats. This card has twenty cats, but this card has ten cats. Which card has ten cats?" All children were correct on these demonstration trials. Then, as in Experiment 1, three runs of training trials ensued. In each trial, the child was shown one of the 6 pairs from the demonstration set and asked, e.g., "which card has ten dogs?" All children got three training runs with feedback, unless they were correct on 5 of 6 trials on each of the first two training runs.

3.1.3. Test trials

There were six pairs of cards for the test trials, depicting the same contrasts between sets of 10 and other quantities, but with different animals and different spatial configurations from the training cards. On test trials, the arrays were equated for total surface area of the individuals (see Fig. 1). As in Experiment 1, there was no feedback on test trials.

3.2. Results

3.2.1. Training trials

Using a 'passing' criterion of 5 or 6 correct out of 6 trials, three children passed both the first and second runs, and thus were not given a third training run. As expected, children easily identify which set was 10 compared to a set of 3 ($M = 95\%$ correct, $b = 2.9$, $SE = 0.7$, $z = 4.1$, $p < 0.001$), but these were CP-knowers and even "three"-knowers can do this by mutual exclusivity from their knowledge of "three." On the critical trained-unknown trials (10 vs. 5/7/15/20/30), children performed well above chance ($M = 67\%$ correct, $b = 0.7$, $SE = 0.2$, $z = 3.9$, $p < 0.001$) on the last two training runs. However, children may have simply learned, trial pair by trial pair, which particular configuration is called "ten." The test trials allow us to assess whether children had created a mapping between the word "ten" and cardinal values of sets of 10, since the test trials involved novel animals in novel spatial configurations, with each pair controlled for total spatial extent.

3.2.2. Test Trials

Children in Experiment 2 robustly succeeded at the 3v10 (trained vs. known) comparison ($M = 95\%$, $b = 13.8$, $SE = 6$, $z = 2.9$, $p = 0.01$), but as above, this success could be driven by the knowledge that sets of three are labeled "three". If success on the Training trials reflect a newly learned mapping between the word "ten" and cardinal values of sets of 10, then children should succeed on critical test trials. Furthermore, any differences in performance across training sets should be strictly a function of ratio: i.e., 10 vs. 30, easiest ratio (1:3); 10 vs. 20 = 10 vs. 5, medium difficulty (1:2); 10 vs. 7 = 10 vs. 15, hardest ratio (1:1.5). Neither of these predictions was born out. Fig. 4 displays the performance on the test trails for each critical test pair (10 vs. 3/5/7/15/20 /30) for each of Experiments 2 and 3, grouped by ratio. Children were at chance

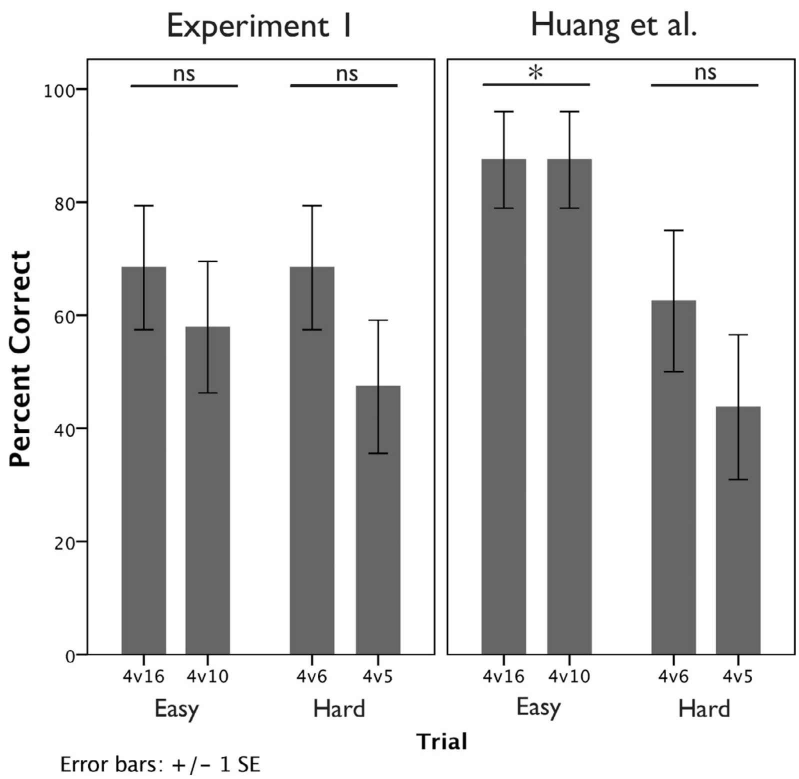

Fig. 3. Performance on the critical test trials (4 vs. 5, 6, 10, 15) in Experiment 1 and in Haung et al. P values represent comparisons to chance using the logistic regression analyses explained in the text: ns (non-significant), * (p < 0.05, 1 tailed), ** (p < 0.01, 1 tailed), *** (p < 0.001, 1-tailed).

for the critical trained vs. unknown test trials overall (10 vs. 3/5/7/15/20 /30): *M* = 56%, *b* = 0.2, *SE* = 0.2, *z* = 1.2, n.s. A mixed model logistic regression examining the fixed effect of ratio difficulty (easy, medium, hard) on performance (correct/incorrect), with subject as a random effect, revealed a marginally significant main effect of ratio ($F(2,97) = 2.99$, $p = 0.055$, 2-tailed), but performance was not a linear function of ratio: performance on the medium difficulty (1:2) ratios were above chance (70% correct, *b* = 0.88, *SE* = 0.41, *z* = 2.1, *p* = 0.02), whereas performance on the easiest ratio (1:3) and the hardest ratios (1:0.7) were each at chance (see Fig. 4). Furthermore, the two medium difficulty ratios (1:2) were not equivalent. Children succeeded at the 10 vs. 5 comparison (80% correct, *b* = 11.64, *SE* = 3.95, *z* = 2.95, *p* = 0.003) but failed at the 10 vs. 20 comparison (60%, *b* = 0.41, *SE* = 0.46, *z* = 0.89, n.s.). The only trained vs. unknown comparison that led to success in Experiment 2 was 10 vs. 5, in contrast to every other trial type (performance below chance at 7 vs. 10 and at chance for the others). Given the failure to pick out the sets of 10 in comparison to sets of 20 and 30, the success at 5 vs. 10, like that of 3 vs. 10, most probably reflects mappings of sets of 5 with the word "five" and sets of 3 to the word "three," plus knowledge that other number words, such as "ten," do not apply to these set sizes. If learning what magnitude "ten" maps onto contributed to this success, children should have also succeeded in saying which card had 10 birds or fish when the other choice was 20 birds or 30 fish.

The finding that young CP-knowers have likely mapped sets of 5 to some verbal numeral, sufficiently to separate it from the label "ten," is a new finding. In previous work, Le Corre and Carey (2007) and Shusterman et al. (2016) found linear positive slopes in some subset-knower's and all CP-knower's verbal estimates for sets from 1 to 4, but they did not test 5. Le Corre and Carey found slopes of 0 for verbal estimates of sets in the range of 6 to 10 in young CP-knowers, and concluded that, prior to the CP-transition, children constructed mappings between numerals and non-verbal number representations *only* for "one" to "four." Since children avoided mapping "ten" to sets of 5, the present finding extends the attested mapping to "five," a theoretically important result, since sets of 5 are beyond the putative range of parallel individuation. We seek to confirm this finding in the next two experiments, and turn to its interpretation in the general discussion.

4. Experiment 3

To explore how abject the failure to map "ten" to sets of 10 is, Experiments 3a and 3b attempted to make *more* salient to children the hypothesis that the mapping between "ten" and sets of 10 should be *numerical*. In Experiment 3a, we de-confounded number and spatial extent even during the training trials (Fig. 1) to highlight that "ten" refers to numerosity, not extent. In Experiment 3b, we supported children's learning of "ten" with their knowledge of counting. Given that CP-knowers know a numerical meaning of "ten" in the context of counting, perhaps highlighting the place of "ten" in the count list would help CP-knowers to draw on their existing knowledge in order to learn a mapping for "ten." Accordingly, in Experiment 3b, we used the procedure and materials of Experiment 3a, but whenever children chose the wrong set on a practice trial, the experimenter and the child together counted both sets.

4.1. Participants

Three-year-old children were screened for knowledge of counting and knower-levels as in Experiment 2. All children classified as CP-knowers participated in the Teach "Ten" task (Experiment 3a, *N* = 19, 9F, *M* = 43.5 months, range: 38–47 months; 9F; Experiment 3b, *N* = 21, 13F, *M* = 41.7 months, range 37–47 months). One additional CP-knower did not complete the teaching task and was excluded from analysis.

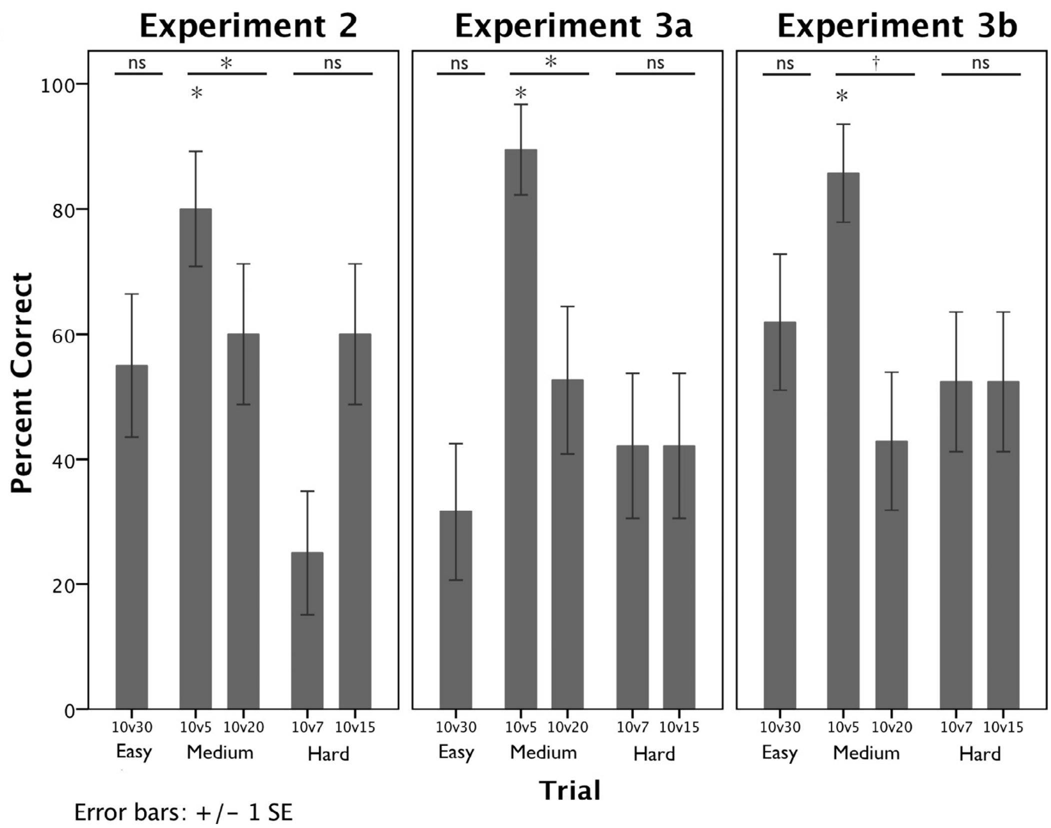

Fig. 4. Performance on critical test trial (10 vs. 5, 7, 15, 20, 30) in Experiments 2 and 3. P values represent comparisons to chance using the logistic regression analyses explained in the text: ns (non-significant), * ($p < 0.05$, 1 tailed), ** ($p < 0.01$, 1 tailed), *** ($p < 0.001$, 1-tailed).

4.2. Procedure

The training and testing portion of **Experiment 3a** were identical to Experiment 2, with the exception that in Experiment 3a, the total surface area of the sets of animals in each pair was equated (see Fig. 1) during the training trials as well as during the test trials. **Experiment 3b** was identical to Experiment 3a, except for additional counting feedback provided for an incorrect answer during the practice runs. If a child did not select the array with ten objects, the experimenter responded with "Good guess, but *this* one has ten. Can you count and make sure?" After the child counted ten, the experimenter responded, "Yes! THAT card has ten; I'll count and see how many this one has," proceeding to count the number of animals on the other card, as experimenter could count faster and more accurately than the child, especially on the larger sets. The experimenter then repeated how many animals were on the other card (e.g., "this card has seven bunnies"), and then pointed to the card with ten and said, "and this card has ten bunnies."

4.3. Results

4.3.1. Training trials

Using a 'passing' criterion of 5 or 6 correct out of 6 trials, three children in Experiment 3a and three children in Experiment 3b passed both the first and second sets of training trials. Thus, as in Experiment 1, most children required three sets of training trials.

As in Experiment 2, children performed above chance on the 10 vs. 3 (trained vs. known) trials on the last two training runs: Experiment 3a, 92% correct, $b = 2.5$, $SE = 0.6$, $z = 4.1$, $p < 0.001$; Experiment 3b, 93% correct, $b = 10.7$, $SE = 3.9$, $z = 2.6$, $p < 0.001$. As in Experiment 2, children were above chance on the critical training trials (trained vs. unknown, 10 vs. 5, 7, 15, 20, 30) in both experiments: Experiment 3a, 66% correct, $b = 0.7$, $SE = 0.2$, $z = 3.7$, $p < 0.001$; Experiment 3b, 70% correct, $b = 0.9$, $SE = 0.2$, $z = 3.9$, $p < 0.001$. Again, young CP-knowers clearly learned to pick the correct set of ten objects from the pairs of sets in the demonstration and practice sets. We must turn to the test trials, which involved new animals in novel configurations, to establish whether they had created a mapping between the numeral "ten" and sets of 10.

4.3.2. Test trials

Children correctly indicated a set of ten objects in the 10 vs. 3 test trials (trained vs. known): Experiment 3a, $M = 89\%$ correct, $b = 12.7$, $SE = 6.3$, $z = 2.0$, $p = 0.02$; Experiment 3b, $M = 95\%$ correct, $b = 13.9$, $SE = 6.0$, $z = 2.3$, $p = 0.01$. The data from the critical test trials are displayed on Fig. 4. In Experiment 3a, as in Experiment 2, children were at chance at the critical test trials (10 vs. 5/7/15/20/30): overall performance, $M = 52\%$ correct, $b = 0.1$, $SE = 0.2$, $z = 0.3$, n.s. In Experiment 3b, performance did not exceed chance on a 2-tailed test, but was significantly above chance on a 1-tailed test, which would be justified by the directional hypothesis. As will be seen below, this borderline success is due entirely to performance on the 10 vs. 5 trials: overall $M = 59\%$ correct, $b = 0.4$, $SE = 0.2$, $z = 1.8$, $p = 0.03$, 1-tailed. And as in Experiment 2, performance on test trials did not scale with ratio. Performance on 10 vs. 30 trials (easy, 1:3; 32% correct in Experiment 3a; 62% in 3b) and 10 vs. 7 and 10 vs. 15 trials (hard, 1:0.7; 42% correct in 3a; 52% in 3b) was at or below chance in both experiments, whereas performance on 10 vs. 5 and 10 vs. 20 (medium 1:2; 71% in 3a; 64% in 3b) was above chance in both experiments, but this was due entirely to performance on the 10 vs. 5 trials (see Fig. 4).

In sum, as is apparent in the analyses reported above, and by inspection of Fig. 4, the pattern of results in each of Experiments 3a and 3b was identical to those of Experiment 2. In spite of deconfounding number and spatial extent during the training trials in both experiments, in spite of engaging counting in Experiment 3b, and in spite of emphasizing the numerical contrasts between the pairs of training trials in both Experiments 2 and 3, young CP-knowers showed no evidence of having mapped "ten" to ANS representations of 10: when asked which set had "ten" animals, they failed on the tests trials to distinguish novel sets of 10 from sets of 20, or even sets of 30, as well as from sets of 7 and 15.

Experiment 3 replicated Experiment 2 with respect to the status of sets of 5. These young CP-knowers robustly demonstrated that they knew the label "ten" does not apply to sets of 5. However, their clear failure to distinguish sets of 10 from sets of 7, 15, 20 and 30 in the test trials suggests that they had not learned a mapping between "ten" and ANS representations of 10.

5. Summary of results

In Huang et al.'s experiment, as well as in our Experiment 1, "three"-knowers easily learned to apply "four" to sets of four in the training trials, as shown by their choices in the novel test trials, in which they applied "four" to sets of four when contrasted with sets of 5, 6, 10, or 15. In contrast, across three Teach "Ten" experiments, children successfully learned labels for sets of "ten" during the training trials, but failed to generalize this learning to test trials involving novel trained-unknown cardinal values. In all three experiments, they failed to apply the label "ten" to novel sets of 10 when these were contrasted with sets of 7, 15, 20, or 30. Furthermore, performance on each of these four comparison ratios in each experiment was at chance, as was overall performance at each comparison when the experiments were combined to increase statistical power. Over all 60 participants, performance on the 10 vs. 7 test trials was 40%, the 10 vs. 15 test trials was 52 %, the 10 vs. 20 test trials was 52% and the 10 vs. 30 was 50%. The one exception was the overall success, over all 60 participants, on the 10 vs. 5 test trials (85%, significantly better than chance, corrected for 5 comparisons[2]).

To further address the hypothesis that the success on the 10 vs. 5 comparison reflected *antecedent* knowledge concerning the cardinal meaning of the word "five," we examined performance on all 60 children in Experiments 2 and 3 on the *very first* training trial when children were asked which of sets of 5 and 10 had "ten." Whereas children overall failed on the first set of training trials, necessitating three repetitions for most children, performance was 70% on the first 5 vs. 10 training trial ($b = 0.85$, $SE = 0.29$, $z = 2.97$, $p = 0.0065$). Merely being told once, in the demonstration run, that "this card has ten, this one has five, and this one has ten" was sufficient for success, suggesting that children already knew that a set of five should be called "five." That is, the 10 vs. 5 comparison, like the 10 vs. 3 comparison, is a trained-known comparison. In contrast, taking all 60 children together, they succeeded at *none* of the other specific critical training contrasts (not 10 vs. 7, 10 vs. 15, 10 vs. 20 nor 10 vs. 30[3]) after seeing just one demonstration of each before the first training run, although they were above chance on *each* of these comparisons over the last two training runs[4]. Their success on these latter trials reflected learning for each picture pair, which one had "ten," as shown by their failure to discriminate novel test pairs on the basis of the trained numerical contrasts.

6. General discussion

Replicating Huang et al. (2010), "three"-knowers could be quickly taught to pick out sets of "four". Exact cardinal meanings for "one" through "four" are all achieved by the end of the subset-knower phase and fully present in CP-knowers (Le Corre & Carey, 2007). Indeed, that "three"-knowers successfully learned something about what "four" meant from the minimal training in Experiment 1, and from only 1 demonstration in Huang et al. (2010), suggests they might have had fragile knowledge of "four" already (see Wagner, Chu, & Barner, submitted for publication). Nonetheless, the training was necessary; children in Experiment 1 did not succeed on the first set of critical practice trials with "four."

The important finding from the present studies is the failure, as a group, of 60 young *CP- knowers* to learn to map "ten" to sets of 10 under identical circumstances in which even younger *"three"-knowers* learn to map "four" to sets of 4. In all three experiments, children succeeded by at least the second and third sets of training trials to associate "ten" with particular arrays, but failed abjectly to generalize this knowledge on the basis of the numerical contrasts between sets of 10 and other set sizes to new exemplars (new kinds, new spatial arrangements).

These results bear on what non-verbal representations underlie the meanings of "one" to "four" for subset-knowers, for they suggest that children – even young CP-knowers – do not have an over-hypothesis that number words pick out set sizes as specified by ANS representations. If this hypothesis were available to them, they should have mapped "ten" to representations in the ANS of sets of 10. Contrasts of 10 vs. 20 and 10 vs. 30 are vastly greater than the resolution of the ANS by the age of the children of Experiments 2 and 3 (Halberda & Feigenson, 2008); indeed the sets of 10 are easily discriminable from the contrasting set in *each* training and test pair. Thus, if the meaning of "four" that allows children to decide that it refers to a set of 4 rather than a set of 7 or 10 is exhausted by ANS representations of the numerosities of sets, and representations of the same sort underlie the meanings of "one" through "three," children who were even older would have been expected to easily learn that "ten" refers to sets of 10 as contrasted with sets of 30. In spite of repeated efforts, Experiments 2 and 3 failed to find any evidence in support of this hypothesis.

These studies provide a new source of data suggesting that the analog symbols that are the output of the ANS system of representation are not likely to underlie the meanings of verbal numerals in the subset-knower stage. Nevertheless, these data do not *force* that conclusion. As Dehaene and Mehler (1992) pointed out, the frequency of usage of number words decreases vastly over "one" through "ten," a phenomenon readily observable in child directed speech. Le Corre et al., 2016, analyzed CHILDES transcripts of parental input to Mandarin and English learners, finding that verbal numerals are frequent in speech to toddlers. In particular, "four" is more frequent, both overall and in cardinal usages in particular, in these corpora than "ten" (cardinal usages in English input: "four" is 0.36/1000 utterances; "ten" is 0.16/1000 utterances; in Mandarin input: "four" is 1.5/1000 utterances; "ten" is 0.47/1000

[2] Performance on 10 vs. 5 on test was significantly above chance ($b = 1.73$, $SE = 0.36$, $z = 4.80$, $p<0.001$), corrected for 5 comparisons. Performance on 10 vs. 7 ($b = -0.41$, $SE = 0.26$, $z = -1.54$, $p = 0.24$), 10 vs. 15 ($b = 0.07$, $SE = 0.26$, $z = 0.26$, ns.), 10 vs. 20 ($b = 0.07$, $SE = 0.26$, $z = 0.26$, ns.), and 10 vs. 30 ($b = 0.0$, $SE = 0.26$, $z = 0.0$, ns.) was not significantly different from chance, correcting for 5 comparisons.

[3] Performance on 10 vs. 5 on the first practice trial was significantly above chance ($b = 0.85$, $SE = 0.29$, $z = 2.97$, $p = 0.0065$). Performance on 10 vs. 7 ($b = 0.07$, $SE = 0.26$, $z = 0.26$, $p = 0.45$), 10 vs. 15 ($b = 0.20$, $SE = 0.26$, $z = 0.77$, $p = 0.45$), 10 vs. 20 ($b = 0.27$, $SE = 0.26$, $z = 1.03$, $p = 0.45$), and 10 vs. 30 ($b = 0.55$, $SE = 0.27$, $z = 2.04$, $p = 0.08$) was not significantly different from chance, correcting for 5 multiple comparisons.

[4] Performance on all trials was significantly above chance on the last two practice trials, correcting for 5 multiple comparisons: 10 vs. 5: $b = 1.44$, $SE = 0.23$, $z = 6.21$, $p < 0.001$; 10 vs. 7: $b = 0.43$, $SE = 0.21$, $z = 2.06$, $p = 0.032$; 10 vs. 15: $b = 0.46$, $SE = 0.20$, $z = 2.25$, $p = 0.032$; 10 vs. 20: $b = 0.71$, $SE = 0.29$, $z = 2.45$, $p = 0.021$; 10 vs. 30: $b = 1.39$, $SE = 0.45$, $z = 3.12$, $p = 0.007$.

utterances. Thus, taking both languages together, "four" is about 2 ½ to 3 times more frequent in the child's input than "ten." Experiments 2 and 3 show that young CP knowers do not work out a mapping between "ten" and the ANS representation of the cardinal value of sets of 10 from the 28 pairings they have of "ten" with sets of 10 (as contrasted with sets of 3, 5, 7, 15, 20, and 30). Perhaps young children, even CP-knowers, simply need more pairings of a verbal numeral with a set size, and more contrasts with other set sizes, in order to map the number word to an ANS value, such that the success in Experiment 1 reflects a past history of many more pairings of "four" with sets of 4 than older children have experienced pairing "ten" with sets of 10. If this hypothesis is right, it should be *possible* to teach 4-year-olds that "ten" means approximately 10 (as specified by the ANS) with more input.

The present experiment cannot rule out this hypothesis, but clearly the child does not have an overhypothesis that supports fast mapping of "ten" to approximately 10. Indeed, Sullivan and Barner (2014) showed that even *6-year-olds* have created mappings between verbal numerals and ANS values only up to "five" or "six." Six-year-olds have heard "ten" in cardinal contexts vastly more than 3-year-olds have heard "four" in cardinal contexts.

The present data, along with Sullivan and Barner's findings, raise the question of why, for young children, ANS values are not readily amenable to labeling with words. Since rats and other animals can formulate non-verbal rules over ANS values, it would be interesting to know whether it is *possible* to teach such rules to subset knowers or young CP-knowers (e.g., press the button N times when cued to do so; hit the button when you see a display of N dots, where N is demonstrated non-verbally). Since rats can do so (Platt & Johnson, 1971), we expect 3-year-old children could. The next question would be whether they just as easily learn could learn that *words* express such values. It is possible that children could learn non-verbal rules about certain quantities, but not word-to-ANS mappings for the same quantities with similar training conditions. Such a finding would suggest strong constraints on word meanings, such that ANS values are not candidate sources of meanings for number words.

One unanticipated finding was that young CP-knowers had apparently mapped "five" to sets of 5 (or at least to some set-size that could not be 10). This finding might seem to pose a problem for the enriched parallel individuation hypothesis concerning the non-verbal representations that support number word meanings, because working memory capacity of children this age is unlikely to be 5. Future work should address whether the linear function between set sizes of 1–4 and numerals "one" through "four" that all CP knowers display extends to 5 - "five." If so, one possible explanation is that parallel individuation has reached a capacity of 5 individuals by 3.5–4 years of age (see Starkey & Cooper, 1995, for suggestive evidence this might be so). Another possible explanation is that children create the long term memory representation of enriched parallel individuation for sets of 5 by chunking representations of 4 and 1, or 3 and 2, and similarly chunk the representations of currently attended sets for purposes of comparison to their long term memory models. These hierarchical set chunking capacities are attested in infants as young as 14 months (Feigenson & Halberda, 2004; Rosenberg & Feigenson, 2013). Such hierarchical representations allow the child to expand their working memory capacity. For example, a child who knows up to "four" could use non-verbal hierarchical chunking to underlie a representation of 4 + 1: ({g h i j}{a}), which the child could potentially use as a basis for labeling the "five."

The data presented here motivate us to reject the ANS value hypothesis concerning the non-verbal representations that underlie the meanings of "one" through "four" in the subset-knower phase. This conclusion, however, does not preclude the possibility that *some mapping* between words and the ANS has been created before children become CP-knowers. Indeed, there is very strong evidence this is so, including the fact that many subset knowers' verbal estimates of the number of individuals in arrays of individuals are positively related to set size (Cheung et al., 2016; Gunderson et al., 2015; Odic et al., 2015; Wagner & Johnson, 2011). Odic et al. (2015) even demonstrated a linear relationship between the number of pats produced and the numbers requested (6, 8, or 10) by all CP-knowers (word-to-ANS mapping). However, the results from the current experiment suggest that whatever mapping there might be, if any, between the ANS and verbal numerals in the subset-knower period, or in the period of transition to becoming a CP-knower, it is unlikely to involve *associative* mappings between the numerals up to "ten" and ANS values. If this type of mapping had already been established, young CP-knowers should have been able to succeed without training in the practice trials in Experiments 2 and 3 when the comparison ratios were between the sets of 10 and the other set were large (e.g., 10 vs. 20 and 10 vs. 30), and they also should have succeeded at these ratios during the test trials after training.

Of course, even adults who create systematic mappings between numerals and ANS values into the hundreds, do not do so by creating individual mappings between "forty- nine", "two hundred twelve," "three-hundred twenty-five," etc., on the one hand, and ANS representations of sets of 49, 212, and 325, etc., on the other. Rather, they have created a structure mapping between the count list as a whole and the ANS as a whole, and they estimate the word that would apply to a given set size by computing a ratio between the set to be estimated and a set for which there is a known ANS value-word mapping, and asking what word stands in the same relation in the count list to the known word in that ANS-value pair. The primary evidence for such a structure mapping derives from calibration effects: telling people that a set of 100 is "one-hundred and fifty" for example, causes people to systematically overestimate the number that would apply to a wide range of other sets (see Izard & Dehaene, 2008). In two elegant papers, Sullivan and Barner (2012), Sullivan and Barner (2014) probed two signatures of bidirectional ANS-word mappings to ask, which, if any, adults and children have constructed. They made two predictions. First, they reasoned that if subjects have an ANS value-to-word mapping between a verbal numeral and an ANS value that supports estimation (either word-to-ANS or ANS-to-word), then the acuity of that estimation should be predicted by the acuity of non-verbal numerical discriminations in that region of the ANS. Second, they predicted that *where* such ANS value mappings have been created, subjects should be relatively impervious to the recalibration effects that reflect structure mappings between the count list and ANS values. By both of these signatures, most adults have created ANS value to word mappings only up to 12-"twelve" or 15-"fifteen" or so, and rely on structure mapping when providing verbal estimations of the cardinal values of larger sets. And five- to seven-year-old children – much older than those in the present studies – appear to have created ANS value-to-word mappings only up to 5-"five"or 6-"six." Like Sullivan and Barner's (2014) findings, the results of the present studies confirm that young children do not readily form direct associations between verbal numerals and ANS representations of cardinal values higher than around 5.

The data from the current paper converge with other arguments that *ANS value-to-word* mappings are not necessary for the transition between being a subset-knower and a CP-knower. However, there are other ways that the ANS might support the transition to CP-knower. Specifically, perhaps a *structural* mapping between the ANS and the count list as a whole is created during the subset-knower period and plays a role in the transition. The simplest structure mapping would be a "later greater" generalization—the generalization that words later in the count list refer to larger numbers as specified by the ANS. Some subset-knowers clearly have such a mapping, indicated in numerical estimation tasks (i.e., quantity-to-word tasks) by small positive slopes between set sizes and verbal estimates in the 6–10 range (e.g., Gunderson et al., 2015) and, in a word-to-quantity task, pat-the-tiger, patting more when asked to pat "ten" times than when asked to pat "six" or "eight" times (Odic et al., 2015). Clearly, if children have created this structure mapping, this might help them in the transition to becoming a CP-knower, which involves learning just this rule (later-greater), as well as the successor rule (next numeral = exactly 1 greater; see Davidson, Eng, & Barner, 2012; Sarnecka & Carey, 2008). However, knowledge of a later-greater rule applied to a subset-knowers' whole count list is unlikely to be necessary for the transition to CP-knowledge, for two reasons. First, not all subset knowers demonstrate knowledge of this rule before they become CP-knowers. Le Corre's and Carey's data (2007; confirmed by Gunderson et al., 2015) suggests that even some young CP-knowers have not created that minimal mapping as applied to their whole count list: they do not consistently provide higher verbal estimates for the cardinal values of sets of 10 than for sets of 6. Relatedly, some children are likely to have the later-greater rule long before making the transition to CP-knower (Gunderson et al., 2015); if this rule were sufficient, then one would expect a rapid transition to CP-knower. Second, the slope of numerals produced as a function of set size probed is usually very shallow, around 0.2—not even close to the slope of 1 that would reflect an accurate mapping between number words and quantities. Given the wide variation in children's understanding of the later-greater rule prior to becoming CP-knowers, it is difficult to see how this mapping can be interpreted as a *necessary* foundation for children's number word meanings.

It is important to note that acquiring a successful later-greater structure mapping logically requires that at least some verbal numerals have been associated with specific ANS values. A child could not learn the generalization that later of two numerals in the count list picks out the larger of two sets (as specified by the ANS) without having mapped at least some small numbers to ANS values. After all, ANS values are automatically computed in the presence of attended sets, and so once the child has learned the meaning of the words "one," "two", "three" and "four" such that they reliably apply to sets of 1–4 respectively, the child is in the position to create these associations, which then could be input into learning the later-greater rule. What we are arguing here is that these associations do not constitute the *meanings* of the verbal numerals in the subset-knower period, and they do not contribute to an over-hypothesis that the meaning of all number words is provided by a specific ANS value.

It is also important to stress that the ANS is foundational to arithmetic in other senses. The ANS is an innately interpreted system of number representation that supports computations of numerical comparison and numerical calculations. Once integrated with numeral representations, it supports intuitions about such calculations and sometimes, in the case of numerical comparison, is automatically engaged (Dehaene, 2011). While it is no doubt important to mathematical reasoning, we conclude that the ANS is not the foundation of the meanings of the first explicit symbols for the positive integers. Further research should continue to seek positive evidence that bears on whether enriched parallel individuation, or some other system of representation, supports children's initial number word meanings.

Acknowledgements

The research reported here was supported by a NSF Reese Program grant to Elizabeth Spelke and Susan Carey. We would like to thank Yeshim Iqbal and Heena Lakhani for their important contributions to the design and data collection in Experiments 1 and 2.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2017.06.022.

References

Almoammer, A., Sullivan, J., Donlan, C., Marušič, F., O'Donnell, T., & Barner, D. (2013). Grammatical morphology as a source of early number word meanings. *Proceedings of the National Academy of Sciences, 110*(46), 18448–18453.

Barner, D. (2017). Language procedures, and the non-perpetual origin of number word meanings. *Journal of Child Language., 44*(3), 553–590.

Barner, D., Libenson, A., Cheung, P., & Takasaki, M. (2009). Cross-linguistic relations between quantifiers and numerals in language acquisition: Evidence from Japanese. *Journal of Experimental Child Psychology, 103*(4), 421–440.

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language, 24*(3), 511–533.

Brady, T. F., & Alvarez, G. A. (2015). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 921.

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition, 118*(1), 32–44.

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

Cheung, M. P., Slusser, E., & Shusterman, A. (2016). A 6-month longitudinal study on numerical estimation in preschoolers. In: Papafragou, A., Grodner, D., Mirman, D., & Trueswell, J.C. (Eds.) (2016). Proceedings of the 38th annual conference of the cognitive science society. Austin, TX: Cognitive Science Society.

Dantzig, T. (1967). *Number: The language of science*. New York: Free Press.

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

Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences, 1156*(1), 232–259.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. Oxford, England: Oxford University Press.

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

Douglas, Bates, Maechler, Martin, Bolker, Ben, & Walker, Steve (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6*(5), 568–584.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Feigenson, L., & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition, 91*, 173–190.

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

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: The MIT Press.

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

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

Gunderson, E. A., Spaepen, E., & Levine, S. C. (2015). Approximate number word knowledge before the cardinal principle. *Journal of Experimental Child Psychology, 130*, 35–55.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the" Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457.

Huang, Y. T., Spelke, E. S., & Snedeker, J. (2010). When is four far more than three? Children's generalization of newly acquired number words. *Psychological Science, 21*, 600–606.

Ifrah, G. (1985). *From one to zero: A universal history of numbers*. New York: Viking Press.

Izard, V., & Dehaene, S. (2008). Calibrating the number line. *Cognition, 106*(3). 1221–1.

Kayne, R. (May 2016). Retrieved from: <http://linguistics.as.nyu.edu/docs/IO/2652/Kayne0516Numeral.pdf>.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the counting principles. *Cognition, 105*, 395–438.

Le Corre, M., Li, P., Huang, B. H., Jia, G., & Carey, S. (2016). Numerical morphology supports early number word learning: Evidence from a comparison of young Mandarin and English learners. *Cognitive Psychology, 88*, 162–186.

Le Corre, M., Van de Walle, G., Brannon, E., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology, 52*(3), 130–169.

Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences, 12*(6), 213–218.

Leslie, A. M., Gelman, R., & Gallistel, C. R. (2007). Where integers come from. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Foundations and future* (pp. 109–138). Oxford: Oxford University Press.

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

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical identity. *Nature, 215*, 1519–1520.

Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition, 138*, 102F121.

Piaget, J. (1952). *The origins of intelligence in children* (Vol. 8, No. 5, pp. 18–1952). New York: International Universities Press.

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences, 14*(12), 542–551.

Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: Effects of error contingencies. *Learning and Motivation, 2*(4), 386–414.

Ramscar, M., Dye, M., Popick, H. M., & O'Donnell-McCarthy, F. (2011). The enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS ONE, 6*(7), e22501.

Rosenberg, R. D., & Feigenson, L. (2013). Infants hierarchically organize memory representations. *Developmental Science, 16*(4), 610–621.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development, 74*(6), 1807–1822.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*(3), 662–674.

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. B. (2007). From grammatical number to exact numbers: Early meanings of 'one', 'two', and 'three' in English, Russian, and Japanese. *Cognitive Psychology, 55*(2), 136–168.

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103*(3), 325–337.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*(3), 418–431.

Siegler, R. S. (1991). *Children's thinking* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition, 38*(2), 179–211.

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the investigation of language and thought*. Cambridge, MA: MIT Press.

Starkey, P., & Cooper, R. G. (1995). The development of subitizing in young children. *British Journal of Developmental Psychology, 13*(4), 399–420.

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences, 110*(45), 18116–18120.

Sullivan, J., & Barner, D. (2012). How are number words mapped to approximate magnitudes? *Quarterly Journal of Experimental Psychology, 66*, 389–482.

Sullivan, J., & Barner, D. (2014). Inference and association in children's early numerical estimation. *Child Development, 85*, 1740–1755.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience, 16*(9), 1493–1504.

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition, 119*(1), 10–22.

Wagner, K., Chu, J., & Barner, D. (submitted for publication). Do numbers begin noisy? A reanalysis of Wynn's Give-a-Number task. submitted for publication, 2017.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*(2), 130–137.

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*(2), 220–251.

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences, 13*(4), 167–174.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(5), 414–420.

Zosh, J. M., & Feigenson, L. (2009). Beyond "what" and "how many": Capacity, complexity, and resolution of infants' object representations. In L. Santos & B. Hood (Eds.), *The origins of object knowledge*. New York: Oxford University Press.