Cognition 138 (2015) 102–121

Contents lists available at ScienceDirect

Cognition

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

Children's mappings between number words and the approximate number system

Darko Odic [a,*], Mathieu Le Corre [b], Justin Halberda [c]

[a] University of British Columbia, Canada
[b] Universidad Autonoma del Estado de Morelos, Mexico
[c] Johns Hopkins University, United States

ARTICLE INFO

Article history:
Received 21 January 2014
Revised 14 October 2014
Accepted 19 January 2015
Available online 24 February 2015

Keywords:
Acquisition of number
Natural numbers
Approximate number system
Cardinality principle

ABSTRACT

Humans can represent number either exactly – using their knowledge of exact numbers as supported by language, or approximately – using their approximate number system (ANS). Adults can map between these two systems – they can both translate from an approximate sense of the number of items in a brief visual display to a discrete number word estimate (i.e., ANS-to-Word), and can generate an approximation, for example by rapidly tapping, when provided with an exact verbal number (i.e., Word-to-ANS). Here we ask how these mappings are initially formed and whether one mapping direction may become functional before the other during development. In two experiments, we gave 2–5 year old children both an ANS-to-Word task, where they had to give a verbal number response to an approximate presentation (i.e., after seeing rapidly flashed dots, or watching rapid hand taps), and a Word-to-ANS task, where they had to generate an approximate response to a verbal number request (i.e., rapidly tapping after hearing a number word). Replicating previous results, children did not successfully generate numerically appropriate verbal responses in the ANS-to-Word task until after 4 years of age – well after they had acquired the Cardinality Principle of verbal counting. In contrast, children successfully generated numerically appropriate tapping sequences in the Word-to-ANS task before 4 years of age – well before many understood the Cardinality Principle. We further found that the accuracy of the mapping between the ANS and number words, as captured by error rates, continues to develop after this initial formation of the interface. These results suggest that the mapping between the ANS and verbal number representations is not functionally bidirectional in early development, and that the mapping direction from number representations to the ANS is established before the reverse.

1. Introduction

While children and preverbal infants of all cultures have an intuitive sense of approximate number (*for review, see* Dehaene, 2009), acquiring an exact verbal count list (e.g., "one, two, three...") that appropriately represents exact cardinalities is a difficult feat that takes children several years to master (Fuson, 1987; Gelman & Gallistel, 1978; Le Corre & Carey, 2007; Wynn, 1992). By adulthood, speakers of languages that have exact number words can translate nearly effortlessly between these two formats, with brief presentations of number symbols (e.g., "7") or words (e.g., "seven") being sufficient to activate a corresponding approximate number system (ANS) representation (Ansari,

* Corresponding author at: Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada.
E-mail address: darko.odic@psych.ubc.ca (D. Odic).

http://dx.doi.org/10.1016/j.cognition.2015.01.008
0010-0277/

2007; Ebersbach & Erz, 2014; Moyer & Landauer, 1967; Pinel, Piazza, Le Bihan, & Dehaene, 2004) and a briefly shown display of dots being sufficient to activate a corresponding number word (Crollen, Castronovo, & Seron, 2011; Ebersbach & Erz, 2014; Izard & Dehaene, 2008; Krueger, 1972). However, before children understand the number words, an interface between number words and the ANS is not possible. How this interface first develops has not yet been determined.

Children's learning of number words is surprisingly slow, and follows a stereotyped progression. Around age 2, children first learn to recite the counting words in order (i.e., "one–two–three…"), but they attach no specific numerical meaning to them (cp. reciting the ABCs or other memorized songs). Then, around 2:6 (years:months), children come to understand what the word "one" means (Le Corre & Carey, 2007; Wynn, 1992). Children can demonstrate this knowledge in multiple ways – e.g., when asked to count a set with only one item children will correctly report that the set contains one item, and when asked to give someone one item children will correctly give only one item (Le Corre, Van de Walle, Brannon, & Carey, 2006; Sarnecka & Carey, 2008). However, during this stage all other number words in the count list remain semantically undifferentiated – e.g., when asked to give two, three or more items a "One-Knower" will grab a random handful of items, apparently failing to differentiate between larger and smaller number words (Le Corre et al., 2006; Wynn, 1992).

Over the following 6- to 12-months, each child progresses through stages of being a "Two-Knower" (i.e., understanding "one" and "two" but not differentiating "three, four, five", etc.), a "Three-Knower" and perhaps a "Four-Knower" before emerging as a child who understands the Cardinality Principle – i.e., understanding that the last number produced when counting a set indicates the cardinality of the set (Gelman & Gallistel, 1978; Le Corre et al., 2006; Wynn, 1992). Children who only understand some subset of the number words, and not the Cardinality Principle, have been called "Subset-Knowers", while children who have fully grasped the Cardinality Principle are called "CP-Knowers" (Cardinality Principle knowers). Importantly, throughout this entire sequence of developmental changes in numerical understanding, which lasts more than a year, Subset-Knowers will often perfectly enumerate sets containing 10 or more items using one-to-one correspondence by serially pointing to each item and counting, "one, two, three…" and they will correctly use the counting words in order and not double count items (Gelman & Gallistel, 1978; Le Corre & Carey, 2007). This striking contrast between children's robust verbal counting behavior and their still emerging conceptual understanding of cardinality is a puzzle: what understanding do these children have for the number words they produce when counting collections of items?

In addressing this question, a recent tension has emerged in the literature, highlighted by two influential papers. On the one hand, Le Corre and Carey (2007) had children look at a brief display of dots and guess their number and found that Subset-Knowers and many CP-Knowers (termed "Nonmappers" by Le Corre and Carey), were unable to produce a number word that approximately matched the number of dots (e.g., when briefly shown nine dots these children produced number words that were no larger than the verbal estimates they produced when shown five dots). This study added to the evidence that young children (including many who understood the Cardinality Principle) have no approximate numerical meaning for these words – not even a sense that arrays of nine dots should map to larger number words than arrays of five dots (*see also* Condry & Spelke, 2008). In contrast, Wagner and Johnson (2011) had children gather some number of objects (e.g., toy fish) in response to a requested number (e.g., "can you put seven fish in the pond") and found that even Subset-Knowers gave approximately the correct number, though they were unable to give exactly the right number (e.g., when asked to give "nine fish" these children tended to give more fish than when asked to give "five fish"). This study provided evidence that young children do have some approximate numerical meaning for these number words, even before fully understanding the Cardinality Principle.

At the heart of this disagreement in the literature is the issue of how children interface their number words with their early approximate number representations – i.e., whether the interface between number words and the ANS is established before or after children fully understand the Cardinality Principle. Although the evidence in the existing papers appears, on its face, incompatible, there is a possibility that they are both correct. Specifically, while Le Corre and Carey (2007) asked children to convert ANS representations into number words (e.g., converting dot arrays into number word responses), Wagner and Johnson (2011) asked children to convert number words into ANS representations (e.g., converting number word requests into fish collections). The appropriate distinction, therefore, may not be that preschool-aged children either have or do not have an interface between their number words and the ANS, but instead that this interface is not functionally bidirectional for the youngest counters, and that one direction of the interface develops prior to the other.

The goal of the present work is to explore the developmental origins of the mapping between ANS representations and number words. One possibility is that the mapping between the ANS and the number words is acquired bidirectionally – that is, children may, at the very same time, come to translate their representations of approximate number into number words and vice versa (Gallistel & Gelman, 1992; Joram, Subrahmanyam, & Gelman, 1998). For example, when briefly shown a set of dots, children might be able to say how many dots there are (i.e., *ANS-to-Word*), and when given a number word (e.g., "six") they might be able to approximately reproduce it (e.g., by tapping their finger approximately 6 times; i.e., *Word-to-ANS*). Both of these directions have been observed in adults, and adults can rather effortlessly translate either from ANS-to-Word or from Word-to-ANS (e.g., for information on Word-to-ANS task, traditionally called magnitude production tasks, see Cordes, Gelman, Gallistel, & Whalen, 2001; Crollen et al., 2011; Krueger, 1972; Whalen, Gallistel, & Gelman, 1999; e.g., for information on ANS-to-Word, traditionally called magnitude estimation tasks see Crollen et al., 2011; Halberda, Sires, & Feigenson, 2006; Izard & Dehaene, 2008; Krueger, 1972).

Another possibility, and one that could reconcile the currently opposing views in the literature, is that the mapping between the ANS and number words is not immediately bidirectional; in other words, while children may be capable of performing one direction of the mapping, they may be unable to perform the other. For example, perhaps a mapping from ANS-to-Word is distinct from, and easier to acquire than, a mapping from Word-to-ANS. A potential basis for this asymmetry might reside in a distinction between the computations required to support mapping from discrete numerical labels to continuous ANS representations compared to the opposite direction. A computational distinction could be the source for the divergent results in the literature. To assess the acquisition of these mapping directions, we relied on a within-subjects design where every child played both an ANS-to-Word game and a Word-to-ANS game in order to determine when each of these mapping directions becomes functional.

While an immediately bidirectional mapping between systems is perhaps intuitively appealing, there are several reasons to anticipate that mappings in these directions, and hence the computations that support them, may be distinct.

First, Dehaene and Cohen (1997) report the case of the patient M.A.R., who, when presented with a number line labeled with 1 and 100 on its ends, was quite successful and accurate at placing a mark in the approximate location that corresponded to a verbal number word given by the experimenter. Thus, M.A.R. successfully demonstrated a Word-to-ANS mapping by translating a number word into an approximate magnitude that could then be mapped to a number line. However, when the experimenter indicated a position on this same number line, M.A.R. was greatly impaired at telling the experimenter what number word corresponded to that part of the line: "he often wrote down (and simultaneously pronounced) completely inappropriate numerals. For instance he said 90, later self-corrected as 30, for a location that actually corresponded to about 5" (Dehaene & Cohen, 1997; p. 237). That is, M.A.R. failed to demonstrate a mapping from ANS-to-Word. His performance is consistent with our suggestion that the mapping procedures between the two systems may be two independent, unidirectional mappings and that we may find that, early in development (as well as in some cases of neurological damage), performance in one direction may be superior to the other.

Second, some recent findings with children also hint at an asymmetry in performance depending on the direction of mapping. Mundy and Gilmore (2009) have demonstrated that 6-year-olds perform more poorly on an ANS-to-Word task than a Word-to-ANS task. Opfer, Thompson, and Furlong (2010) found that, for 4-year-old children who do not demonstrate a left-to-right bias in their counting behavior, counting some number of chips (1–9) can be quite accurate, while grabbing an appropriate number of chips in response to a verbal label can be quite inaccurate. And, as discussed above, Wagner and Johnson (2011) found that Subset-Knowers demonstrate an appropriate mapping from Word-to-ANS while Le Corre and Carey (2007) reported evidence that children even older than those in the Wagner and Johnson sample were unable to demonstrate a mapping from ANS-to-Word. These findings help motivate the question whether one or the other of these mapping directions will be functional earlier in development, but they fail to answer this question because of conflicting results and the diversity of methods and ages across these papers.

When examining the mapping between the ANS and number words (or symbols) in adults, researchers have usually focused on one of three dependent measures: the slope of the responses, the variability in responses, and the response error rate (the difference between the response and the target; see Castronovo & Göbel, 2012; Castronovo & Seron, 2007; Cordes et al., 2001; Crollen et al., 2011; Le Corre & Carey, 2007; Mejias, Grégoire, & Noël, 2012; Mejias & Schiltz, 2013). Linear slopes are calculated by regressing each observer's response to either the number of dots presented (in ANS-to-Word) or the verbal label provided to the observer (in Word-to-ANS), against the answer provided by the observer. A successfully *formed mapping* should show slopes that are significantly positive (i.e., greater than 0), indicating higher responses to higher numbers. Slopes not significantly different from 0 indicate that the observer's responses do not vary with the number of dots presented and that they are guessing randomly. The second variable of interest – the variability of estimates – is often indexed through the coefficient of variance (CV), calculated as the standard deviation of the responses divided by the mean for every target value. Because ANS representations show scalar variability, a mapping between number words and the ANS results in a constant CV value for any target value (i.e., higher estimates also show higher variability). Cordes and colleagues (2001) have demonstrated that CV values only stay constant across target values when participants use the mapping between the ANS and number words, and that CV is non-constant across target values when participants use non-ANS representations, such as counting the number of dots. Finally, the third measure – error rate – is useful for assessing the *accuracy of mapping*, with negative values indicating greater under-estimation and positive values indicating over-estimation (Castronovo & Göbel, 2012; Castronovo & Seron, 2007; Crollen et al., 2011; Mejias et al., 2012; Mejias & Schiltz, 2013).

In the present work, we investigate how slopes, CV, and error rates change as children acquire number words and form mappings to the ANS. We have chosen to focus on the mapping between the ANS and number words – rather than, more generally, on the mapping between the ANS and exact number representations, or between the ANS and numeric symbols (e.g., "7"). Given the evidence that the number words, symbols, and exact number representations may be somewhat independent (Dehaene, Piazza, Pinel, & Cohen, 2003; Gallistel, 2007), we felt it important to restrict our investigation to the number words themselves as these have been the focus of the majority of previous work – and specifically the conflicting results from Le Corre and Carey (2007) and Wagner and Johnson (2011). Similarly, we focus on the mapping between the number words and the ANS as opposed to any other numerically relevant representations. For example, a wealth of evidence – from multiple labs, species, and developmental stages – reveals a deep distinction between the "small number range" (generally 1–4) and the "large number range" (generally 5 and greater) in cognition and perception (Feigenson, 2005; Feigenson, Dehaene, & Spelke, 2004; Trick, Enns, & Brodeur, 1996; Trick & Pylyshyn, 1994). This research demonstrates the development of number words between "one" and "four" indicates a mapping to a *parallel individuation system* and not to the ANS (Carey, 2009; Le Corre & Carey, 2007). For this reason, we focus on performance with larger numbers (i.e., above five) when assessing the mapping between the ANS and the number words; we rely on performance in the small number range (1–4) to assess whether children understood and were engaged in the task.

To place the challenge of mapping between ANS representations and exact number representations in a broader theoretical context, we will discuss the development of these mappings as one example from a broader class of similar challenges involving mappings between continuous and discrete quantities. ANS representations fall into the class of "analog magnitude" representations, which includes other continuous quantities such as surface area, length, time and brightness (Cantlon, Platt, & Brannon, 2009; Feigenson, 2007; Odic, Libertus, Feigenson, & Halberda, 2013; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Walsh, 2003). In other words, although the ANS may represent something discrete (i.e., collections of objects), it does so continuously, by representing number as distributed Gaussian random variables that exhibit scalar variability and obey Weber's law. In contrast, number words are compositional with other natural language symbols (e.g., "John has three dogs") and are represented discretely along a linear scale (e.g., Crollen et al., 2011; Hockett, 1987; Izard & Dehaene, 2008; Landau, 2004; Senghas, Kita, & Özyürek, 2004). The formal challenges faced in translating from a continuous representation to a discrete one (and vice versa) are thereby at the heart of our investigation into the development of the mapping between the ANS and number words, and any evidence for an asymmetry in the development of the mapping between these systems can also serve to inform theorizing beyond the case study of children's representation of number (e.g., for a similar argument in the domain of spatial language, see Landau & Jackendoff, 1993).

2. Experiment 1

2.1. Methods

In order to assess mappings from ANS-to-Word and mappings from Word-to-ANS we relied on methods from previously published studies that focus on ANS representations in isolation from counting abilities. Every child participated in both the ANS-to-Word task and the Word-to-ANS task in addition to a standard verbal Counting Assessment (i.e., "What's on this card").

For the ANS-to-Word task, we relied on the "Fast Cards" game previously published by Le Corre and Carey (2007; see also Baroody & Gatzke, 1991; Davidson, Eng, & Barner, 2012; Lipton & Spelke, 2005). In this method, on each trial children are briefly shown a card that depicts some number of dots (1–10) on each trial and children are asked to estimate how many dots were on the card. If children have successfully mastered a mapping from ANS-to-Word, the numerosity of the words they produce will increase with the number of dots on the card (i.e., a positive slope), and the coefficient of variance will remain constant. The accuracy of this mapping can be further established with error rates, with lower error rates indicative of more accurate mapping. Nearly identical methods – known as magnitude estimation tasks – have been used extensively with adults throughout the literature on number (Crollen et al., 2011; Ebersbach & Erz, 2014; Halberda et al., 2006; Izard & Dehaene, 2008; Jevons, 1871; Krueger, 1972; Mejias & Schiltz, 2013).

For the Word-to-ANS task, we modified another method that has been used extensively with adults. In the "Pat the Tiger" game, children were presented with a different stuffed animal on each trial. Each animal liked to be patted a certain number of times on the head (the experimenter would puppeteer the animal asking the child e.g., "Hi! Can you pat me on the head *seven* times really really fast?"). Children were encouraged to pat very fast and to only try to get close to the right number, which, in adults, engages the mapping from Word-to-ANS in isolation of verbal counting (Cordes et al., 2001; Whalen et al., 1999). If children have successfully mastered a mapping from Word-to-ANS, the number of pats they produce will increase with the number word requested (i.e., a positive slope), and the coefficient of variance will remain constant. The accuracy of the mapping can be further assessed with error rates. Nearly identical tasks – known as magnitude production tasks – have been used extensively with adults throughout the literature on number (Cordes et al., 2001; Crollen et al., 2011; Krueger, 1972).

Note that all our analyses treat the small number range (i.e., 1–4) separately from the large number range (i.e., 6–10) because previous work in adults and children has suggested that small numbers may not be mapped to the ANS, but to a subitizing or parallel individuation system instead (Feigenson et al., 2004; Le Corre & Carey, 2007; Svenson & Sjöberg, 1983; Trick & Pylyshyn, 1994). As discussed above, we report on the data from the small number range in both our tasks as a method-check (e.g., did children understand the task) and as a replication of previous results, while children's performance in the large number range is the focus of the present mapping questions.

2.2. Participants

62 preschool-aged children (36 girls) between the ages of 2:7 and 4:6 (average age 3:6) were run. An additional 24 children were excluded due to either not completing the task (11), parental interference (2), technical issues (e.g., sound did not record) (9), or performance higher than 3.0 standard deviations from the mean of the sample (2[1]). All

[1] These two children did not seem to want to play the games, and consistently gave answers that were in the teens or twenties, even when shown one dot or asked to pat one time.

children were from the Baltimore general community and were tested, individually, in the Johns Hopkins Lab for Child Development. They received a small prize for participating.

3. Materials and procedures

Children were tested in a room at the laboratory, seated at a small table across from the experimenter. The parent, if present, was seated approximately 3 feet behind the child and was instructed not to aid the child during the task. The entire session was recorded on video. The children always did the Counting Assessment task first, and then, in counterbalanced order across children, the Fast Cards and Pat the Tiger tasks second and third.

3.1. Counting Assessment task: What's on this Card

Materials consisted of two decks of 8.5′ × 11′ laminated cards with pictures of animals on them. Each card depicted from 1 to 10 animals (e.g., a card with 1 fish, a card with 7 giraffes).

In order to assess the counting ability of each child, we administered a version of the "What's on this Card" task that has been used throughout the number literature (Gelman, 1993; Halberda, Taing, & Lidz, 2008; Le Corre & Carey, 2007; Le Corre et al., 2006). Children were first presented with a card that had a single fish on it, and were asked "What's on this card?". Once they had named the item on the card (e.g., "A fish!"), the experimenter said "That's right, it's *one* fish", putting emphasis on the number word. The experimenter then continued presenting cards from a deck depicting from 1 to 9 animals in a pseudorandom order asking each time, "What's on this card?" Children, on each trial, counted the animals and reported the number of animals on the card; children had to count on every trial and were not allowed to estimate the number of animals. Children who progressed through all 9 cards in this deck and provided a cardinal number word on each trial were categorized as CP-Knowers and the task ended. Children were allowed to make up to 2 miscounts across these trials (e.g., failing to count one of the items or double-counting one of the items), so long as they provided a cardinal number that reflected the last number in their count as their estimate. This precaution was taken to ensure that we did not unnecessarily penalize children for miscounting though, in practice, miscounts were very rare. Typically, each child completed 9–10 cards and the counting task was only administered once.

A typical response on each trial was for the child to count and then repeat the cardinal value at the end of their counting (e.g., "one, two, three, four, five. Five monkeys!"). If the child did not provide a number word estimate or count on a particular trial, the experimenter would ask "so what's the number" and if needed would ask them to count the animals. If the child did not repeat a cardinal value at end of their count, they were asked "Do you remember what number you counted?" Children were never asked a "how many" question, as this has been shown to invalidate the counting assessment (Sarnecka & Carey, 2008). If the child gave an incorrect answer (e.g., counted to five, but then gave a different cardinal number estimate; e.g., "one, two, three, four, five. Two monkeys!"), the experimenter switched to a second deck of cards that depicted from 1 to 4 animals per card. Within this deck, the experimenter began with a card depicting 1 animal and, if the child answered correctly, the experimenter would give a card with two animals, etc., continuing up in numerosity through the 1 to 4 deck, and then shifting to the 1 to 9 deck. If the child answered incorrectly on any particular trial, the experimenter would give a card with a lower numerosity. Once the experimenter had run out of cards, either completing the 5 to 9 deck or the 1 to 4 deck, a final card showing sixteen animals was shown to the child; children were allowed to freely count the animals and we recorded the highest number they counted to (up to sixteen) without making errors. The "knower-level" for the child (i.e., the highest number word that the child appeared to have an exact meaning for) was recorded as the highest number the child could name correctly at least 3 out of 4 times (i.e., One-, Two-, Three-, Four-Knower or CP-Knower). We found no evidence for knower-levels between Four-Knowers and CP-Knowers – which would have been detectable as correct performance on all 1 to 4 animal cards but incorrect performance at some higher numerosities in the 1 to 9 deck (*but see* Wagner & Johnson, 2011).

3.2. ANS-to-Word task: Fast Cards (FC)

Materials consisted of 14 laminated 8.5′ × 11′ cards with pictures of black dots on them. The 14 cards had either one, two, three, four, six, eight, or ten dots (2 cards per number); 1–4 acted as the "small range", and 6–10 as the "large range" (Le Corre & Carey, 2007). The overall area was controlled across cards such that the amount of filled area taken up by the one-dot card was the same as the total combined area for the ten-dot card.

This procedure was identical to that used by Le Corre and Carey (2007). The child was told that they would see some cards with dots on them, and that they would not be able to count them, but they would need to give their best guess. The experimenter would then show the child a card containing some dots for about one second, ensuring that the child saw them, after which the experimenter placed the card face-down in their lap and would wait for the child to respond. If the child did not respond, the experimenter would show the card again for 1 s (fewer than 5% of trials). Likewise, in cases where the experimenter judged that the child adopted a routine to answering (e.g., answered 1 for first, 2 for second, 3 for third card) or where the guess was unrealistic (e.g., higher than one hundred), the experimenter stressed the instructions again to the child and re-presented the card (about 2% of trials). There were two orders of cards presented across children. Each number was presented twice in a pseudo-random fixed order, yielding a total of fourteen trials.

3.3. Word-to-ANS task: Pat the Tiger (PTT)

Materials consisted of 20 different small plush toy animals (approx. 4′ × 5′).

The child was told that they would see some animals who like to be patted on the head. The experimenter would take out a random animal from a bin of 20 distinct plush toys and would puppeteer the animal saying: "Hi! Can you pat me on the head X times really really fast?", where X was one of the target numbers. The child was asked to pat either one, two, three, four, six, eight, or ten times in a pseudo-random fixed order; the 1–4 numbers acted as the "small range" and the 6–10 as the "large range". Each number was presented twice, yielding fourteen trials. If the child did not pat the animal, the experimenter would repeat the question. To prevent counting, children were asked to pat the animal really fast without counting. If the experimenter judged that the child either counted the pats or was tapping without paying attention, they would repeat the trial, stressing the "really really fast" when necessary (fewer than 8% of trials). Subsequent analyses, presented in the results, verified that children did not count. Children's responses were videotaped and coded for the number of pats. There were two orders of pat numbers presented across children. These orders perfectly matched the orders used in the ANS-to-Word task, and were crossed with the order the child received in the ANS-to-Word task. In this way, the ANS-to-Word task and the Word-to-ANS task were perfectly matched in trial order, the numbers probed and the number of trials per number probed.

3.4. Vocabulary checklist

Parents were asked to complete the MacArthur-Bates CDI during their visit in order to assess vocabulary size (Fenson et al., 2007). No differences were found in the present study as a function of vocabulary size and so we do not focus on these data in the present paper. For example, there was no correlation between vocabulary size and exact number knowledge once age was controlled for ($r(59) = .19$; $p = .14$).

4. Results

To assess the *formation of the mapping*, we report linear slopes of the child's response (i.e., the number they said in the Fast Cards task, and the number of times they patted in the Pat the Tiger task) versus the target number. Slopes that are significantly higher than 0 are evidence for the existence of a mapping (i.e., that children produce higher numbers for higher target values). Slopes that are not significantly different from 0 are suggestive that the child is guessing and has no established mapping between the ANS and number words. In order to assess the formation of the mapping between the ANS and number words, we focus on the large number slopes (6–10); small number slopes (1–4) are used as a method check to verify that children understood the task. Thus, our analyses focus on planned t-tests comparing subjects' 6–10 slopes to the chance level (i.e., zero), along with ANOVAs to investigate interactions (*see also* Le Corre & Carey, 2007).

To assess the *variability of the mapping*, we report CV values (e.g., the standard deviation for the responses divided by the mean). To calculate CV values with a low number of trials, we used the PsiMLE method (www.panamath.org/psimle; Odic, Im, Eisinger, Ly, & Halberda, under review). One word of caution: for children with slopes of 0 (i.e., children who have no mapping and are guessing) it is not entirely clear what a CV value indicates and these values should be interpreted with caution.

To assess the *accuracy of the mapping*, we report signed error rates (i.e., the response minus the target value). Error rates closer to 0 are indicative of a more accurate mapping. Decades of work has shown that adults and older children performing ANS-to-Word tasks produce negative error rates (i.e., under-estimation) and Word-to-ANS tasks produce positive error rates (i.e., over-estimation; Crollen et al., 2011; Crollen & Seron, 2012; Ebersbach & Erz, 2014; Krueger, 1972; Whalen et al., 1999). Hence, although scores closer to 0 indicate a more objectively accurate mapping, slightly negative scores in the ANS-to-Word task and slightly positive scores in the Word-to-ANS task are more consistent with adult-like performance. Also note that, like with CV, when a linear slope is near zero (i.e., no mapping), an error rate may not reflect mapping performance and must be interpreted with caution (e.g., normally distributed random guesses can result in an observed error rate of 0).

Task order had no effect on performance, as estimated by a 2 (Task) × 2 (Order) Repeated Measures ANOVA on the average numerical guess of each child during the ANS-to-Word (FC) and Word-to-ANS (PTT) tasks. There was no main effect or interaction with Order, and all future analyses excluded Order. As discussed in more detail below, there was a main effect of Task, with the average pats for Word-to-ANS ($M = 5.37$; $SE = 0.21$) being higher than the average guess for ANS-to-Word ($M = 4.04$; $SE = 0.12$; $F(1,62) = 56.61$; $p < .01$).

As in Le Corre and Carey (2007), we first investigated whether children's performance differed by Knower-Level. We used the What's On This Card performance to classify children into groups of One-Knowers ($N = 11$), Two-Knowers ($N = 7$), Three-Knowers ($N = 8$), and CP-Knowers ($N = 36$). To increase sample size, and thereby statistical power, we next determined whether children from the various Knower-Levels (i.e., One-Knower, Two-Knower, etc.) could be collapsed into larger groups based on similar performance; Le Corre and Carey (2007) used a similar approach. These group analyses revealed that all groups differed, with the exception of the Two- and Three-Knowers. That is, a 2 (Knower-Level: Two-Knower, Three-Knower) × 7 (Number Probed: 1, 2, 3, 4, 6, 8, 10) Mixed Measures ANOVA on children's numerical responses in both the Fast Cards and Pat the Tiger tasks did not reveal any effects of Knower-Level nor interactions involving Knower-Level for Two- and Three-Knowers. All other groups were significantly different. For this reason we collapsed Two- and Three-Knowers into a single larger group for the remaining analyses.

4.1. Formation of the ANS-to-Word and Word-to-ANS Mapping (Slopes)

Data from both the ANS-to-Word task and the Word-to-ANS task are shown in Figs. 1 and 2. A 3 (Knower-Level:

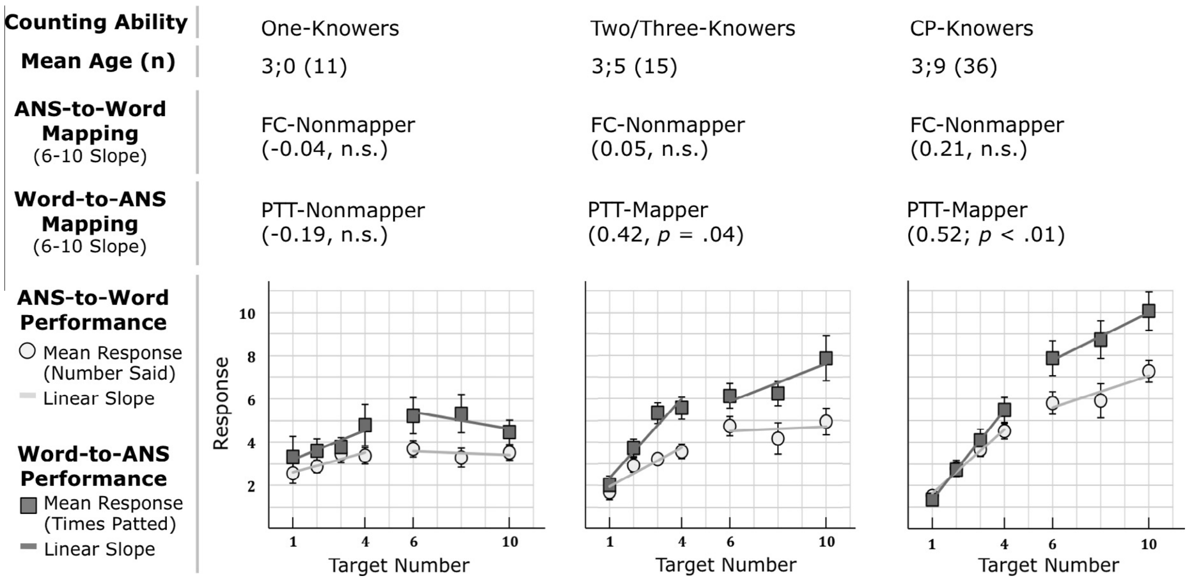

Fig. 1. Performance on the ANS-to-Word (Fast Cards) and Word-to-ANS (Pat the Tiger) task by counting ability group. Each point is the average guess for that group, and the lines are the best-fitted linear slopes. Error bars are SEM.

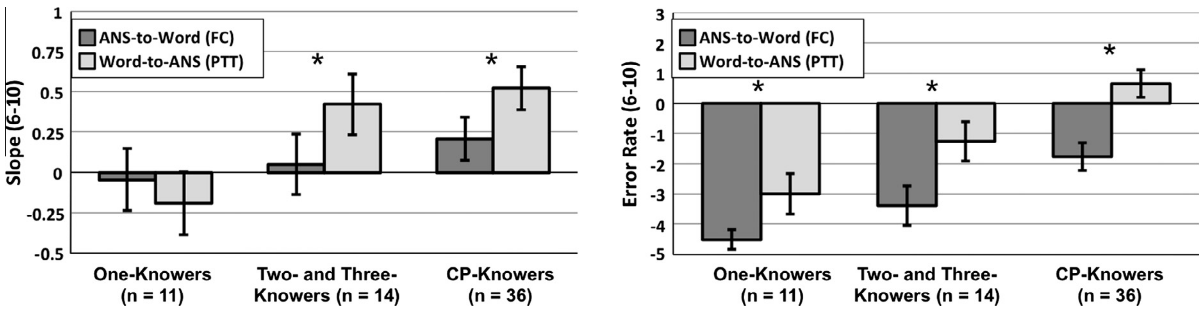

Fig. 2. The best-fit slopes and error rates for each Knower Level for the 6–10 range on each task. Error bars are SEM, and stars indicate significant difference between the two tasks at $p < .05$.

One-Knowers, Two/Three-Knowers, CP-Knowers) × Task (Fast Cards, Pat the Tiger) ANOVA over the 6–10 slopes revealed a main effect of Task ($F(1,59) = 5.89$; $p = .02$), a marginal effect of Knower-Level ($F(2,59) = 2.95$; $p = .06$) and no significant interaction ($F(2,59) = 1.67$; $p = .20$). Hence, the children in our sample showed significantly higher slopes in the Pat the Tiger task, suggesting an earlier development of the Word-to-ANS mapping. To further evaluate this possibility, we carried out a series of planned contrasts to investigate, for each Knower Level, whether their 1–4 and 6–10 slopes were significantly different from zero in each task.

In the ANS-to-Word (FC) task, One-Knowers failed to show significantly positive slopes in both the 1–4 range ($M = 0.31$; SE = 0.23; $t(10) = -0.81$; $p = .22$; see Fig. 1) and the 6–10 range ($M = -0.05$; SE = 0.06; $t(10) > -1$; $p = .48$), suggesting no ANS-to-Word mapping. In the Word-to-ANS (PTT) task, One-Knowers had significantly positive slopes in the 1–4 range ($M = 0.45$; SE = 0.15; $t(10) = 2.62$; $p < .05$) but not in the 6–10 range ($M = -0.19$; SE = 0.19; $t(10) > -1$; $p = .34$), suggesting no Word-to-ANS mapping in the 6–10 range, either. Together, these results suggest that by the time children are One-Knowers, they have not attained any mapping between number words and the ANS (Fig. 2).

In the ANS-to-Word (FC) task, the combined sample of Two- and Three-Knowers showed positive slopes in the 1–4 range ($M = 0.59$; SE = 0.12; $t(14) = 4.84$; $p < .001$; Fig. 1), but not in the 6–10 range ($M = 0.05$; SE = 0.14; $t(14) < 1$; $p = .73$; Figs. 1 and 2), suggesting no ANS-to-Word mapping. Importantly, in the Word-to-ANS (PTT) task Two- and Three-Knowers showed positive slopes in both the 1–4 range ($M = 1.20$; SE = 0.20; $t(14) = 5.76$; $p < .001$; Fig. 1) *and* the 6–10 range ($M = 0.42$; SE = 0.19; $t(14) = 1.87$; $p < .05$; Figs. 1 and 2). The difference between the two tasks was significant ($t(14) = 2.67$; $p < .05$; Fig. 2). This is consistent with Two- and Three-Knowers having a successful mapping from Word-to-ANS, but not from ANS-to-Word (Fig. 2).

Finally, in the ANS-to-Word (FC) task the CP-Knowers showed significantly positive slopes in the 1–4 range ($M = 1.02$; SE = 0.10; $t(35) = 10.55$; $p < .001$; Fig. 1) but *not* in the 6–10 range ($M = 0.21$; SE = 0.13; $t(35) = 1.66$; $p = .11$; Figs. 1 and 2). This suggests that CP-Knowers, as a group, much like the Two- and Three-Knowers, have not formed the ANS-to-Word mapping. In contrast, in the Word-to-ANS (PTT) task CP-Knowers showed significantly positive slopes in the 1–4 range ($M = 1.36$; SE = 0.15; $t(36) = 8.71$; $p < .001$; Fig. 1) and in the 6–10 range ($M = 0.52$; SE = 0.13; $t(36) = 3.91$; $p < 0.001$; Figs. 1 and 2). The difference between the two task was significant ($t(14) = 4.55$; $p < .01$; Fig. 2). As in the case of Two- and Three-Knowers, this suggests that children formed the Word-to-ANS mapping prior to the ANS-to-Word mapping (Fig. 2).

4.2. Variability of the ANS-to-Word and Word-to-ANS mapping (CV)

To assess the variability of each mapping, we used the PsiMLE method to calculate coefficients of variance. This freely available method (www.panamath.org/psimle) uses maximum-likelihood estimation to calculate a global CV by pooling evidence across all available trials and target numbers, maximizing power and reliability (Odic et al., under review).

In the FC task, the average CV for One-Knowers was 0.25 (SE = 0.07), for Two/Three-Knowers was 0.19 (SE = 0.04) and for CP-Knowers was 0.26 (SE = 0.02). In the PTT task, One-Knowers had an average CV of 0.25 (SE = 0.06), Two/Three-Knowers of 0.21 (SE = 0.03) and CP-Knowers of 0.22 (SE = 0.02). A 3 (Knower-Level: One-Knowers, Two/Three-Knowers, CP-Knowers) × Task (Fast Cards, Pat the Tiger) ANOVA over the 6–10 CV values revealed no main effect of Knower-Level ($F(2,59) = 1.15$; $p = .32$), nor Task ($F(1,59) < 1$; $p = .51$), nor an interaction ($F(2,59) < 1$; $p = .72$). Hence, there does not appear to be observable development of CV values as children are becoming more experienced counters within the range available in our sample.

4.3. Accuracy of the ANS-to-Word and Word-to-ANS Mapping (Error Rates)

The average error rates for both the FC and PTT tasks are shown in Fig. 2. A 3 (Knower-Level: One-Knowers, Two/Three-Knowers, CP-Knowers) × Task (Fast Cards, Pat the Tiger) ANOVA over the 6–10 error rates revealed a main effect of Task ($F(1,59) = 34.150$; $p < .001$), a main effect of Knower-Level ($F(2,59) = 16.20$; $p < .001$) and no significant interaction ($F(2,59) < 1$; $p = .67$). As in the case of slope, we further investigated the effect of Task through a series of planned contrasts.

An optimal error rate is zero and, in the literature, adults show positive error for Word-to-ANS and negative error for ANS-to-Word. In our sample, One-Knowers showed significantly more negative error rates in the ANS-to-Word (FC) task ($M = -4.51$; SE = 0.33) than in the Word-to-ANS (PTT) task ($M = -3.0$; SE = 0.67; $t(11) = 2.29$; $p < .05$; Fig. 2). Similarly, Two/Three-Knowers showed significantly more negative error rates in the ANS-to-Word (FC) task ($M = -3.39$; SE = 0.66) than in the Word-to-ANS (PTT) task ($M = -1.25$; SE = 0.66; $t(14) = 2.92$; $p < .01$; Fig. 2). Finally, CP-Knowers also showed significantly more negative error rates in the ANS-to-Word (FC) task ($M = -1.75$; SE = 0.46) than in the Word-to-ANS (PTT) task ($M = 0.66$; SE = 0.46; $t(34) = 4.55$; $p < .01$; Fig. 2). Interestingly, the pattern of error rates in CP-Knowers is relatively close to the pattern in adults and older children of slight under-estimation in ANS-to-Word tasks and slight over-estimation in the Word-to-ANS tasks (Crollen et al., 2011; Ebersbach & Erz, 2014). These differences are further evidence for the asymmetry we found in the formation of the interface (i.e., slopes): children are more accurate in the Word-to-ANS compared to the ANS-to-Word task. As we discuss in the General Discussion, these results also point to a developmental trajectory whereby children first establish a mapping between number words and the ANS (evidenced in slopes), and then slowly refine the mapping accuracy over time to resemble adult responses (evidenced in error rates).

4.4. Two Populations of CP-Knowers

Previous work by Le Corre and Carey (2007) revealed that the CP-Knowers group may be comprised of two distinct populations: "CP-Mappers", who have a mapping between the ANS and number words (i.e., a positive 6–10 slope on the ANS-to-Word task), and "CP-Nonmappers", who do not have a mapping between the ANS and number words (i.e., a non-positive 6–10 slope on the ANS-to-Word task). We investigated whether the same might be true of our sample of CP-Knowers.

We first examined whether – as in Le Corre and Carey (2007) – our CP-Knowers comprised two groups in the Fast Cards task (i.e., FC-Mappers and FC-Nonmappers). Consistent with the prediction of two separate groups, a Kolmogorov–Smirnov test showed that the 6–10 slopes of our sample of CP-Knowers in the ANS-to-Word task violated normality (KS(38) = 0.154; $p < .05$). In line with previous work, we set a criterion of a slope $\geqslant 0.3$ for separating FC-Mappers from FC-Nonmappers.[2] This resulted in two groups: FC-Mappers ($N = 14$) and FC-Nonmappers ($N = 24$). This split data is shown in Figs. 3 and 4.

In the ANS-to-Word task (FC), both FC-Mappers and FC-Nonmappers had significantly positive slopes for the 1–4 range ($M = 1.02$; SE = 0.07; FC-Mappers: $t(13) = 15.375$; $p < 0.001$; FC-Nonmappers: $M = 0.99$; SE = 0.14; $t(23) = 6.87$; $p < 0.001$; Fig. 3). Because our division was based on the 6–10 slopes, the FC-Mappers showed, as expected, positive 6–10 slopes ($M = 0.92$; SE = 0.08), and FC-Nonmappers did not ($M = -0.21$; SE = 0.11; Figs. 3 and 4). In direct contrast to this, both groups of children showed significantly positive slopes in the Word-to-ANS (PTT) task (FC-Nonmappers: $M = 0.38$, SE = 0.12, $t(21) = 3.12$, $p < .01$; FC-Mappers: 0.74, SE = 0.28, $t(13) = 2.60$, $p < .05$, Figs. 3 and 4). Indeed – FC-Nonmappers – who fail to show positive slopes in the ANS-to-Word (FC) task, showed significantly higher, non-zero slopes in the Word-to-ANS (PTT) task ($t(14) = -2.11$; $p < .03$). This powerfully suggests that, even among children who have

[2] This criterion was used by both Le Corre and Carey (2007) and Davidson et al. (2012). We note that our results remain qualitatively unchanged when any criterion in the range of 0.3–0.5 is used.

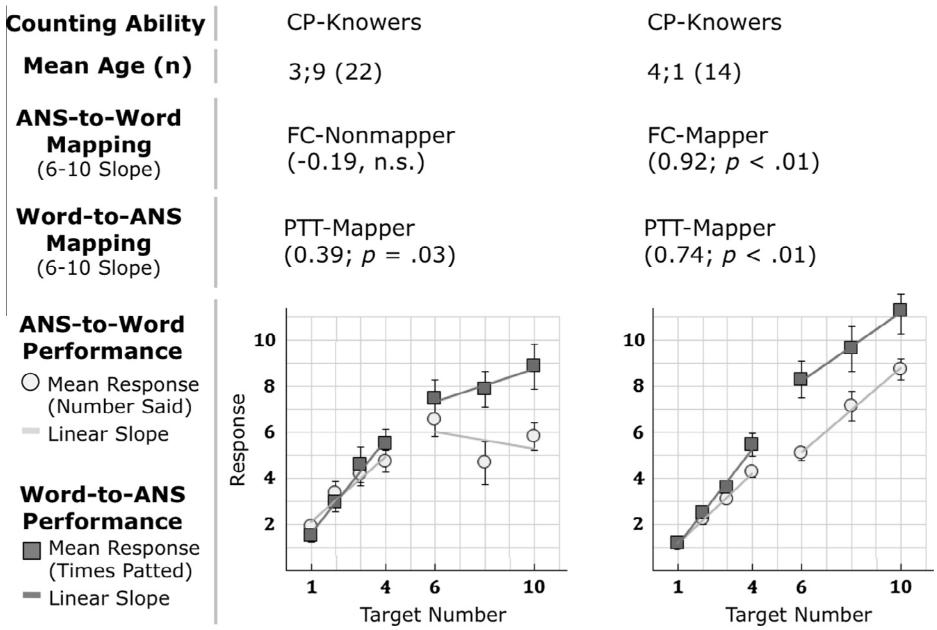

Fig. 3. Performance on the ANS-to-Word (FC) and Word-to-ANS (PTT) tasks split into the two populations of CP-Knowers: FC-Mappers and FC-Nonmappers. Each point is the average guess for that group, and the lines are the best-fitted linear slopes. Error bars are SEM.

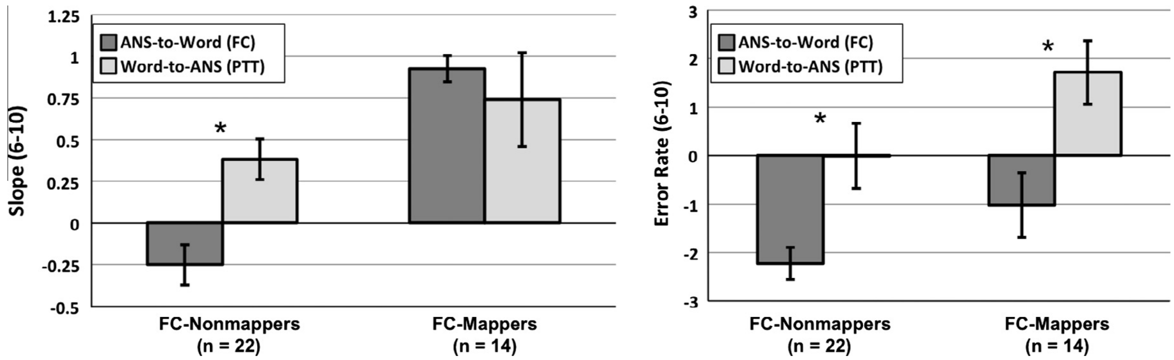

Fig. 4. The best-fit slopes and error rates for the FC-Mappers and FC-Nonmappers (the split CP-Knower group) the 6–10 range on each task. Error bars are SEM, and stars indicate significant difference between the two tasks at $p < .05$.

all learned how to count, there is an asymmetry in the mapping whereby children demonstrate a superior, earlier mapping from Word-to-ANS than from ANS-to-Word.

Next, we investigated whether the CP-Knowers could be divided into two groups based on their Word-to-ANS (PTT) 6–10 slopes. In contrast to the pattern of results in the ANS-to-Word (FC) task, the Kolmogorov–Smirnov test for the Word-to-ANS (PTT) task revealed that the 6–10 slopes of the CP-Knower sample did *not* deviate from a normal distribution (KS(36) = 0.12; $p = .20$). This suggests that the CP-Knowers comprise a single population in the ANS-to-Word (PTT) task – i.e., PTT-Mappers. Hence, while we found evidence for FC-Mappers and FC-Nonmappers, we found no evidence for PTT-Mappers and PTT-Nonmappers.

We also examined the development of error rates for FC-Mappers and FC-Nonmappers. As shown in Fig. 4, FC-Nonmappers showed significantly more negative error rates in the ANS-to-Word (FC) task ($M = -2.22$; SE = 0.52) than in the Word-to-ANS (PTT) task ($M = -0.01$; SE = 0.57; $t(13) = 3.14$; $p < .01$). Similarly, FC-Mappers showed more negative error rates in the ANS-to-Word (FC) task ($M = -1.03$; SE = 0.36) than in the Word-to-ANS (PTT) task (M = 1.71, SE = 0.69; $t(21) = 3.23$; $p < .01$). This replicates the pattern observed with the combined CP-Knower sample and suggests development toward the adult-like error pattern (i.e., under-estimation in ANS-to-Word and over-estimation in Word-to-ANS).

One error rate data point is worth discussing further – on the face of it, the FC-Nonmappers had a more accurate mapping in the Word-to-ANS task (error rate of 0.01) than the FC-Mappers (error rate of 1.71). Though at first counter-intuitive, it is important to keep in mind that adults performing Word-to-ANS tasks strongly over-estimate and show positive error scores similar to values we observe here (e.g., Crollen et al., 2011; Ebersbach & Erz, 2014; Whalen et al., 1999). Hence, our results point to a secondary developmental process that occurs *after* the interface between number words and the ANS is first formed: children continue to refine their mapping toward adult-like values, and already begin showing signatures of over-estimation by their mid-fours. Hence, while FC-Nonmappers are objectively more accurate on the Word-to-ANS task than the FC-Mappers, it is the FC-Mappers who appear further developed in their error rates when compared to typical adult performance.

4.5. Alternative strategies analyses

To ensure that the good performance in the Word-to-ANS (PTT) task derives from a functioning Word-to-ANS mapping, it is important to establish that children were not verbally counting during the task. At a first pass, children in the PTT task appeared to enjoy the challenge of trying to pat fast and typically patted at a very fast rate. Overt verbal counting was explicitly discouraged and trials on which it occurred were repeated or dropped from analysis (approximately 1% of trials). However, it would be nice to provide further evidence from children's own performance to suggest that they were not counting, even covertly "under their breath", while they patted. Below, we present three separate analyses that suggest that children in Word-to-ANS (PTT) task did not count, covertly or otherwise.

When adults count items in the large number range (i.e., more than 4 items), they do so at a rate of 250–300 ms/item (Landauer, 1962; Trick & Pylyshyn, 1994). In 6- and 7-year-old children, these large number counting rates are substantially slower, with rates around 550–1000 ms/item (Svenson & Sjöberg, 1978; Trick et al., 1996). Hence, one indicator of counting would be children's patting speeds, as rates faster than 550 ms/item would likely be too quick for children to covertly or overtly count. We coded patting rates from the session videos for each child by taking the total time they took to pat and dividing it by the number of times they patted. Time started at the moment they first touched the plush toy and stopped at the moment they took away their hand from the toy on their final pat. For each child, we averaged the patting rate for numbers 6, 8, and 10, as our interest was only in the large number range. A 4 (Knower-Level: One-Knower, Two/Three-Knower, FC-Mapper, FC-Nonmapper) ANOVA on average patting rate found no effect of Knower-Level ($F(1,56) < 1$), with One-Knowers having an average patting rate of 383 ms (SE = 28 ms), Two/Three-Knowers with an average of 396 ms (SE = 23 ms), FC-Nonmappers with 416 ms (SE = 24 ms) and FC-Nonmappers with 384 ms (SE = 18 ms). These patting rates are faster than even the most optimistic estimate for 6-year-old children's counting speeds, and suggest that children in our sample were likely not counting.

In adults, counting shows two other behavioral signatures. A seminal study by Cordes and colleagues (2001) used a Word-to-ANS task and asked adult observers to tap a button a given number times while either overtly or covertly counting or while being prevented from counting by verbal interference (repeating the word "the"). Their results showed two behavioral signatures that differentiate counting from non-counting responses during Word-to-ANS tasks. First, Cordes et al. (2001) found that tapping slowed down with higher numbers as a function of number whenever adults engaged in either overt or covert counting (*see also* Whalen et al., 1999 *and* Landauer, 1962). On the other hand, tapping speed stayed constant for trials where counting was prevented via verbal interference. Second, Cordes et al. (2001; see also Le Corre & Carey, 2007 *and* Davidson et al., 2012) found that the CV values decrease as the square root of the number presented when adults engaged in covert or overt counting, while CV stayed constant when counting was prevented.

In order to demonstrate that children in the Word-to-ANS task did not count, we checked for both of the counting signatures reported by Cordes and colleagues (2001). First, we analyzed patting speeds from the video taped sessions for PTT-Mappers via a 3 (Knower-Level: Two- and Three-Knowers, FC-Mappers, FC-Nonmappers) × 6 (Number: 2, 3, 4, 6, 8, 10) repeated measures ANOVA and revealed a main effect of Number ($F(5,230) = 3.462$; $p < .01$) toward an *increase* of patting speed with larger numbers (i.e., faster tapping for 10 than 6) and no effect of Knower-Level ($F(1,46) < 1$; $p = .85$) nor an interaction ($F(10,230) = 1.51$; $p = .14$) with Knower-Level. This is the opposite direction of that predicted for counting and is robust evidence that children did not engage in counting, covert or otherwise, when patting during the Word-to-ANS (PTT) task.

Lastly, we examined CV values. Unfortunately, the PsiMLE method maximizes reliability by collapsing across target values; this prevents us from examining the slope of CV values across different numbers. Hence, instead, we calculated, for each number bin (i.e., 6, 8 and 10) the standard deviation and divided it by the mean. This method has a high correlation with the PsiMLE values ($r = .65$) and was previously used by Le Corre and Carey (2007). A 3 (Knower-Level: Two- and Three-Knowers, FC-Mappers, FC-Nonmappers) × 3 (Number Requested: 6, 8, 10) ANOVA performed on the CV values revealed no effect of Number Requested ($F(2,94) = 1.655$; $p > 0.19$) nor an effect of Knower-Level ($F(4,94) < 1$), nor any interactions, suggesting constant CVs.

Together, children's patting speed, patting rate as a function of target number, and CVs replicate the non-counting patterns of Cordes et al. (2001), and suggest that children in our PTT task did not count.

5. Discussion

In the first experiment, we found that children formed a successful mapping from Word-to-ANS (Pat the Tiger task) significantly before demonstrating a mapping from ANS-to-Word (Fast Cards task). More specifically, while we replicated the findings of Le Corre and Carey (2007), finding that many CP-Knowers do not have a mapping between the ANS and exact number words, these same children showed an existing mapping from exact number words to the ANS, as evidenced by significantly positive slopes in the large number range in the PTT task. These results suggest that the mapping between the ANS and number words is not immediately bidirectional, and that children acquire a mapping from Word-to-ANS developmentally prior to a mapping from ANS-to-Word. These results were replicated in the accuracy of the mapping: error rates were significantly lower in the Word-to-ANS task compared to the ANS-to-Word task for both Two/Three-Knowers and CP-Knowers. Finally, we found that error rates continued developing after the interface is formed and began approaching adult-like values (i.e., under-estimation in ANS-to-Word and over-estimation in Word-to-ANS) among our oldest kids. Finally, we found no development of CV values in either task.

Although we performed analyses to verify that children in the Word-to-ANS task did not count, additional methodological concerns include differences between the Fast Cards and Pat the Tiger tasks. Besides the mapping direction, our Word-to-ANS (PTT) task in Experiment 1 also involves serial event processing (i.e., patting with the hand) while our ANS-to-Word (FC) task involves parallel object processing (i.e., viewing dots on the page). These differences might have played some role in determining why one task was easier than the other. Because our focus in Experiment 1 was on tasks that were previously established in the literature, we cannot currently adjudicate between the direction of mapping or task demands being the source of the differences we observed in children's performance.

In Experiment 2, we extended our approach by creating two tasks that differed *only* in the mapping direction required. This was accomplished by using the Pat the Tiger (PTT) task as our Word-to-ANS task – where the child was told a number and had to pat – and by creating the "Fast Pats" (FP) task – where the child watched the experimenter pat the animal and had to verbally estimate the number of pats. In doing so, we emulated every aspect of children's own patting performance (e.g., the experimenter patted the animal at the speed and rate that children showed in Experiment 1). Similar methods have appeared in the literature (Spaepen, Coppola, Spelke, Carey, & Goldin-Meadow, 2011). In this way, the only difference between the two tasks was whether the child guessed the number of pats, or was told the number of times to pat the animal.

6. Experiment 2

6.1. Participants

For efficiency, we restricted our sample to only CP-Knowers – because Experiment 1 demonstrated that the asymmetry between ANS-to-Word and Word-to-ANS is observed in the within the sample of CP-Knowers. Twenty-three children (18 girls) between the ages of 3:5 and 4:6 (average age 4:1) were run. An additional 2 were excluded for not being CP-Knowers (one was a Two-Knower and one a Three-Knower). All children were from the Baltimore general community and were tested, individually, in the Johns Hopkins Lab for Child Development. They received a small prize for participating.

6.2. Methods and procedure

Our Counting Assessment (What's on this Card) and Word-to-ANS (Pat the Tiger) tasks were identical to that of Experiment 1, except that the numbers probed in the Pat the Tiger task were changed to two, three, four, six, eight, ten, and twelve. This change was made possible by testing older children who knew higher number words, and allowed us additional trials to estimate the linear slopes in the large number range. As in Experiment 1, all children first received the Counting Assessment, followed by the Word-to-ANS and ANS-to-Word tasks in counterbalanced order.

6.3. ANS-to-Word task: Fast Pats (FP)

This task was nearly identical in structure to the Word-to-ANS task (PTT), but with the reversed direction of mapping. Children were shown the same plush toy animals used in the PTT task and the experimenter explained that each animal would whisper to the experimenter how many times it wanted to be patted. On each trial, the experimenter took a different plush animal, held it to their ear to "hear" the whisper, and then patted the animal on the head the target number of times while the child watched. The child was asked to estimate the number of times the experimenter patted the animal on the head. Children were given neutral positive feedback after every trial.

Trial orders, animals, and target numbers of pats were identical in the Word-to-ANS (PTT) and ANS-to-Word (FP) tasks: e.g., either two, three, four, six, eight, ten, or twelve pats, with each number tested twice. Because Experiment 1 revealed that children's patting speed in the PTT task was approximately 3 pats per second, experimenters were trained to pat the animals in the FP task at a rate of about 3 pats per second. This is a rate that is too fast for children to precisely count (Svenson & Sjöberg, 1978; Trick et al., 1996), but easily within the range for supporting ANS estimations of serial events (Droit-Volet, Clément, & Fayol, 2008). Importantly, we did not provide arrhythmic or time anti-correlated patting trials, because the goal of the Fast Pats task was for the experimenter to pat in a way that was nearly identical to children's own style of patting in Experiment 1. Additionally, the availability of time-based cues should, if used, make the task easier for children, working directly against our hypothesis. Subsequent video coding checked that the experimenters patted each animal the target number of times and at the proper rate, and any trials in which they did not were removed from analyses (less than 0.5% of trials).

7. Results

As in Experiment 1, we classified children into Knower-Level categories based on the WOC task. Because we focused on CP-Knowers, our sample was aimed at older children. Out of 25 children, 23 were classified as CP-Knowers; the other two children were excluded from further analyses. As in Experiment 1, the relevant measures were slope, CV, and error rate for the 6–12 range.

To determine whether task order affected results, we took the average numerical guess of each child during the ANS-to-Word (Fast Pats) task and the average number of pats produced in the Word-to-ANS (Pat the Tiger) task and computed a 2 (Task) × 2 (Order) Repeated Measures ANOVA. As in the first experiment, there was a main effect of Task, with the average pats for Word-to-ANS ($M = 9.11$; SE = 1.02) being higher than the average guess for ANS-to-Word ($M = 5.60$; SE = 0.50; $F(1,21) = 9.89$; $p < .001$). This is consistent with the adult state where ANS-to-Word tasks give rise to under-estimation while Word-to-ANS tasks give rise to under-estimation (Cordes et al., 2001; Crollen et al., 2011; Ebersbach & Erz, 2014). There was no main effect or interaction with Order. Therefore, all future analyses excluded Order.

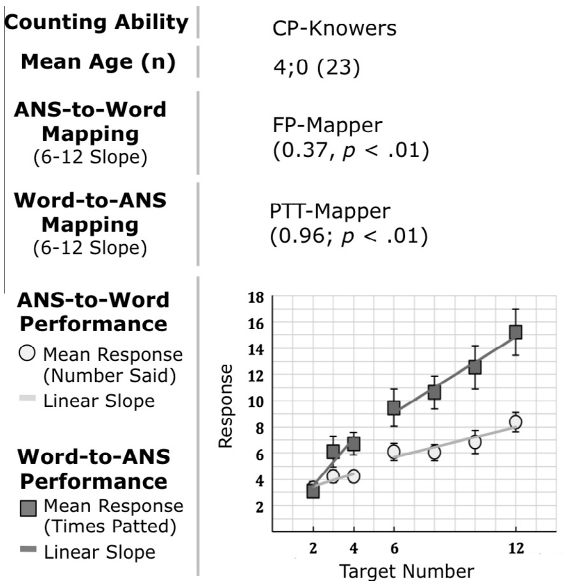

Fig. 5. Performance on the ANS-to-Word (Fast Pats) and Word-to-ANS (Pat the Tiger) for the CP-Knowers group. Each point is the average guess for that group, and the lines are the best-fitted linear slopes. Error bars are SEM.

7.1. Formation of the ANS-to-Word and Word-to-ANS Mapping (Slopes)

We determined the slope of the best-fit line for the target value versus the number given by the child for both the small number range (2–4) and the large number range (6–12). The average slopes for both the ANS-to-Word and Word-to-ANS tasks are shown in Figs. 5 and 6.

We found that CP-Knowers showed significantly positive 2–4 slopes in both the ANS-to-Word task ($M = 0.52$; SE = 0.18; $t(23) = 2.88$; $p < .05$) and in the Word-to-ANS task ($M = 1.77$; SE = 0.29; $t(23) = 3.98$; $p < .001$), suggesting that the children understood both the tasks.

If the Word-to-ANS mapping forms prior to the ANS-to-Word mapping, we should find significantly higher slopes in the Word-to-ANS (PTT) task compared to the ANS-to-Word (FP) task. Consistent with this prediction, a 2-way repeated-measures ANOVA (Task: PTT, FP) over the 6–12 slopes showed a main effect of Task ($F(1,22) = 10.15$; $p < .01$), with the Word-to-ANS (PTT) slopes ($M = 0.96$; SE = 0.18) higher than the ANS-to-Word (FT) slopes ($M = 0.37$; SE = 0.09; Figs. 5 and 6).

7.2. Variability of the ANS-to-Word and Word-to-ANS Mapping (CV)

The average CV value in the ANS-to-Word (FP) task ($M = 0.28$; SE = .02) was comparable to the Word-to-ANS task ($M = 0.27$; SE = 0.04). Thus, in replication of Experiment 1, there does not appear to be a significant amount of development or difference in CV values between the two mapping directions.

7.3. Accuracy of the ANS-to-Word and Word-to-ANS Mapping (Error Rates)

The average error rates are shown in Fig. 6. The average error rate for the CP-Knowers, as a single group, was strongly negative in the ANS-to-Word task ($M = -2.22$; SE = 0.56) and strongly positive in the Word-to-ANS task ($M = 3.02$; SE = 1.42), consistent with performance in Experiment 1 (and with typical adult performance). These error rates were significantly different from each other

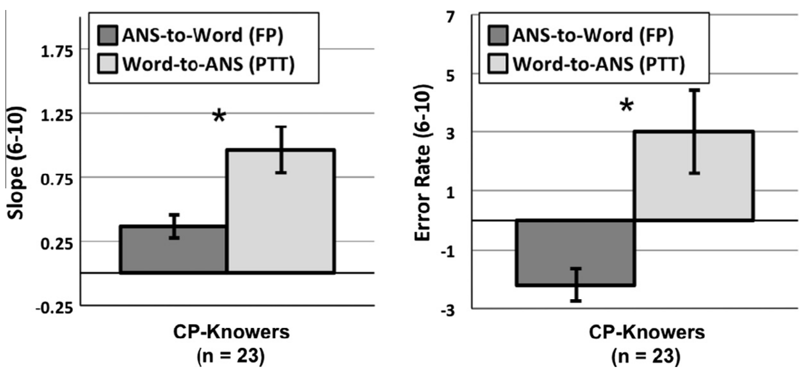

Fig. 6. The best-fit slopes and error rate for the CP-Knowers in the 6–10 range on each task. Error bars are SEM, and stars indicate significant difference from 0 at $p < .05$.

($t(23) = 3.46$; $p < .05$), thereby further replicating the results of Experiment 1. The higher rate of over- and under-estimation compared to Experiment 1 is likely due to both testing values up to 12 (where over- and under-estimation was highest) and from testing older children.

7.4. Two Populations of CP-Knowers

As in Experiment 1, we replicated previous approaches (Davidson et al., 2012; Le Corre & Carey, 2007) and divided the group of CP-Knowers into two groups based on their ANS-to-Word (i.e., Fast Pats) performance: children with a 6–12 slope lower than 0.3 were classified as FP-Nonmappers (N = 10; Average Age = 4;1), and children with 6–12 slopes higher than 0.3 were classified as FP-Mappers ($N = 13$; Average Age = 4;1). Results remained qualitatively unchanged if the slope criterion was raised to 0.5.

FP-Mappers had positive slopes in the 2–4 range on the ANS-to-Word task (Mean = 0.72, SD = 1.02, $t(12) = 2.54$, $p < .05$) and the Word-to-ANS task (Mean = 1.95, SD = 1.63, $t(12) = 4.23$, $p < .001$; Fig. 7). FP-Nonmappers had a significantly positive 2–4 slope on the Word-to-ANS task (Mean = 1.53, SD = 1.06, $t(9) = 4.07$, $p < .001$) and a marginally positive 2–4 slope in the ANS-to-Word task (Mean = 0.35, SD = 0.47, $t(9) = 2.12$, $p = .053$, Fig. 7). This marginal effect was driven by one child with a strongly negative 1–4 slope (−0.65) and removing this child did not alter any of the subsequent results, and so they were retained in the analysis to maximize power.

As expected by our division criteria (and shown in Figs. 7 and 8), the average FP-Mapper large number slope for the ANS-to-Word (FP) task was higher than 0 (Mean = 0.66; SD = 0.31), while it was not higher than 0 for the FP-Nonmappers (Mean = −0.01; SD = 0.23; Figs. 7

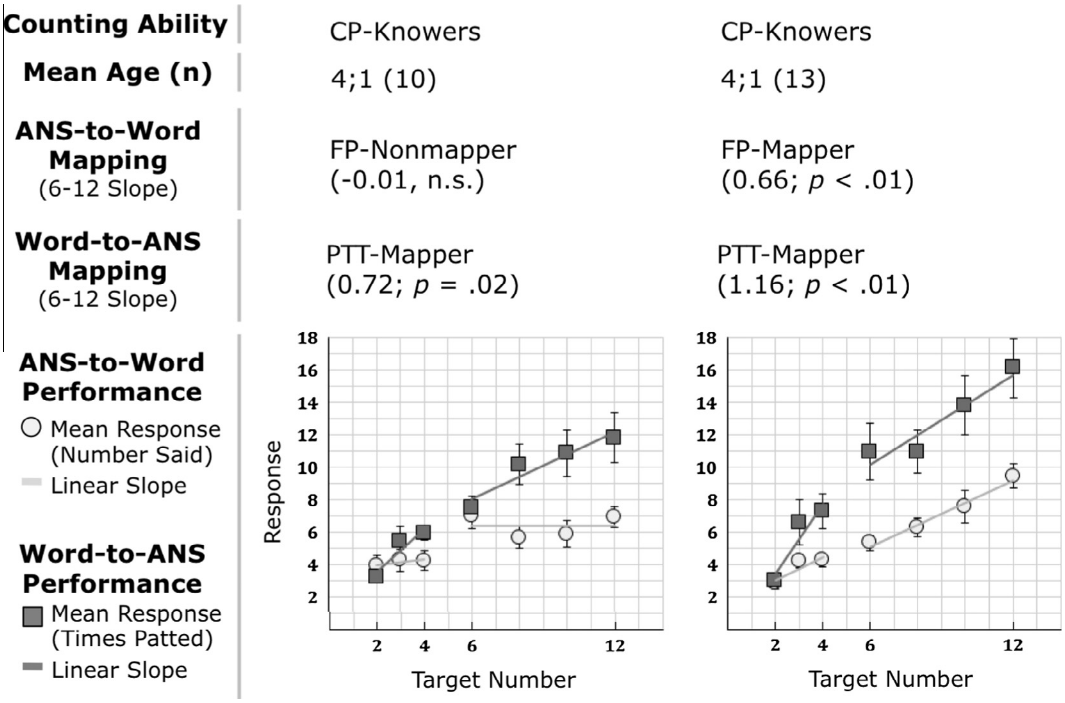

Fig. 7. Performance on the ANS-to-Word (FT) and Word-to-ANS (PTT) tasks split into the two populations of CP-Knowers: FT-Mappers and FT-Nonmappers. Each point is the average guess for that group, and the lines are the best-fitted linear slopes. Error bars are SEM.

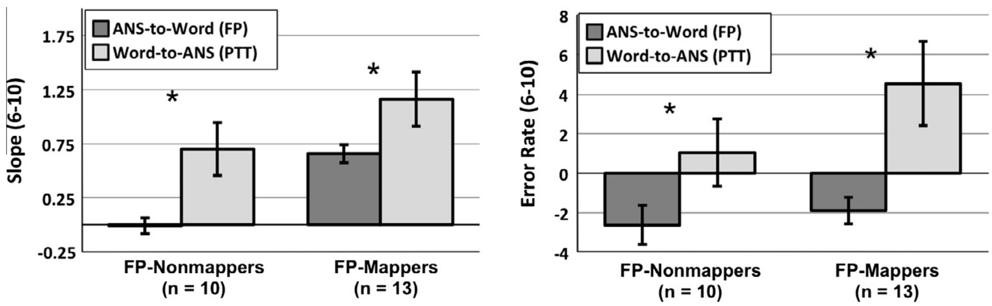

Fig. 8. The best-fit slopes and error rates for the FT-Mappers and FT-Nonmappers (the split CP-Knower group) the 6–10 range on each task. Error bars are SEM, and stars indicate significant difference between the two tasks at $p < .05$.

and 8). Additionally, in replication of Experiment 1, we found that both groups had significantly positive large number slopes in the Word-to-ANS (PTT) task, with FP-Mappers having an average slope of 1.16 (Figs. 7 and 8; SD = 0.91; $t(12) = 4.49$; $p < .001$), and FP-Nonmappers of 0.72 (SD = 0.78; $t(9) = 2.78$; $p < .02$). This means that both FP-Mappers and FP-Nonmappers successfully demonstrated a mapping from Word-to-ANS in the PTT task (i.e., both groups were PTT-Mappers). Furthermore, we found no significant difference between the PTT task large number slopes for FP-Mappers versus FP-Nonmappers (Fig. 8; $t(9) = 1.32$; $p = .22$). These results replicate the results of Experiment 1, while controlling for task demands, and further support to the claim that a mapping from Word-to-ANS emerges developmentally prior to the ANS-to-Word mapping.

When we consider error rates, a similar story emerges. FP-Mappers, just like the FC-Mappers in Experiment 1, under-estimated in the ANS-to-Word task ($M = -1.91$; SE = 0.66) and over-estimated in the Word-to-ANS task ($M = 4.53$; SE = 1.59; Fig. 8). Both of these results are consistent with typical performance in adults. On the other hand, FP-Nonmappers had relatively accurate error rates in the Word-to-ANS task ($M = 1.06$; SE = 1.28) and under-estimated in the ANS-to-Word task ($M = -2.63$; SE = 0.99; Fig. 8). This pattern replicates Experiment 1 – while FP-Nonmappers actually performed better by an objective criterion, it is the FP-Mappers who are performing more similarly to an adult pattern. In this way, error rates also replicate the results of Experiment 1 and provide further evidence that the accuracy of the mappings continue to develop even once the initial interface between number words and ANS representations has been established.

7.5. Alternative strategies analyses

To verify that the FP-Nonmappers in the Word-to-ANS task did not count, we once again performed analyses on patting rates and CVs. As a reminder, if children are counting, their patting rates should be slower than about 550 ms, their patting rates should decrease (i.e., slow down) with higher numbers requested, and their coefficients of variation (CV) should decrease with higher numbers (Cordes et al., 2001; Svenson & Sjöberg, 1983; Trick et al., 1996). In our analyses, we did not find evidence for any of these three signatures. The average patting rate for FP-Nonmappers was far faster than the estimated counting rate for 4-year-olds ($M = 353$ ms; SE = 45 ms), and was not significantly different from that of the FP-Mappers ($M = 368$ ms; SE = 29 ms; $t(10) < 1$). A 2 (Mapper-Level: FP-Mapper, FP-Nonmapper) $\times$ 4 (Number Requested: 6, 8, 10, 12) Mixed-Measure ANOVA performed on individual children's tapping speeds revealed no main effect of Number Requested ($F(3,63) < 1$; $p = .91$), nor an interaction between Mapper-Level and Number Requested ($F(3,63) = 1.01$; $p = .39$). Finally, a 2 (Mapper-Level: FP-Mapper, FP-Nonmapper) $\times$ 4 (Number Requested: 6, 8, 10, 12) Mixed-Measure ANOVA performed on individual children's CV values revealed no main effect of Number Requested ($F(3,54) < 1$), nor an interaction between Mapper-Level and Number Requested ($F(3,54) < 1$). These results all suggest that children in the Word-to-ANS task did not overtly or covertly count.

8. General discussion

In two experiments, we investigated the development of a mapping between two systems of number representation – the ANS and number words. We tested children who were acquiring the meanings of exact number words and gave them both an ANS-to-Word mapping task (Fast Cards in Experiment 1, and Fast Pats in Experiment 2) and a Word-to-ANS mapping task (Pat the Tiger). We investigated the formation of the mapping (slopes), the variability of the mapping (CV) and the accuracy of the mapping (error rates). In both experiments, we found that children attained proficiency at ANS-to-Word earlier than Word-to-ANS (in both slopes and error rates), and that this may account for the contradictory findings in the literature (Le Corre & Carey, 2007; Wagner & Johnson, 2011). Specifically, we found that both Two/Three-Knowers and CP-Knowers showed significantly positive slopes and lower error rates in the Word-to-ANS task compared to the ANS-to-Word task. Furthermore, we found that the group of CP-Knowers who fail to show an existing mapping in the ANS-to-Word task (i.e., FC- and FP-Nonmappers) nevertheless show a functioning mapping in the Word-to-ANS task (i.e., PTT).

Details of children's performance suggest that children engaged in both tasks (e.g., slopes >0 in the 1–4 range) and that the differences we observed in the 6–10 and 6–12 range are not simply because one task is "better" or "easier" than the other – though this, in itself, would have been an important result to demonstrate in the literature. Additionally, these differences are unlikely to have arisen from task demands – our second experiment gave children two serial patting tasks that only differed in the direction of mapping. Finally, we found that the accuracy of the mapping between number words and the ANS continues to develop even after the initial interface is first formed, as seen in error rates.

Why are children presenting with an asymmetry in the mapping between the ANS and number words? We explore two explanations for this result, though we caution that additional work will be required to fully explain the patterns observed here.

As discussed in the introduction, a commonly observed result in adults is that participants under-estimate number in ANS-to-Word tasks (i.e., magnitude estimation), and over-estimate number in Word-to-ANS tasks (i.e., magnitude production); our error rate data replicate this result in children as the average verbal estimation (i.e., in Fast Cards and Fast Pats) was significantly lower than the average number of pats produced (i.e., in Pat the Tiger). Recently, Crollen and colleagues (2011) have suggested that this difference might be driven by the ANS representations being log-distributed, and number words being linearly-distributed; mappings from log to linear (i.e., ANS-to-Word) will produce under-estimation (e.g., log of 50 aligns with linear value of 10), while mappings from linear to log (i.e., Word-to-ANS) will produce over-estimation (e.g., linear value of 10 aligns with log value of 50). One possibility, therefore, may be that our data do not demonstrate an asymmetry in mapping, but rather show that FC/FP-Nonmappers substantially under-estimate on Fast Cards and Fast Pats tasks (hence appearing to have flat slopes), while over-estimating on the Pat the Tiger task (hence having positive slopes).

Two patterns in our data strongly suggest, however, that this explanation cannot account for the difference between FC/FP-Nonmappers and FC/FP-Mappers. Under the log-linear explanation, the mappings are two sides of an identical mapping function; hence, the more log-compressed the ANS scale, the more under/over-estimation there should be, and the magnitude of under-estimation should perfectly correlate with the magnitude of over-estimation. This prediction was verified by Crollen and colleagues (2011), who found that adult observers' error rates on ANS-to-Word tasks strongly negatively correlate with individual error rates on Word-to-ANS tasks: the more over-estimation, the more under-estimation. However, this result is not observed in our data. First, although the Word-to-ANS slopes increase with age, so do the ANS-to-Word slopes; this is exactly the opposite of the prediction made by the log-linear account, because higher over-estimation in Word-to-ANS should be coupled with higher under-estimation in ANS-to-Word. Second, we examined the individual slope and error rate correlations among the FC/FP-Nonmappers: in Experiment 1, we found a non-significant correlation in the wrong direction (Slopes: $r(13) = .25$; $p = .49$; Error Rates: $r(13) = .18$; $p = .55$), and in Experiment 2 we found no significant correlation (Slopes: $r(9) = -.06$; $p = .86$; Error Rates: $r(9) = .05$; $p = .89$). Together, these results strongly suggest that the log-linear explanation cannot explain the observed asymmetry.

Instead, our results appear most consistent with a developmental phenomenon wherein children form a mapping from Word-to-ANS prior to a mapping from ANS-to-Word. Here, we explore possibilities for why such a developmental progression may be natural given the unique challenges posed by each of these mappings. We begin by hypothesizing that the mapping between the ANS and the number words may be two distinct unidirectional mapping functions and not a single bidirectional function. Subsequently, we connect the findings reported here to adult work on ANS and number words, and to the issue of mapping continuous representations to discrete ones more broadly.

There may be a fundamental difference between the procedures that support the translation from a noisy continuous representation (e.g., the ANS) into a discrete, precise representation (e.g., a number word) and the procedures required to support the reverse. Izard and Dehaene (2008) discuss the case of converting a continuous ANS representation into a discrete number word (ANS-to-Word mapping; see also Gallistel & Gelman, 1992). Research on the ANS suggests that approximate number is one of many systems represented as continuous Gaussian representations that obey Weber's law (Bueti & Walsh, 2009; Cantlon et al., 2009; Odic, Libertus, et al., 2013). Such representations can emerge either from a linearly ordered mental number line with linearly decreasing standard deviations or a logarithmically compressed mental number line with constant standard deviation (*for discussion see* Gallistel & Gelman, 1992). Here, we will rely on figures consistent with a linearly ordered number line, though results would be the same in spirit for a logarithmically compressed number line. Fig. 9 displays the hypothesized approximate number representations of the ANS with linearly increasing standard deviations as number increases – thereby instantiating Weber's law. The ANS representation of, for example, *seven dots*, is active across the entire segment of the mental number line that selectively codes for seven items, and also, to a slightly lesser degree, the segments that code for five, six, eight and nine items (Izard & Dehaene, 2008; Nieder & Miller, 2004; Piazza et al., 2004). The ANS representation of e.g., *approximately seven* is the entire continuous curve that is not only active in one spot, but instead spans from roughly 4 to roughly 10. In this way, although the ANS represents something discrete (i.e., the number of objects), it does so through a continuous, real-numbered representation.

Because ANS representations are not associated with a single numerical value along the mental number line, the challenge of translating from a continuous ANS representation of e.g., *approximately seven* into a discrete representation of "seven" involves sampling – the mental computations must somehow reduce the information of the distributed activation into a single value. Izard and Dehaene (2008) suggested that this might happen by instantiating a series of "response bins" that map one-to-one with the exact number words (Crollen et al., 2011; Gallistel & Gelman, 1992; Izard & Dehaene, 2008; Joram et al., 1998). That is, a visual display of 9 dots will activate some corresponding Gaussian along the ANS mental number line (Fig. 10). The subject would then take one or more discrete samples from this continuous activation and determine in which response bin the samples fall. With the response bin activated (either from one sample or the average of multiple samples) the subject could then reply with the verbal label for that response bin (Fig. 10). With the bins properly aligned with the approximate mental number line, the subject could successfully translate the

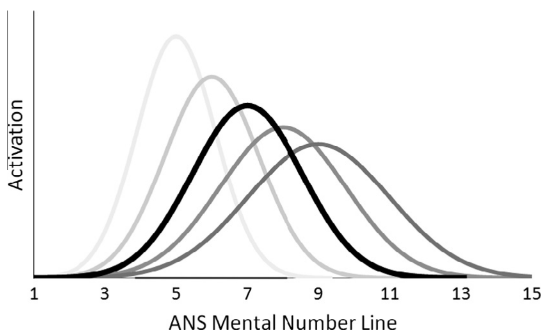

Fig. 9. The hypothesized activation pattern of the ANS. Each activation is continuous and normally distributed (i.e., Gaussian). Highlighted is the activation for seeing 7 dots. Note that although we label the *X*-axis with numbers for convenience, the activation of the ANS is continuous, and no such units actually exist.

continuous ANS Gaussian activation from ANS-to-Word (see also Crollen et al., 2011).

Thus, two potential challenges arise for mapping from ANS-to-Word: sampling and choosing among multiple bins. Children below the FC/FP-Mapper level might have an imprecise sampling procedure for converting the continuous ANS representation into a discrete response bin (or they may not have developed this procedure at all) and, as a result, they cannot perform well on the ANS-to-Word task. Alternatively, all children might be capable of properly sampling, but, in the face of multiple activated bins, children may find it difficult to select *only one* number bin and give up on choosing from these activated options. Under this view (i.e., difficulty selecting among multiple activated response bins), children's performance on Fast Cards and Fast Pats might also suffer if they can recognize that there is only one correct answer to the question ("How many dots were on the card?") and that the many response bins activated on any one trial means that they are unlikely to answer correctly. Presently, our data cannot adjudicate between the sampling and selection proposals. A more formal account of these proposals might be constructed and submitted to empirical test in future work. But, whatever the correct account of children's difficulty in mapping from ANS-to-Word, this difficulty will have repercussions, in practice, for how they use ANS representations to inform their growing sophistication with exact number representations and number words.

We suggest that going from a number word to an approximate ANS value (Word-to-ANS mapping) is a different challenge from the reverse. In this situation, the child is given a precise number word and is asked to somehow reproduce it (e.g., by rapidly tapping). A production problem like this is not naturally solved by a sampling procedure akin to that just described, since the ANS representation must be generated via tapping before it could be sampled and compared to the target value; this would require sampling after each individual tap, thus making a sampling procedure cumbersome.

In contrast to a sampling procedure, one plausible model for a Word-to-ANS mapping involves the building up of an approximate signal as the reproduction behavior (e.g., patting) continues until the representation of the approximate number of pats meets some associatively learned criterion (Cordes et al., 2001; Meck & Church, 1983; Fig. 11). The requested number word may associatively map to a region of the ANS mental number line with either greater (e.g., Fig. 11, "two") or lesser precision (e.g., Fig. 11, "ten"). Such mappings could be learned piecemeal – for example, a child might build a sense of where "ten" falls on the continuum before they have a strong sense of the ordering between "eight" and "nine" (Condry & Spelke, 2008; Nicoladis, Pika, & Marentette, 2010; *and the present data*). When asked to pat "ten times", the subject would activate the corresponding goal region of the ANS number line that is associated with the number word (Fig. 11). With this criterion region activated, the subject would begin patting, and the Gaussian activation representing the currently patted number would grow as patting continues (e.g., in Fig. 11 the Gaussian activation would move rightward as patting occurs). The current activation could be compared to the criterion activity throughout the patting sequence, and the subject would stop whenever the current activation falls within – or moves past – the goal region. This comparison of current activation to goal activation requires the same operation that is performed when discriminating two sets of dots using the ANS, an ability that even rats and newborns, who lack number words entirely, are capable of (Barth et al., 2006; Meck & Church, 1983; Xu & Spelke, 2000).

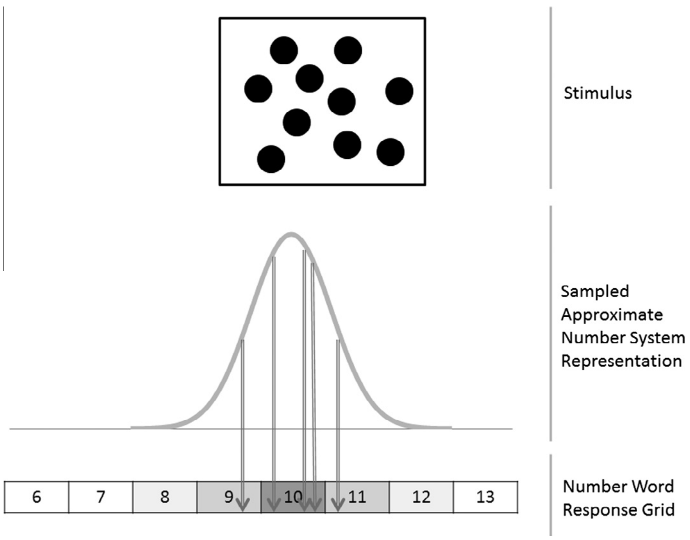

Fig. 10. The hypothesized ANS-to-Word mapping procedure. The top panel illustrates the stimulus which would be briefly displayed. The second panel is the internal representation of approximate number, along a mental number line. The final panel is the response grid that is used in the sampling procedure and that, in turn, actives several numbers, to different degree. Each sample is illustrated by the arrows; given these samples, the most likely number word produced would be "ten".

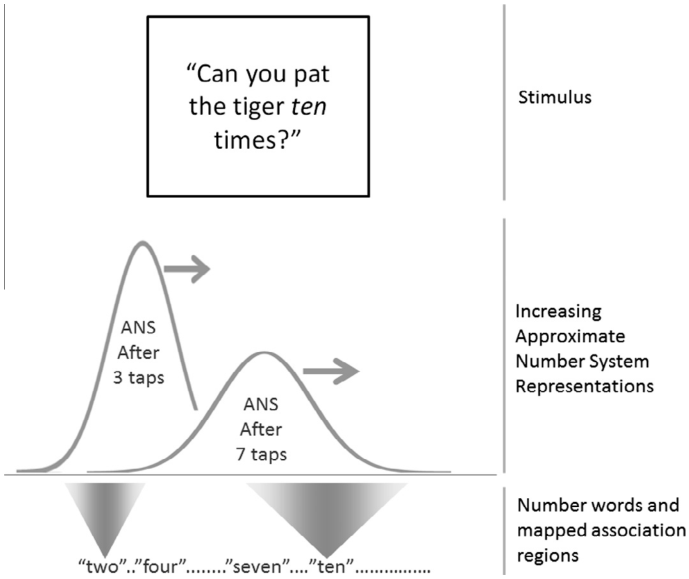

Fig. 11. The hypothesized Word-to-ANS mapping procedure. The top panel illustrates the stimulus – a verbal request for a number. The second panel is the internal representation of the ANS that is increasing (i.e., moving right) with more taps. The third panel represents the number word and the sections of the ANS line that are associated with those words. Notice that the precision can vary. When the activation is aligned with the standard, the tapping stops.

In this model, the child may gradually acquire a rough sense of where along the ANS continuum various number words tend to fall, and number words can capitalize on this functional mapping between long-term memory representations for a criterion and the ANS (Fig. 11). In this way, the challenge of mapping from Word-to-ANS is distinct from any sampling procedure and instead may involve the associative learning of where number words map to the ANS number line. Overall, this model of the Word-to-ANS mapping does not require sampling, can be learned piecemeal for any number word, can have mappings adjusted and improved with increasing experience, and predicts overpatting in Word-to-ANS tasks (e.g., through patting until crossing a criterion). While models of a Word-to-ANS task that involve sampling procedures are certainly possible, we believe that such models would fail to capture the distinction between children's failure in an ANS-to-Word task and their success in Word-to-ANS tasks.

As a result of these differences in the mapping procedures, children may require *both* an ANS-to-Word and a Word-to-ANS mapping before they can demonstrate the full interface between the ANS and the number words. An absence of bidirectionality and the requirement of distinct procedures is consistent with our findings that one direction (i.e., Word-to-ANS) emerges developmentally earlier than the other. Future work will be needed to further develop and verify such models.

Le Corre and Carey (2007) focused their discussion on a claim that ANS representations appear to not play a role in children's acquisition of the exact number concepts (because children became CP-Knowers before demonstrating an ANS-to-Word mapping). It is possible that the successful Word-to-ANS mapping demonstrated here could be productive in helping children to understand the exact number concepts, but a causal mechanism remains to be specified. The evidence from the present study suggests that before children have become CP-Knowers, they are able to map from a discrete number word representation (e.g., "ten") to a region on the continuous ANS mental number line. It may be possible to construct learning proposals in which this association plays a role in the acquisition of the meaning of the number words and new numerical understandings (e.g., Verguts & Fias, 2008 propose that learning number symbols may improve the precision of ANS representations). Such proposals could be a focus of future work.

Our proposal of distinct computations that support the ANS and number word interface is similar to at least some explanations of the production-comprehension asymmetry in children's language acquisition (i.e., the fact that children can understand many more words and grammatical constructions than they can produce; Clark & Hecht, 1983). For instance, when shown an array of plush animals, children asked to point to *doggie* (i.e., a comprehension "Word-to-Animal" task) will only point to plush dogs, but when asked to name various animals (i.e., a production "Animal-to-Word" task), these same children may overextend *doggie* to almost all other animals (Clark, 1983). Several explanations of such over-extensions in language production suggest that young children struggle when forced to choose a single lexical entry for a complex category, and that one strategy may be over-using a single term for a variety of concepts (Clark & Hecht, 1983; Thomson & Chapman, 1977). On the other hand, when children are provided with a lexical item (e.g., point to *doggie*), they are already given the single relevant concept (i.e., *doggie*) and must choose to point at only those items that are a close match to that concept. This kind of explanation is quite similar to one version of our account: an ANS-to-Word mapping is challenging to children because they must select only a single word to describe an entire continuously distributed representation, and a Word-to-ANS mapping is easier to master because children are given the single relevant item (e.g., "ten") and must tap until they reach a representation that is sufficiently similar to this target region of ANS activation.

Regardless of the computations that support the initial formation of the interface between number words and the ANS, there is clear evidence for further refinement of the mapping. In our data, we found that error rates – which index how accurately children calibrate number words to the ANS scale – were different between our two tasks for both Two/Three-Knowers and CP-Knowers. Specifically, in the Word-to-ANS task, while Two/Three-Knowers underestimated, the FC/FP-Nonmappers were very accurate in their responses, while the FC/FP-Mappers significantly over-estimated (as is typical in adults). Hence, the data suggest that after children first acquire the Word-to-ANS interface, they then continue to refine their mapping toward an adult-like state (and, because adults over-estimate when making this mapping, this gives rise to a stage in which children are actually more accurate in their mapping than adults). Our data are also consistent with a recent finding by Ebersbach and Erz (2014), who found that children continue to develop the accuracy of their calibration for a long time in both production and comprehension tasks. Interestingly, in their data, children eventually performed *better* (that is, more accurately) at the ANS-to-Word task compared to the Word-to-ANS task. The kind of asymmetry in the mapping that we are proposing is entirely consistent with this finding: while the ANS-to-Word direction may be more difficult to initially acquire, the development of accurate calibration once the interface is in place may actually be easier.

The present work is also related to aforementioned work by Crollen and colleagues (2011; *see also* Castronovo & Seron, 2007) on the mapping between ANS and exact number representations in adults. The data presented here supports at least part of their hypothesis, as we found that children in both experiments significantly over-estimated in the Word-to-ANS task compared to the ANS-to-Word task. Our work further explores the model of Crollen et al. (2011) by demonstrating that the two directions of mapping do not develop at once. Additionally, our work makes predictions for future adult work; for example, the selection problem of choosing only one bin should also affect adult performance, and adults may show different amounts of internal precision (i.e., CV) in estimation tasks compared to production tasks.

We close by considering number words as a case study for the more general challenge of learning to interface discrete representations (e.g., words) with continuous representations (e.g., ANS). Across multiple literatures, there is evidence for many other contents that rely on continuous representations similar to the ANS; these include representations of approximate area (Brannon, Lutz, & Cordes, 2006; Odic, Libertus, et al., 2013; Odic, Pietroski, Hunter, Lidz, & Halberda, 2013), length (Droit-Volet et al., 2008), time (Droit-Volet et al., 2008; Meck & Church, 1983; Walsh, 2003), speed (Möhring, Libertus, & Bertin, 2012), and many more (Cantlon et al., 2009; Dehaene & Brannon, 2011; Feigenson, 2007). Each of these is available prelinguistically and humans in many cultures eventually master an ability to map from these representations into discrete values and vice versa (e.g., "the speed limit is 130 km/h", "these crayons are various shades of red"). It remains a difficult challenge to determine how such mapping functions operate. In the present study, we found that a mapping from Word-to-ANS may be functional earlier than a mapping from ANS-to-Word. This developmental progression may highlight the extent to which the interfaces between approximate representations (e.g., ANS) and discrete linguistic representations (e.g., number words) may depend on an ability to use language to "point down" toward regions within the prelinguistically-existing approximate spaces rather than on an ability to abstract away from, grow out of, or otherwise increase the precision of pre-existing continuous representations. We suggest that children's success at using the number words to point down to goal regions in the ANS (e.g., when mapping from Word-to-ANS), and their failure to discretize continuous ANS activations into response bins (e.g., when mapping from ANS-to-Word), highlights the importance of mapping from discrete to continuous representations, and the difficulties of mapping from continuous to discrete.

Acknowledgements

This research was funded by NICHD R01 HD057258 awarded to J.H., and by a NSERC PGS-M awarded to D.O. Thank you to all the parents and children to participated in this study. Thank you to Meg Bowen, Andrea Stevenson, and Amy Lai for help in data collection.

References

Ansari, D. (2007). Does the parietal cortex distinguish between "10", "Ten", and Ten Dots? *Neuron, 53*(2), 165–167. http://dx.doi.org/10.1016/j.neuron.2007.01.001.

Baroody, A. J., & Gatzke, M. R. (1991). The estimation of set size by potentially gifted kindergarten-age children. *Journal for Research in Mathematics Education, 22*(1), 59–68. http://dx.doi.org/10.2307/749554.

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. S. (2006). Non-symbolic arithmetic in adults and young children. *Cognition, 98*(3), 199–222.

Brannon, E. M., Lutz, D., & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. *Developmental Science, 9*(6), F59–F64. http://dx.doi.org/10.1111/j.1467-7687.2006.00530.x.

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1525), 1831–1840. http://dx.doi.org/10.1098/rstb.2009.0028.

Cantlon, J. F., Platt, M., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences, 13*(2), 83–91. http://dx.doi.org/10.1016/j.tics.2008.11.007.

Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE, 7*(4), e33832. http://dx.doi.org/10.1371/journal.pone.0033832.

Castronovo, J., & Seron, X. (2007). Numerical estimation in blind subjects: Evidence of the impact of blindness and its following experience. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1089–1106.

Carey, S. (2009). *The Origin of Concepts*. New York: Oxford University Press.

Clark, E. V. (1983). Meanings and concepts. *Handbook of Child Psychology, 3*, 787–840.

Clark, E. V., & Hecht, B. F. (1983). Comprehension, production, and language acquisition. *Annual Review of Psychology, 34*(1), 325–349.

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General, 137*(1), 22–38. http://dx.doi.org/10.1037/0096-3445.137.1.22.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review, 8*(4), 698–707.

Crollen, V., Castronovo, J., & Seron, X. (2011). Under- and over-estimation: A bi-directional mapping process between symbolic and non-symbolic representations of number? *Experimental Psychology, 58*(1).

Crollen, V., & Seron, X. (2012). Over-estimation in numerosity estimation tasks: More than an attentional bias? *Acta Psychologica, 140*(3), 246–251.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123*(1), 162–173. http://dx.doi.org/10.1016/j.cognition.2011.12.013.

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

Dehaene, S., & Brannon, E. (Eds.). (2011). *Space, time and number in the brain: Searching for the foundations of mathematical thought* (1st ed. Academic Press.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex, 33*, 219–250.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*(3–6), 487–506.

Droit-Volet, S., Clément, A., & Fayol, M. (2008). Time, number and length: Similarities and differences in discrimination in adults and children. *Quarterly Journal of Experimental Psychology, 61*(12), 1827–1846.

Ebersbach, M., & Erz, P. (2014). Symbolic versus non-symbolic magnitude estimations among children and adults. *Journal of Experimental Child Psychology, 128*, 52–68.

Feigenson, L. (2005). A double-dissociation in infants' representations of object arrays. *Cognition, 95*(3), B37–B48. http://dx.doi.org/10.1016/j.cognition.2004.07.006.

Feigenson, L. (2007). The equality of quantity. *Trends in Cognitive Sciences, 11*(5), 185–187. http://dx.doi.org/10.1016/j.tics.2007.01.006.

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314. http://dx.doi.org/10.1016/j.tics.2004.05.002.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-bates communicative development inventories, user's guide and technical manual* (2nd ed.). Baltimore: Brookes.

Fuson, K. C. (1987). *Children's counting and concepts of number*. Springer.

Gallistel, C. R. (2007). Commentary on Le Corre & Carey. *Cognition, 105*(2), 439–445. http://dx.doi.org/10.1016/j.cognition.2007.01.010.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1–2), 43–74. http://dx.doi.org/10.1016/0010-0277(92)90050-R.

Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. *Learning and Motivation, 30*, 61–96.

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

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17*(7), 572–576. http://dx.doi.org/10.1111/j.1467-9280.2006.01746.x.

Halberda, J., Taing, L., & Lidz, J. (2008). The development of "most" comprehension and its potential dependence on counting ability in preschoolers. *Language Learning and Development, 4*(2), 99–121.

Hockett, C. F. (1987). *Refurbishing our foundations: Elementary linguistics from an advanced point of view*. John Benjamins Pub Co.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*(3), 1221–1247. http://dx.doi.org/10.1016/j.cognition.2007.06.004.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature, 3*(67), 281–282. http://dx.doi.org/10.1038/003281a0.

Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. *Review of Educational Research, 68*(4), 413–449. http://dx.doi.org/10.3102/00346543068004413.

Krueger, L. E. (1972). Perceived numerosity. *Attention, Perception, & Psychophysics, 11*(1), 5–9.

Landau, B. (2004). Perceptual units and their mapping with language: How children can (or can't?) use perception to learn words. *Weaving a Lexicon*, 111–148.

Landau, B., & Jackendoff, R. (1993). Whence and whither in spatial language and spatial cognition? *Behavioral and Brain Sciences, 16*(02), 255–265.

Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills, 15*(3), 646. http://dx.doi.org/10.2466/pms.1962.15.3.646.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*(2), 395–438.

Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology, 52*(2), 130–169. http://dx.doi.org/10.1016/j.cogpsych.2005.07.002.

Lipton, J. S., & Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76*(5), 978–988.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9*(3), 320–334.

Mejias, S., Grégoire, J., & Noël, M.-P. (2012). Numerical estimation in adults with and without developmental dyscalculia. *Learning and Individual Differences, 22*(1), 164–170.

Mejias, S., & Schiltz, C. (2013). Estimation abilities of large numerosities in Kindergartners. *Frontiers in Psychology*, *4*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3756289/.

Möhring, W., Libertus, M. E., & Bertin, E. (2012). Speed discrimination in 6- and 10-month-old infants follows Weber's law. *Journal of Experimental Child Psychology, 111*(3), 405–418. http://dx.doi.org/10.1016/j.jecp.2011.11.002.

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

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology, 103*(4), 490–502.

Nicoladis, E., Pika, S., & Marentette, P. (2010). Are number gestures easier than number words for preschoolers? *Cognitive Development, 25*(3), 247–261.

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America, 101*(19), 7457.

Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology, 49*(6), 1103–1112.

Odic, D., Pietroski, P., Hunter, T., Lidz, J., & Halberda, J. (2013). Young children's understanding of "more" and discrimination of number and surface area. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 451–461.

Opfer, J. E., Thompson, C. A., & Furlong, E. E. (2010). Early development of spatial-numeric associations: Evidence from spatial and quantitative performance of preschoolers. *Developmental Science, 13*(5), 761–771.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*(3), 547–555. http://dx.doi.org/10.1016/j.neuron.2004.10.014.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron, 41*(6), 983–993.

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

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science, 305*(5691), 1779–1782. http://dx.doi.org/10.1126/science.1100199.

Spaepen, E., Coppola, M., Spelke, E. S., Carey, S., & Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences, 108*(8), 3163.

Svenson, O., & Sjöberg, K. (1978). Subitizing and counting processes in young children. *Scandinavian Journal of Psychology, 19*(1), 247–250. http://dx.doi.org/10.1111/j.1467-9450.1978.tb00327.x.

Svenson, O., & Sjöberg, K. (1983). Speeds of subitizing and counting processes in different age groups. *The Journal of Genetic Psychology, 142*(2), 203–211. http://dx.doi.org/10.1080/00221325.1983.10533512.

Thomson, J. R., & Chapman, R. S. (1977). Who is "Daddy"revisited: The status of two-year-olds' over-extended words in use and comprehension. *Journal of Child Language, 4*(03), 359–375.

Trick, L. M., Enns, J. T., & Brodeur, D. A. (1996). Life span changes in visual enumeration: The number discrimination task. *Developmental Psychology, 32*(5), 924–932.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*(1), 80–102.

Verguts, T., & Fias, W. (2008). Symbolic and nonsymbolic pathways of number processing. *Philosophical Psychology, 21*(4), 539–554. http://dx.doi.org/10.1080/09515080802285545.
Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition, 119*(1), 10–22. http://dx.doi.org/10.1016/j.cognition.2010.11.014.
Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*(11), 483–488. http://dx.doi.org/10.1016/j.tics.2003.09.002.
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. (1992). Children's acquisition of the number words and the counting system∗ 1. *Cognitive Psychology, 24*(2), 220–251.
Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11. http://dx.doi.org/10.1016/S0010-0277(99)00066-9.