Received: 10 August 2015 | Accepted: 21 July 2016
DOI: 10.1111/desc.12495

PAPER

WILEY Developmental Science

Attaching meaning to the number words: contributions of the object tracking and approximate number systems

Kristy vanMarle[1] | Felicia W. Chu[1] | Yi Mou[1] | Jin H. Seok[1] | Jeffrey Rouder[1] | David C. Geary[1,2]

[1]Department of Psychological Sciences, University of Missouri, Columbia, Missouri, USA

[2]Interdisciplinary Neuroscience Program, University of Missouri, Columbia, Missouri, USA

Correspondence:
Kristy vanMarle, Department of Psychological Sciences, 210 McAlester Hall, University of Missouri, Columbia, MO, 65211, USA.
Email: vanmarlek@missouri.edu

Abstract

Children's understanding of the quantities represented by number words (i.e., cardinality) is a surprisingly protracted but foundational step in their learning of formal mathematics. The development of cardinal knowledge is related to one or two core, inherent systems – the approximate number system (ANS) and the object tracking system (OTS) – but whether these systems act alone, in concert, or antagonistically is debated. Longitudinal assessments of 198 preschool children on OTS, ANS, and cardinality tasks enabled testing of two single-mechanism (ANS-only and OTS-only) and two dual-mechanism models, controlling for intelligence, executive functions, preliteracy skills, and demographic factors. Measures of both OTS and ANS predicted cardinal knowledge in concert early in the school year, inconsistent with single-mechanism models. The ANS but not the OTS predicted cardinal knowledge later in the school year as well the acquisition of the cardinal principle, a critical shift in cardinal understanding. The results support a Merge model, whereby both systems initially contribute to children's early mapping of number words to cardinal value, but the role of the OTS diminishes over time while that of the ANS continues to support cardinal knowledge as children come to understand the counting principles.

RESEARCH HIGHLIGHTS

- Analog magnitudes (ANS) and object tracking (OTS) both contribute to cardinal number learning.
- ANS plays a larger role than OTS in predicting conceptual shift to CP-knower.
- Novel multiplicative analysis used to quantify interaction between ANS and OTS over time.
- Influence of OTS on cardinal knowledge wanes over time in preschoolers

1 | INTRODUCTION

The question of when and how children come to understand counting as a principled system has been a matter of study for decades, and remains vigorously debated. Gallistel and Gelman (1992) proposed that children's understanding of counting emerges through bidirectional mappings between the verbal count list ('one', 'two', 'three', …) and the nonverbal magnitudes available in the approximate number system (ANS). The ANS is a core system that represents number as analog magnitudes. In this system, the ability to discriminate two numbers depends on their ratio, rather than their absolute difference, in accord with Weber's Law (Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004; Gallistel & Gelman, 1992). The ANS is present at birth in humans (Izard, Sann, Spelke, & Streri, 2009), and is thought to implicitly embody features that provide children a foothold on the counting principles (i.e., cardinality, stable order, one-to-one correspondence, etc.) (Gallistel & Gelman, 1992).

More recent work, however, suggests an alternative account in which the verbal labels are mapped onto episodic object representations in another core mechanism – the object tracking system (OTS). This system consists of a set of indexes that 'point' to objects in the world, keeping track of them as they move through space and undergo occlusion (Kahneman, Treisman, & Gibbs, 1992; Le Corre & Carey, 2007; Pylyshyn & Storm, 1988). Importantly, the OTS has a limited capacity. It can track only as many objects as it has indexes, which is about 3 in infants and children, and about 4 in adults (Feigenson et al., 2004). At least two sets of indexes can be held in memory and compared on the basis of one-to-one correspondence, allowing infants to discriminate and order two sets of up to three items each (Feigenson & Carey, 2003, 2005; Feigenson, Carey, & Hauser, 2002; vanMarle, 2013).

At the heart of this debate lie empirical findings describing the interesting, and perhaps counterintuitive, developmental progression of children's learning of cardinal value. Wynn's (1990, 1992a) now classic findings suggest that although toddlers may be able to recite the verbal count list up to 'ten' as early as 2 years of age, they may take 1.5–2 additional years to attach meaning (i.e., cardinal value) to the verbal labels (cf. Gelman, 1972, 1993, 2006; also Zur & Gelman, 2004). This progression is revealed in the *GiveN task* in which children are asked to give an experimenter a particular number of objects. Initially, children can give exactly 'one' when requested, but give a random number of objects for all larger set sizes. Such a child is described as a 'one-knower'. Children who can give exactly 'one' and 'two' items correctly are considered 'two-knowers', and children who respond correctly for set sizes 'one', 'two', and 'three' are labeled 'three-knowers'. Individual children tend to progress through these 'knower-levels' as follows: after becoming one-knowers, they take approximately 3–6 months before becoming two-knowers, and several additional months before they become three-knowers. After they come to understand the meaning of 'four', they appear to induce the meaning of the rest of the count list, suddenly transitioning from three-knowers to 'cardinal principle-knowers' (CP-knowers) who can correctly give the exact number requested up to the limit of their count list (Carey, 2004; Le Corre & Carey, 2007; Wynn, 1992a).

This protracted developmental progression is inconsistent with Gallistel and Gelman's (1992) proposal that the number words acquire meaning by being mapped onto ANS representations (see Wynn, 1992a, for a similar argument against Gelman and Gallistel's 1978 counting principles theory). If they did, the task should be an easy one because the two systems are thought to share the same structure. Like ANS representations, the verbal count list has a stable order, is applied in a one-to-one fashion to the counted items, and honors the cardinality principle (i.e., the last item counted represents the cardinal value of the set) (Gallistel & Gelman, 1992). Because there are no discontinuities in how small (i.e., 1, 2, and 3) and large (> 3) numbers are represented in the ANS, there should not be discontinuities in children's learning of the meanings of the number words, like we see in the GiveN task. In addition, the imprecision of ANS representations poses a problem for explaining how children come to understand number words as denoting exact numerosities (Le Corre & Carey, 2007; Leslie, Gallistel, & Gelman, 2007).

More recently, Carey (2004) and Le Corre and Carey (2007) proposed that the developmental progression is an example of genuine conceptual change, and that the OTS is the representational system underlying this change. On their view, the meanings of children's first few number words derive from their representations of small sets of individual items in the OTS.[1] The argument is that children learn the first few number words as natural language quantifiers by extension, mapping the verbal labels onto sets of individual items. For example, a child attending to two objects will use the OTS to set up a working memory model, i_a, i_b, that can be stored in long-term memory and, over time, becomes associated with the heard verbal label 'two'. These mappings for the words 'one', 'two', and 'three' then go on to support the induction of the successor principle – that moving to the *next word* in the count list is equivalent to *adding one individual item* (Carey, 2004; Le Corre & Carey, 2007). On this account, the slow and piecemeal acquisition of meaning reflects the time it takes to develop strong associations between the heard labels and models of sets of individual objects in long-term memory; the discontinuity emerges with children's induction of the counting principles, especially the successor principle.

Despite the better fit the OTS-only model provides to the developmental pattern revealed by the GiveN task, it too falls short as a full explanation of how children learn to count. In particular, Gallistel (2007) raised concerns about Le Corre and Carey's (2007) study, as well as with the general model forwarded by Carey (2004). The majority of Gallistel's objections rest on the fact that the OTS creates symbols to represent individual objects, but has no symbol to represent the total number of items in the set, that is, its cardinality. Without a representation of cardinality, the OTS lacks critical numerical content, making it unclear how the OTS could imbue the count words with quantitative meaning.

At this time, it appears that neither system alone can explain how children fix meaning to the verbal count list. There is a third view, however, proposed initially by Spelke and Tsivkin (2001a), and more recently Spelke (2011), in which both the ANS and the OTS play a role. As with Carey (2004), Spelke regards children's acquisition of counting as an instance of conceptual change, but one in which language (i.e., the verbal count list) brings together the two core systems to create a new concept. The ANS generates cardinal values for a wide range of magnitudes, but with comparatively little precision; the OTS produces exact representations, but only for small numbers of individuals and without cardinal value. When combined through language, they together support a new system of verbal counting, allowing the child to represent and generate sets of any size with precise cardinal values (Spelke, 2011). To date, however, this view (which we call *Combine*) stands as a theoretical model that, to our knowledge, has not been empirically tested.

As with Le Corre and Carey's (2007) OTS-only model, the Combine model assumes that the sudden shift to CP-knower status is the result of the child inducing two principles – the successor principle and the cardinal principle. To induce the successor principle, the child comes to notice that *adding (another) one* object index (a function assumed to be inherent in the OTS) produces a set whose cardinality is named by the *next number* in the count sequence. Inducing the cardinality principle involves the realization that each count word labels 'a set of individuals with a unique cardinal value' (Spelke, 2011, p. 305). On the OTS-only model, neither of these inductions is possible because there is no cardinal representation of the set of items active in the OTS, making it impossible to notice that the cardinal value changes when an index is added. On the Combine model, however, the child uses language to link the two systems. Once linked, when the child sees a set of two objects, the OTS, ANS, and the verbal label 'two' are all simultaneously activated. After the first three or four words are linked in this way, the child can induce the principles because the word accesses both precise sets of individuals and cardinality, capitalizing on the strengths and overcoming the limitations of each system.

Although an improvement over the two single-mechanism models, one potential problem with the Combine model is that it adopts the OTS 'as is' from the views of Carey and colleagues (Carey, 2004; Feigenson et al., 2002; Feigenson et al., 2004; Le Corre & Carey, 2007), thus incorporating the limitations of that view. For instance, although inducing the successor principle in this model makes use of the cardinal meaning available in the ANS, there is still the problem that the *add (another) one* function itself does not have numerical content. Without numerical meaning, *adding another one* connotes only a spatial relation in the OTS (i.e., place 'this' near 'that'). As discussed by Leslie et al. (2007), having an exact representation of two individuals by virtue of their spatial locations (i.e., their active indexes) is not equivalent to having an exact representation of 'twoness', nor is *add another one* equivalent to '1 + 1'; assuming numerical content here is begging the question, making it unclear whether simply combining the two systems really solves the problem. Although language could be the glue that gives the sets of indexes numerical meaning, the process by which this might happen (beyond mere association) has yet to be specified. Moreover, from a purely theoretical perspective, because the ANS already instantiates the successor function and the cardinality principle in structure and process, it may be more parsimonious to conclude that the ANS bears most of the weight in the child's transition to CP-knower.

Indeed, although we favor a dual-mechanism approach, we believe the Combine model may also fall short in some respects. We, therefore, suggest a modified dual-mechanism view, the *Merge* model, which differs from Spelke's (2011) model in two ways. First, on our reading of Spelke's model, the combining of the ANS and OTS through the acquisition of verbal counting predicts that both systems should continue to support reasoning about cardinality over time; once the systems are linked (the OTS, ANS, and verbal count list), the links remain intact into adulthood (e.g., Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Spelke & Tsivkin, 2001b). The Merge model, in contrast, proposes that while the OTS may play a significant role early on (i.e., by offering precision, as a potential source of the successor principle), its influence will diminish over time. Thereafter, although both systems may continue to coexist and be utilized, the ANS may be preferred for reasoning about cardinality during childhood, and possibly into adulthood, although the latter is debated (e.g., Burr, Turi, & Anobile, 2010; Cordes, Gelman, Gallistel, & Whalen, 2001; Hyde & Spelke, 2009, 2010; Revkin, Piazza, Izard, Cohen, & Dehaene, 2008).

Second, while the Combine model proposes that the transition to CP-knower is the result of the induction of the cardinal and successor principles, we propose instead that the transition results from the OTS moving into the background once the capacity limit is breached, which allows the child to capitalize on the quantitative knowledge inherent in the ANS. Many studies suggest that the OTS often trumps the ANS in infancy, particularly in contexts involving three-dimensional objects being tracked through time and space (as in the GiveN task and other common counting situations faced by children) (for review, see Mou & vanMarle, 2014). One reason the OTS competes successfully with the ANS early in development may be its precision, at least for sets within its capacity. However, the precision of the ANS improves dramatically throughout infancy and early childhood (Halberda & Feigenson, 2008; Izard et al., 2009; Lipton & Spelke, 2003), and becomes precise enough to provide reliable discrimination within the small number range (successful at 3:4 ratios) by the ages tested here (3–4 years). At this point, children may shift towards relying primarily on the ANS, giving them access to cardinal representations within and beyond the limit of the OTS, as well as the successor function, and allowing them to recognize the structural congruency between the ANS and the counting sequence. In other words, because the ANS appears to already possesses the properties necessary for the child to understand counting (i.e., cardinality, successor function, ordinal structure, etc.), these principles do not need to be induced, only mapped to counting words. And hence, there is no need for an induction-based conceptual change, but rather an explicit recognition that counting words map onto and refine their intuitive number sense. The OTS is necessary, therefore, primarily as a source of precision early on, possibly by 'calibrating' the first few magnitudes in the ANS and giving rise to the notion of 'exact integer'.[2]

1.1 | The present study

Previous arguments for the ANS-only or OTS-only models of the emergence of cardinal knowledge have relied on indirect data (e.g., Carey, 2004; Gallistel & Gelman, 1992; Le Corre & Carey, 2007). That is, children's performance on some tasks (e.g., GiveN) has been interpreted as being consistent or inconsistent with a given model, and none to our knowledge have independently examined children's OTS abilities and whether they are related to cardinal knowledge. The present study thus builds on this earlier work by directly testing whether measures of the ANS, OTS, or both are related to the development of cardinal knowledge. Moreover, we assessed children's general cognitive abilities, including IQ, executive function, and preliteracy skill, and collected demographic information (parental education, income, race/ethnicity). Thus, unlike previous studies, we were able to test the relationship between ANS, OTS, and cardinality above and beyond the contribution of these other factors that are known to influence children's mathematical development (Bull, Espy, & Wiebe, 2008; Clark, Sheffield, Wiebe, & Espy, 2013; Geary, 2011; LeFevre et al., 2010).

As part of an ongoing longitudinal study that is focused on identifying the early quantitative knowledge that contributes to later mathematical achievement, children completed ANS, OTS, and GiveN tasks at two time points during their first year of preschool. We were able, therefore, to examine concurrent relations between the two systems and cardinal knowledge, as well as the relative contributions of the two systems to change in cardinal knowledge over time, allowing us to determine which of the four models – ANS-only, OTS-only, Combine, or our Merge model – best explains the development of cardinal knowledge in children.

Because neither the ANS nor the OTS alone can support the acquisition of cardinal knowledge, we did not expect to find evidence for either single-mechanism model. Instead, we expected our data to support one of the two dual-mechanism views, each of which makes different predictions about the role of the OTS as summarized in Table 1. If the Combine model is correct, then both the ANS and OTS should be related to cardinal knowledge at both time points, and both should predict the likelihood of transitioning from 'non-CP-knower' at time 1 (T1) to 'CP-knower' at time 2 (T2). In contrast, the Merge model predicts that both systems should be related to cardinal knowledge at T1, but only the ANS should remain related at T2. And, because the ANS is the primary source of cardinal knowledge, it alone should predict the transition to CP-knower.

2 | METHOD

For a full description of the methods used, please refer to Chu, vanMarle, and Geary (2013). For space considerations, here we just include details relevant to the present study.

2.1 | Participants

Two hundred and twenty-eight children were recruited from the Title I preschool program within the public school system in Columbia, MO. Children were recruited in three cohorts (A, B, and C), entering in the fall of 2011, 2012, and 2013. We excluded data from 14 children due to low (< 70) intelligence scores and 16 other children who had moved, refused to participate, or were otherwise unable to complete most of the tasks. Data from the remaining 198 (96 boys) children were used in the analyses. At their first assessment, children averaged 3 years 10 months of age (range: 3y2 m–4y4 m), and 4 years 2 months (range 3y6 m–4y10 m) at their second assessment.

2.1.1 | Demographic information

We collected parent reported demographic information for a subset (*n* = 154) of our sample, and not all provided responses to all questions. In terms of ethnicity, our sample was 84% non-Hispanic, 10% Hispanic/Latino, and 6% reported their child's ethnicity as 'unknown'. In terms of race, our sample was 55% White, 23% Black, 8% Asian, 13% more than one race, and 1% reported their child's race as 'unknown'. Our sample was primarily lower- to middle-class economically. Self-reported total household income was: $0–$25k (36%), $25k –$50k (25%), $50k–$75k (25%), $75k–$100k (13%), $100k–$150k (1%), $150k or more (1%). Forty percent of respondents reported receiving food stamps, and 11% reported receiving housing assistance.

2.1.2 | Parental education

Parental education for mothers and fathers, respectively, was: some high school (12%, 15%), complete HS/GED (52%, 43%; with 54% and 41% reporting some college), bachelor degree (26%, 19%), and postgraduate degree (10%, 23%).

Maternal and paternal education levels were highly correlated, $r_{146} = .67$, $p < .0001$, and thus we created a mean parental education variable ($\alpha = .80$). This was then used to create three groups: no information provided (*n* = 43), high school (*n* = 87), and college (*n* = 68). This contrast was used as a covariate in the reported analyses.

2.2 | Cognitive and achievement measures

In addition to the three tasks of interest (described below), children completed standardized measures of intelligence (Wechsler Preschool and Primary Scale of Intelligence-III) (Wechsler, 2002), executive function (Conflict EF scale) (Beck, Schaefer, Pang, & Carlson, 2001), and preliteracy skills (Phonological and Literacy Screening) (Invernizzi, Sullivan, Meier, & Swank, 2004).

2.3 | Quantitative measures

The primary measures of interest consisted of three tasks – ANS, OTS, and Cardinal Knowledge – given as part of a larger quantitative

TABLE 1 Predicted relationship between each of the two systems (ANS and OTS) and cardinal knowledge over time for the four models

	Single-mechanism		Dual-mechanism	
	ANS-only (ANS/OTS)	OTS-only (ANS/OTS)	Combine (ANS/OTS)	Merge (ANS/OTS)
Time 1	+/−	−/+	+/+	+/+
Time 2	+/−	−/+	+/+	+/−
Transition to 'CP-knower'	+/−	−/+	+/+	+/−

Note. '+' indicates that the system should be positively related to cardinal knowledge and reliably predict CP-status; '−' indicates that the system should not be related to cardinal knowledge or predict CP-status.

battery (including nine additional tasks) that children completed twice over the course of their first year of preschool.

2.3.1 | Analog number precision (ANS)

ANS acuity was measured using the Panamath program (Halberda & Feigenson, 2008) presented on a laptop computer. Children were simultaneously presented with two dot arrays (a group of blue dots and a group of yellow dots, presented on different sides of the screen) and asked to report which array 'ha[d] more dots'. The experimenter pressed a button indicating the array the child selected. Across trials, arrays were controlled for non-numerical cues, such as density and surface area, ensuring that children's responses were based on number, per se. The data yielded both accuracy (% correct) and an estimate of the Weber fraction (ω), which is essentially a measure of the precision of the underlying magnitude representations. Because previous work (Inglis & Gilmore, 2014; vanMarle, Chu, Li, & Geary, 2014) showed that, compared to ω, accuracy was a more stable predictor of various quantitative skills and mathematics achievement, we focused on accuracy here.[3] Children in the first cohort (A) completed 24 test trials. Based on results for this cohort, we added six relatively easy trials for cohorts B and C (30 trials total). The ratio of blue:yellow dots was determined randomly on each trial and varied between 1.29 and 3.38 (cohort A), and 1.29 and 4.0 (cohorts B and C). All dot displays contained more than 3 dots, to discourage verbal counting, and were displayed for 2533 ms, followed by the response period. Children were not provided feedback on their accuracy.

2.3.2 | Object tracking ability (OTS)

To measure children's object tracking abilities, we used the *magic box* task. The task was specifically designed as an analogue to Wynn's (1992b) classic addition/subtraction task which is believed to engage the object tracking system (Kibbe & Leslie, 2013; Leslie, Xu, Tremoulet, & Scholl, 1998; Simon, 1997) and may also be considered a variant of Starkey's (1992) search box task (vanMarle & Wynn, 2001). Importantly, because the task requires children to track small sets of bounded 3-D objects as they move through space and undergo occlusion, it should facilitate engagement of the OTS (Kibbe & Leslie, 2013; Mou & vanMarle, 2014).

Children were first introduced to a puppet and a colorful box and told that the puppet sometimes does magic tricks on toys when they are inside the box. Following four familiarization trials involving non-numerical, but still 'magical' changes (e.g., a cow turned into a frog), children completed eight test trials. In each test trial, children watched the experimenter hide 0, 1, or 2 objects in the box, and then either add or remove an object from the hidden set before closing the lid. Children were never allowed to see the resulting set. Finally, the lid was opened to reveal the correct or incorrect number of objects, and children had to report whether the puppet had done a magic trick. Four items were tested (0 + 1, 1 + 1, 1−1, 2−1), resulting half of the time in numerically correct outcomes (not magical) and half the time in numerically incorrect outcomes (magical). The dependent measure was accuracy (% correct) on the test trials.

Because the maximum number of objects children had to track (3 identical objects, in 2−1 = 2) was within the capacity limit of the OTS (Cheries, Wynn, & Scholl, 2006; Feigenson & Carey, 2003, 2005; Feigenson et al., 2002; Kibbe & Leslie, 2013; Wynn, 1992b; Zosh & Feigenson, 2009), the individual differences measured by this task were not necessarily a reflection of the *number of items* that could be tracked. Indeed, although the capacity-limited nature of the OTS is often considered its signature property, there are many processes involved in tracking that could affect children's performance. In order to successfully track a set of individuals through occlusion (as in Wynn, 1992b, and our magic box task), children must assign one and only one index to each tracked item, maintain active indexes over time and occlusion, and then reassign each index to its object once visible. Poor implementation of any of these processes will affect performance. Indeed, capacity limits seen in adults are thought to reflect not only a limit on the number of objects that can be tracked, but also individual differences in selecting and maintaining targets (i.e., individuation and sustained attention) (Drew & Vogel, 2008), as well as an individual's ability to store only relevant feature information and avoid storing irrelevant information in working memory (Vogel, McCollough, & Machizawa, 2005). Thus, our measure of OTS ability likely reflects individual differences in the integrity and application of these processes, as revealed through overall task accuracy, rather than individual differences in capacity limits.

2.3.3 | Knowledge of cardinality

We measured children's cardinal knowledge using the GiveN task (Wynn, 1990). Children were asked to put exactly 1, 2, 3, 4, 5, or 6 toy cookies on a plate for the puppet to eat. Using the standard titration method, children began at set size 1, advanced to the next set size following a correct response, and went down one set size following an incorrect response. This continued until they failed on 2 of 3 attempts, and the highest set size they correctly gave at least twice was considered the highest set size for which the child understood cardinality (Le Corre & Carey, 2007).

2.4 | Procedure

Children were tested individually in 35-minute sessions six times during their first year of preschool. The three quantitative tasks (ANS, OTS, and Cardinality) were all part of the same testing session and were completed twice, at the beginning of the Fall (T1) and Spring (T2) semesters. Intelligence, executive function, and preliteracy were assessed in a single session at the end of the Fall semester, and mathematics achievement was assessed in a single session toward the end of the school year.

Testing sessions were videotaped for the purposes of coding and measuring reliability. Data were coded from the videotaped records and computer output (ANS task). To determine reliability, trained observers who were naïve to the purposes of the study watched the recorded sessions and recoded data from the GiveN and magic box tasks for 24 randomly selected participants (12% of the sample). Data from the original coding and the recoding were correlated separately for each task at each time point. Reliability scores between the original and recoded data for the two tasks across the two time points were very high, ranging between .84 and 1. Thus, the original codings were used for all analyses. The test–retest reliability was highly significant for ANS (r_{169} = .41, p < .0001), OTS (r_{171} = .39, p < .0001), and Cardinality (r_{187} = .66, p < .0001). Split-half reliability for the ANS task used here has been reported elsewhere as .54 (Odic, Libertus, Feigenson, & Halberda, 2013), and for the OTS task was calculated for our sample to be .34 and .66 at T1 and T2, respectively. Given the ages being tested (3–4 years) and the substantial developmental change that occurs over this period, these reliabilities reflect an acceptable degree of consistency for the measures.

3 | RESULTS

In general, children performed well on all three tasks at both time points. Accuracy on the ANS task was reliably above chance at T1 (M = 67%, SD = 17.15; t_{186} = 13.68, p < .0001) and T2 (M = 71%, SD = 20.09; t_{179} = 14.12, p < .0001), as was accuracy on the OTS task (T1: M = 67%, SD = 17.97, t_{182} = 13.00, p < .0001; T2: M = 71%, SD = 19.01, t_{183} = 15.00, p < .0001). Because number and surface area of the dot arrays in the ANS task were confounded for half the trials and equated for the other half, we were able to examine whether performance differed when non-numerical cues were controlled. Paired samples t-tests comparing performance on confounded and equated trials were not significant at either T1 (M_{con} = 66%, SD = 19.84, M_{eq} = 67%, SD = 19.03,; t_{184} = −.527, p = .599, two-tailed) or at T2 (M_{con} = 71%, SD = 22.10, M_{eq} = 71%, SD = 20.32,; t_{184} = −.376, p = .707, two-tailed), although performance was significantly above chance (50%) at both time points, and for both trial types (all ts > 11.2, all ps < .0001, two-tailed). This suggests that non-numerical cues had little influence on children's performance in our study.[4]

Both tasks showed comparable performance and variability around the means. To examine change in performance over time, we conducted a 2 × 2 repeated measures ANOVA with Time (T1 and T2) and Task (ANS and OTS) as within-subject factors. Results revealed a main effect of Time (F[1, 156] = 14.07, p < .0001), but not Task (F[1, 156] = .12, p = .735), suggesting that although overall performance increased from T1 (M = 68%) to T2 (M = 72%), children performed equivalently on the ANS and OTS tasks. Moreover, the Time × Task interaction was not significant (F[1, 156] = 1.23, p = .270), indicating that the magnitude of change was similar for both tasks. On the Cardinality task, children correctly gave an average of 3.25 items (SD = 1.99) at T1, and 3.98 items (SD = 1.95) at T2, showing a significant improvement across time (t_{186} = 6.14, p < .0001). Importantly, performance on the ANS and OTS tasks was not related at T1 (pr_{123} = −.06, p = .512) or T2 (pr_{121} = .09, p = .331), controlling for sex, race, parental education, age, IQ, executive function, and preliteracy, suggesting that the two tasks were tapping distinct sets of skills.

3.1 | Does ANS and/or OTS predict CP-knower status?

To examine whether ANS, OTS, or both systems were related to children's cardinal knowledge, we explored the concurrent relations between the ANS and OTS to Cardinality by conducting separate binary logistic regressions for each time point (all p-values are one-tailed). Initial analyses with just ANS and OTS entered as predictors of CP-knower status (*non-CP-knower* = GiveN score of 0–3, or *CP-knower* = GiveN score of 4–6) revealed a pattern consistent with the Merge model. At T1, both ANS (β = .781, p < .0001) and OTS (β = .381, p = .014) significantly predicted Cardinality. However, at T2, ANS (β = .723, p < .0001) was still a significant predictor of Cardinality, while OTS was not (β = .222, p = .098).

As noted earlier, in addition to the variables of interest, we also collected demographic information (race, parental education) and assessed several domain-general cognitive abilities (IQ, executive function, and preliteracy skills), allowing us to make an especially strong test of the models. As shown in Table 2, the pattern was the same as in the initial analyses with inclusion of all covariates (sex, race, age, IQ, EF, preliteracy, and parental education). Both the ANS (β = .742, p < .004) and OTS (β = .552, p = .018) significantly predicted CP-knower status at T1; at T2, however, ANS accuracy remained a significant predictor (β = .628, p = .039), but OTS accuracy did not (β = .259, p = .164).[5] Note also that age (β = .601, p = .042), IQ (β = .723, p = .022), and preliteracy skill (β = .714, p < .012) were all significant predictors of CP-knower status at T1, but at T2, only age (β = 1.048, p = .002) and preliteracy (β = 1.172, p = .001) remained significant. Together, these analyses are consistent with the Merge model and indicate that both the ANS and OTS variables were related to CP-knower status at T1, above and beyond the contribution of demographic factors and domain-general cognitive abilities. Then, by T2, the role of the OTS diminished, whereas the ANS continued to be predictive.

3.2 | Does ANS and/or OTS predict the transition to CP-knower?

To resolve the critical question of whether the ANS, OTS, or both, drive the transition to CP-knower status, we first identified children who were non-CP-knowers at T1, and then categorized them into two groups: *Change* (n = 37, moved from '1–3-knower' at T1 to '4–6-knower' at T2) or *No Change* (n = 73, '1–3-knower' at T1 and T2). A binary logistic regression with T1 ANS and OTS as predictors of group membership (Change or No Change), and with sex, race, and age, as well as IQ, EF, preliteracy, and parental education simultaneously entered as covariates, revealed that a 1 SD increase in ANS accuracy at T1 (β = .99, p = .009) was associated with a 7.3-fold increase in the odds of being in the Change group at T2. The corresponding odds for the OTS task were 4.3, but the relationship was not significant (β = .41, p = .153). Importantly, this relationship was not driven simply by children's knower-level at T1. When included as an additional covariate, T1 knower-level was not itself significant (β = .66,

TABLE 2 Parameter estimates of binary logistic regressions. ANS and OTS accuracy predicting CP-knower status ('non-CP-knower' or 'CP-knower'), shown for the basic and full models at T1 and T2 (*p*-values for ANS and OTS are one-tailed because the Merge model makes directional predictions)

	Time 1			Time 2		
	Variable	Estimate of β	*p*	Variable	Estimate of β	*p*
Basic model	ANS	.781	**.000**	ANS	.723	**.000**
	OTS	.381	**.014**	OTS	.222	.098
	Constant	−.385	.022	Constant	.345	.038
Full model	ANS	.742	**.004**	ANS	.628	**.039**
	OTS	.552	**.018**	OTS	.259	.328
	Sex	.101	.842	Sex	−.338	.537
	Race	–	.638	Race	–	.467
	Age	.601	**.042**	Age	1.048	**.002**
	Parent Educ.	–	.780	Parent Educ.	–	.602
	EF	.352	.236	EF	.014	.963
	IQ	.723	**.022**	IQ	.194	.618
	Preliteracy	.714	**.012**	Preliteracy	1.172	**.001**
	Constant	−19.243	1.0	Constant	1.21	.132

$p = .07$), while ANS ($\beta = 1.19$, $p = .005$), but not OTS ($\beta = .44$, $p = .146$), continued to reliably predict change in CP-knower status.

The fact that ANS accuracy predicted children's transition to CP-knower is strong evidence that the ANS is involved; however, the lack of an effect for OTS does not necessarily mean that it was not involved. That is, if all the variation in OTS accuracy was over a finer span at the higher end of the performance range, then we might fail to find a statistical relationship between OTS accuracy and the transition to CP-knower.[6] In such a case, children's OTS accuracy might be strong enough to support the transition despite the variability being too small for us to detect the relationship. However, our data suggest this was not the case. As can be seen in Figure 1 and Table 3, children's performance on the OTS and ANS tasks showed comparable variability over a very large range, and this was true for both CP-knowers and non-CP-knowers. It is also clear from the data that variability in OTS accuracy was not restricted relative to variability in ANS accuracy, and furthermore, the fact that the range of scores was similar for both CP-knowers and non-CP-knowers (and even greater for CP-knowers in some cases) argues strongly against this possible alternative.

The large degree of overlap between the tasks and across CP-knower groups also argues against another alternative explanation, which is that the obtained pattern might be driven by CP-knowers alone. Negen and Sarnecka (2015) recently published data suggesting that non-CP-knowers may not understand the instruction in the ANS task to choose the set with 'more dots'. Specifically, children in their dataset who were not yet CP-knowers often performed at chance, or chose on the basis of surface area, rather than number. According to these authors, moving from non-CP-knower to CP-knower leads to an increase in performance not because better ANS acuity drives the transition, but instead because CP-knowers suddenly understand the task. However, one-sample *t*-tests comparing CP-knowers and non-CP-knowers ANS accuracy to chance (50%), showed that both groups performed significantly above chance at both T1 ($M_{non\text{-}CP} = 61\%$, $SD = 15.12$, $t_{106} = 7.69$, $p < .0001$, two-tailed; $M_{CP} = 75\%$, $SD = 16.93$; $t_{744} = 12.91$, $p < .0001$, two-tailed) and T2 ($M_{non\text{-}CP} = 63\%$, $SD = 15.48$, $t_{77} = 7.69$, $p < .0001$, two-tailed; $M_{CP} = 78\%$, $SD = 19.93$, $t_{102} = 14.07$, $p < .0001$, two-tailed). This suggests that both groups of children understood the task. In addition, the large degree of overlap in performance between our CP-knowers and non-CP-knowers, as well as

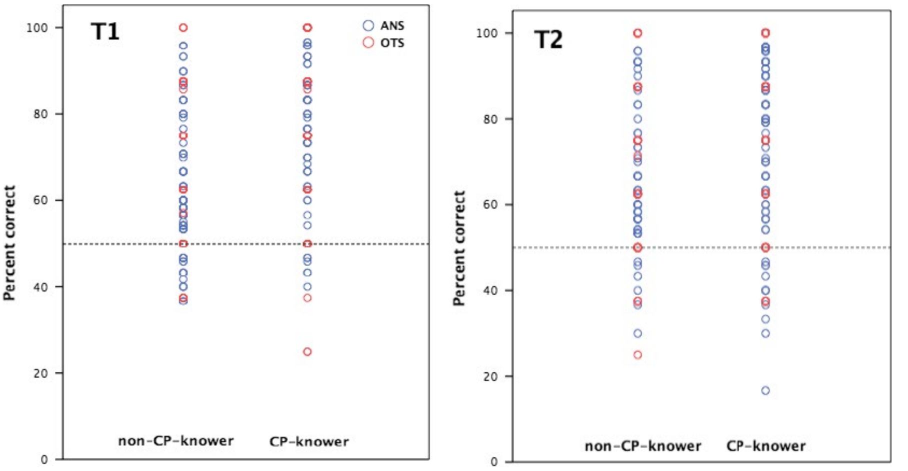

FIGURE 1 Scatter plots of percent correct on ANS and OTS tasks at T1 (left panel) and T2 (right panel) as a function of CP-knower status. Range of scores and variability are highly similar across both tasks, regardless of CP-knower status

TABLE 3 Descriptive statistics for ANS and OTS tasks at T1 and T2, as a function of CP-knower status. Despite overall better performance for CP-knowers, there is substantial overlap in performance between the two groups for both tasks, and at both time points

	T1				T2			
	ANS		OTS		ANS		OTS	
	CP−	*CP+*	*CP−*	*CP+*	*CP−*	*CP+*	*CP−*	*CP+*
N	109	75	105	75	76	101	76	107
Mean	62%	75%	64%	71%	64%	76%	67%	74%
Median	58%	77%	63%	75%	60%	83%	63%	75%
Std. Deviation	15.14	16.93	16.26	19.56	15.65	20.03	18.35	19.18
Variance	229.3	286.7	264.5	382.6	244.8	401.13	336.8	367.89
Range	59	60	80	75	70	83	75	63
Minimum	37%	40%	20%	25%	30%	17%	25%	38%
Maximum	96%	100%	100%	100%	100%	100%	100%	100%
Interquartile range	50%–71%	63%–88%	50%–75%	50%–88%	53%–75%	63%–95%	50%–88%	50%–88%

Note. 'CP−' indicates 'non-CP-knower' and 'CP+' indicates 'CP-knower'.

control of IQ and executive functions (which should influence individual differnces in task understanding), suggests that at least in our sample, this is not the case. Instead, we would argue that the variability in both groups is meaningful, and that chance performance is likely reflecting children's poor ANS acuity rather than their misunderstanding the task instructions. The same logic applies to the OTS task; but the fact that variability overlaps for this task too actually provides evidence in favor of our Merge model, precisely because OTS accuracy was not predictive of the transition to CP-knower despite substantial variability in performance.

3.3 | Characterizing the role of OTS

As shown in Table 1, the predictions made by the Merge model are identical to those made by the ANS-only model, save for the expectation that both systems will be related to cardinal knowledge at T1. The involvement of the OTS at T1, however, is consistent with at least three different possibilities regarding the nature of its role. One possibility is that OTS ability might have a relatively constant effect across children's ANS ability, such that a given increase in OTS ability for different levels of ANS will lead to similar increases in the probability that a child will be a CP-knower. Here, the OTS can be seen as playing a *complementary* role in determining children's cardinal knowledge. Another possibility is that OTS ability has a varying effect depending on the child's ANS ability. For instance, finding that OTS is more influential for children of lower ANS ability relative to those of higher ANS ability would be consistent with the OTS playing a *compensatory* role in rescuing the numerical representations of children hampered by poor ANS ability. On the other hand, finding that OTS ability is more influential for children of higher ANS ability might suggest that OTS' contribution, whether it be precision or otherwise, is *amplificatory* in nature; that is, OTS becomes increasingly important for children whose ANS ability is high.

Two models – one that tests the complementary possibility and another that tests the compensatory vs. amplificatory possibilities – were fitted to the data at both time points in order to adjudicate between these various possibilities. The first model is a basic logistic regression with the ANS and OTS variables predicting whether a child is a CP-knower or non-CP-knower. Consistent with the results above, both ANS and OTS were found to significantly predict CP-knower status at T1 (p < .001 for ANS, p < .042 for OTS), whereas only ANS was significant at T2 (p < .001). As can be seen in the upper-left panels of Figure 2 (T1) and Figure 3 (T2), this model assumes a relatively constant effect of OTS across different levels of ANS, and therefore suggests a *complementary* role for OTS.

The second model is a novel multiplicative model that is sensitive to the nature of the interaction between the ANS and OTS variables, and can reveal whether it is *compensatory* or *amplificatory* by quantifying the relative contributions of the two systems. Formally, this model is stated as

$$p_i = [a_i]^{\alpha}[t_i]^{\beta}$$

where p_i is a number between 0 and 1 that represents the probability that the ith child will be a CP-knower, a_i is the ith child's accuracy (i.e., proportion correct) on the ANS task, t_i is the ith child's accuracy (i.e., proportion correct) on the OTS task, and α and β are weights which determine the influence of ANS and OTS, respectively, on the probability of the child being a CP-knower. A state of no influence is represented by a zero value for these weights, and greater positive values indicate greater influence. Details about the model and its analysis are provided in the Appendix.

We analyzed the multiplicative model in the Bayesian framework. Outputs are the posterior distributions of the weights α and β at T1 and T2. These distributions, shown in Figure 4, provide the plausible values of the weights. As can be seen, at T1 much of the posterior mass is away from zero for both α and β, providing evidence for the influence of both ANS and OTS. Moreover, there is a larger influence of ANS than OTS as evidenced by the larger values

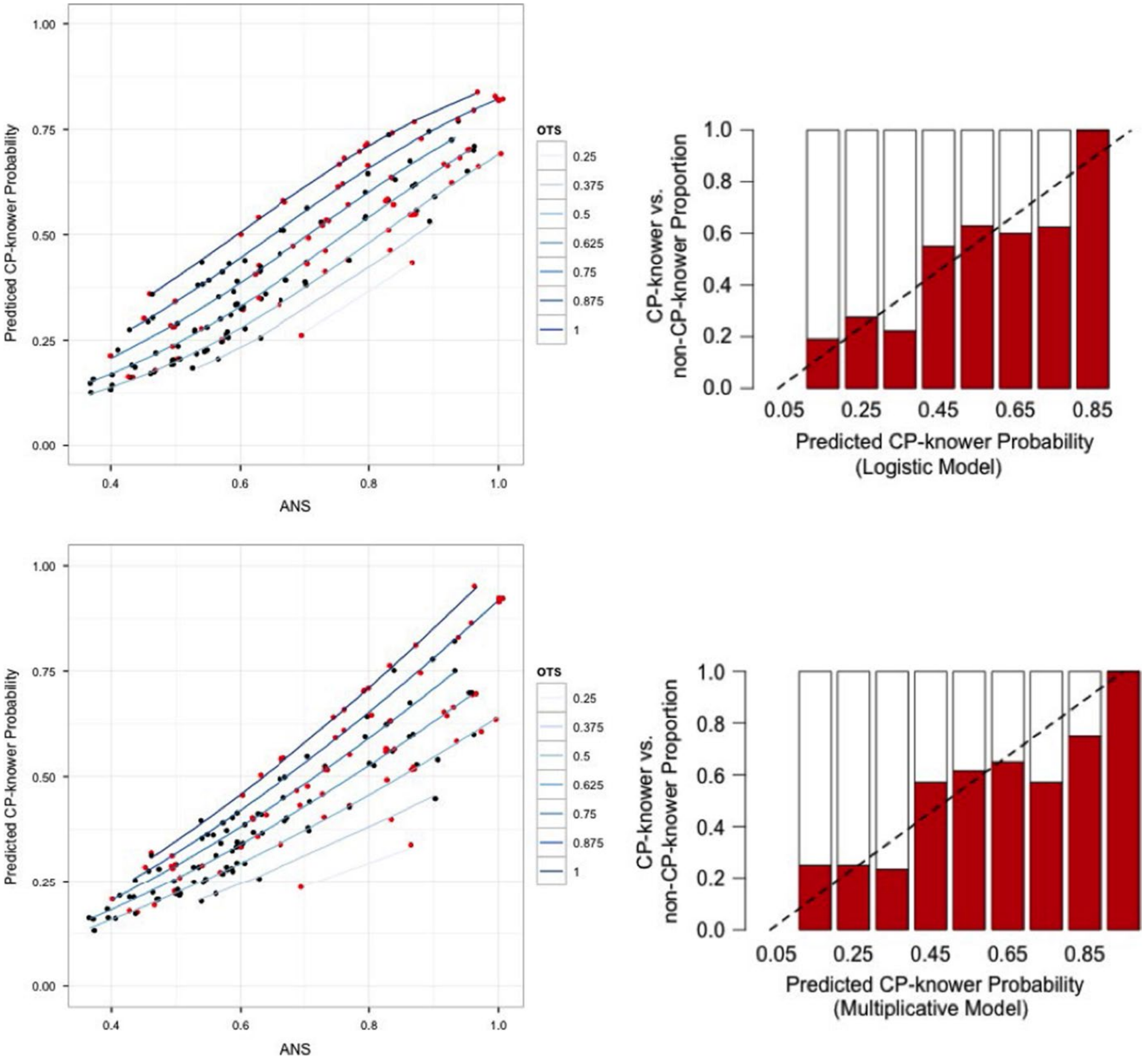

FIGURE 2 On the left, the y-axis shows the probability of being a CP-knower as predicted by the logistic model (upper panel) and the multiplicative model (lower-left panel) plotted against ANS performance for different levels of OTS at T1 on the x-axis (red points indicate CP-knowers and black points indicate non-CP-knowers). On the right, proportion of CP-knowers (red portion) vs. non-CP-knowers (white portion) for every .1 interval of CP-knower probability predicted by the logistic model (upper-right panel) and the multiplicative model (lower-right panel) at T1. Dotted lines represent ideal predictions

of α than β. The general pattern is similar at T2, except that the influence of OTS is substantially reduced. The posterior mean of β, the best point estimate, is reduced to .36 at T2 from .64 at T1. This is a 44% reduction in influence, and it is the size of this reduction that supports the argument of the waning role of OTS in cardinal knowledge.

The logistic regression model and the novel multiplicative models are not the same and may make somewhat different predictions about the patterns of data. Therefore, relative model fit was assessed by comparing the expected percent correctly predicted (ePCP;[7] Herron, 1999) by the multiplicative model to that of the logistic regression model. Similar ePCP values for the two models suggest that the multiplicative and the logistic regression model fit the data equally well at T1 (ePCP = .58 for the multiplicative model, ePCP = .58 for the logistic model) and also at T2 (ePCP = .59 for the multiplicative model, ePCP = .58 for the logistic model). Moreover, although the models are not the same, they made very similar predictions about the proportion of CP-knowers versus non-CP-knowers (upper- and lower-right panels of Figures 2 and 3). The multiplicative model may be viewed as more parsimonious, however, since it

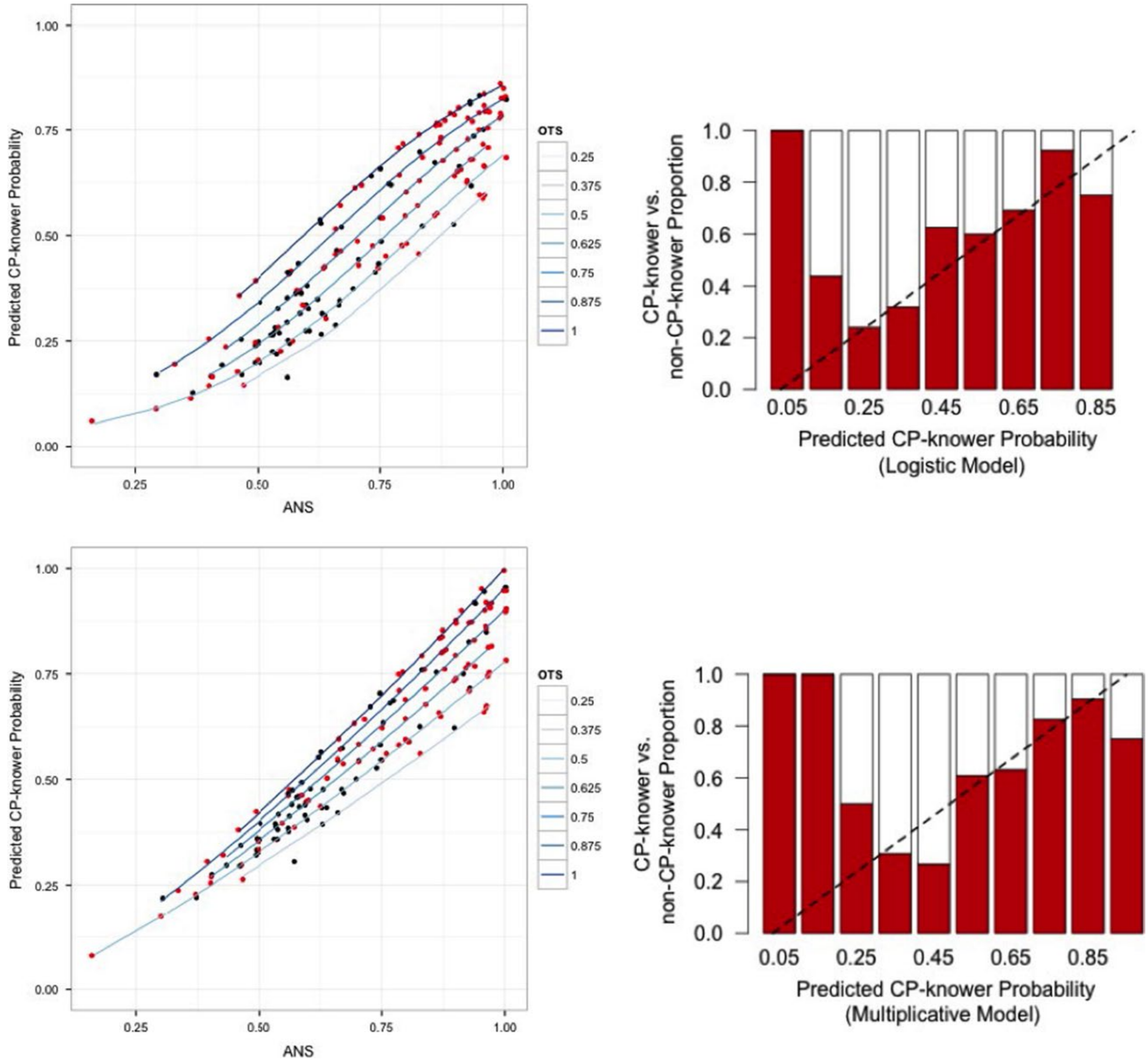

FIGURE 3 On the left, the y-axis shows the probability of being a CP-knower as predicted by the logistic model (upper panel) and the multiplicative model (lower-left panel) plotted against ANS performance for different levels of OTS at T2 on the x-axis (red points indicate CP-knowers and black points indicate non-CP-knowers). On the right, proportion of CP-knowers (red portion) vs. non-CP-knowers (white portion) for every .1 interval of CP-knower probability predicted by the logistic model (upper-right panel) and the multiplicative model (lower-right panel) at T2. Dotted lines represent ideal predictions. Note that the poor proportional fit for the first two CP-knower probability intervals (between 0 and .2) are due to two children who were CP-knowers despite extremely poor ANS and OTS performance

achieves comparable fit with one less parameter (i.e., no intercept term). Note that the poor proportional fit for the first two CP-knower probability intervals (between 0 and .2) are due to two children who were CP-knowers despite extremely poor ANS and OTS performance.

Visual inspection of the multiplicative model at T1 (lower-left panel of Figure 2) allows a model-based determination of the role of OTS. The rightward flare of the fan-shaped plot indicates, according to the model, that the OTS is more influential among children with higher ANS ability; this is consistent with the OTS playing an *amplificatory* role in children's numerical representations (a leftward flare would be indicative of a *compensatory* role of the OTS). As can be seen, the relation of ANS to CP-knower probability plotted across the different levels of OTS at T1 makes clear the dual contribution of both ANS and OTS at the earlier time point. The change in slopes across OTS performance levels suggests that ANS performance is increasingly related to CP-knower status as performance in OTS increases. In contrast, at T2 (lower-left panel of Figure 3), there is relatively little change and tighter grouping in the slopes for different OTS levels, which suggests that the influence of OTS on the relation between ANS and CP-knower status has waned. These results confirm the regression analyses in the preceding section and lend further support for the Merge model of cardinal knowledge acquisition.

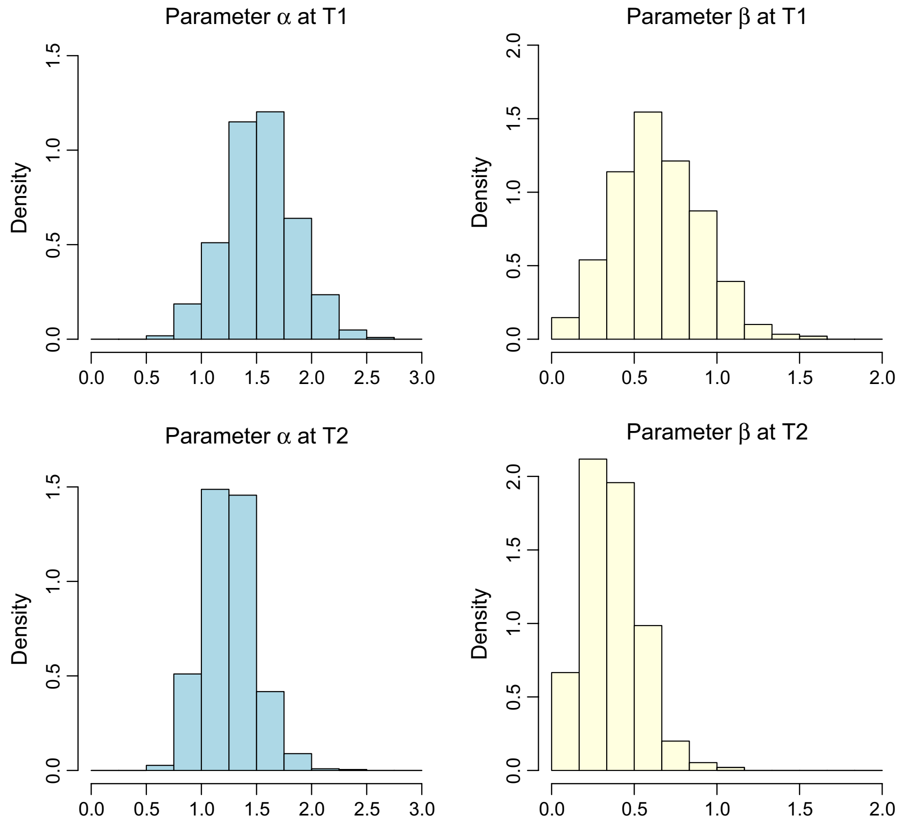

FIGURE 4 Posterior distribuions of the parameters *α* and *β* at T1 (top row) and T2 (bottom row), where *α* and *β* denote the influence of ANS and OTS, respectively, on cardinal knowledge. As can be seen, the distribution for *α* is most dense around 1.5 at T1 and remains mostly above .5 at T2. In contrast, although the T1 distribution for *β* is centered between 1 and 1.5, it shifts strongly to the left at T2, indicating the reduction in its influence

3.4 | Do gains in cardinal knowledge predict gains in ANS and OTS accuracy?

Mussolin, Nys, Content, and Leybaert (2014) recently reported that preschoolers' initial performance on the GiveN task predicted later ANS accuracy, but not the reverse, such that initial performance on the ANS task did not predict later GiveN performance. Based on this, they suggested that gains in cardinal knowledge lead to improved ANS accuracy, rather than the other way round as we have suggested. Because the relationship we find between ANS acuity and change in CP-knower status is essentially correlational, our earlier regression analyses do not necessarily rule out the possibility that change in CP-knower status may in fact be driving improvements in ANS acuity, and not OTS accuracy. To explore this alternative, we first computed cross-lagged partial correlations (following Mussolin et al., 2014, also Williams, 1959, and Steiger, 1980), controlling for sex, race, parental education, age at T1, IQ, EF, preliteracy, and the relevant autoregressor. Neither T1 ANS accuracy ($pr_{125} = -.01$, $p > .05$, two-tailed) nor T1 GiveN performance ($pr_{114} = .165$, $p > .05$, two-tailed) was significantly related to the other variable at T2, and the partial correlations themselves did not differ ($t(124) = 1.80$, $p > .05$, two-tailed). Although our partial correlations were in the same direction as reported by Mussulin et al. (a larger coefficient for cardinality predicting ANS accuracy than the reverse), our larger sample (twice the size of theirs) should have afforded us ample power to detect such a relationship if it was there, suggesting that gains in cardinal knowledge were not related to improvements in ANS accuracy over time in our dataset.

Nonetheless, we further expolored this alternative by conducting two additional regression analyses predicting change in accuracy from T1 to T2 separately for the ANS and OTS measures. In the first analysis, CPchange (whether or not a child moved from CP-non-knower at T1 to CP-knower at T2) was entered along with sex, race, parental education, age at T1, IQ, EF, preliteracy, and the autoregressor (T1 ANS accuracy) to predict change in ANS accuracy from T1 to T2 (T2 ANS accuracy − T1 ANS accuracy); CPchange was not significant ($\beta = .257$, $p = .324$). The second analysis was identical, except with CPchange and the other variables predicting change in OTS accuracy from T1 to T2 (T2 OTS accuracy − T1 OTS accuracy). Again though, becoming a CP-knower between T1 and T2 did not predict change in OTS accuracy ($\beta = .034$, $p = .901$). These results provide further support that in our sample, changes in ANS acuity drove improvements in cardinal knowledge, and not the other way round.

4 | DISCUSSION

Children's surprisingly slow and stage-like progression through the knower-levels using the GiveN task (Wynn, 1992a) sparked a debate that remains lively despite intense research efforts over the last two decades. The initial impact of Wynn's findings was partly due to the fact that they did not accord with the then dominant ANS-only view (Gallistel & Gelman, 1992). But even without Wynn's results, the ANS-only view is not a complete explanation of how the verbal count list acquires meaning; the ANS embodies the necessary principles, but the representations are imprecise. Carey's (2004) and Le Corre and

Carey's (2007) more recent OTS-only view provides a better account for children's progression in the GiveN task; however, there are other reasons to question whether the OTS alone is sufficient for children to fix the meaning of the number words (see Gallistel, 2007 and Leslie et al., 2007, 2008). In particular, the lack of cardinal representations means that the OTS cannot provide semantic meaning to the count list, even for sets that fall within its capacity (cf. Le Corre & Carey, 2008). Indeed, Carey (2009) explicitly points out that because the OTS lacks cardinal representations, it must be augmented by other mechanisms, namely the set-based notions that underlie the singular/plural distinction in natural language. Given these limitations, we were not surprised that our data did not support either of the single-mechanism views.

The present study is not only the first to directly test the ANS-only and OTS-only models in the same dataset, but it also provides the first empirical evidence for a dual-mechanism approach to understanding how children learn to count. More importantly, we were able to determine which of the two dual-mechanism views best described change in cardinal knowledge over time. The developmental pattern we observed favored the Merge model over the Combine model. Both systems were related to CP-knower status at T1, above and beyond the influence of domain-general cognitive abilities and demographic factors; but as predicted by the Merge model, only the ANS remained a significant predictor at T2. And critically, the ANS, but not the OTS, reliably predicted children's developmental transition to CP-knower. It is notable that performance on the ANS and OTS tasks was not correlated at either time point. This may reflect trivial differences in the task demands, or a more fundamental difference in the functional roles played by the two systems. Hyde and Mou (2015), for example, argue in a recent review that the extant evidence suggests that the ANS and OTS, despite both being core mechanisms, have distinct neural and behavioral signatures. If so, then it is perhaps less surprising that we find children's performance on the two tasks to be unrelated.

If the Merge model is correct, then how can we account for Le Corre and Carey's (2007) two main findings suggesting that (1) children are not immediately able to estimate the magnitude of sets beyond 'four' upon becoming CP-knowers, and (2) that estimation errors for set sizes within the capacity of the OTS ('one' to 'four') do not show the typical variability signature of the ANS? With respect to the first finding, we do not consider it at all surprising that it takes children time (~6 months, according to their data) to map the labels in the 'six' to 'ten' range to the ANS. The Merge model claims that children rely on both the ANS and OTS for learning the meanings of the first few number words ('one' to 'four'); but after that point, they shift to relying primarily on the ANS as they continue to learn the meanings of the words beyond 'four'. Our model does not claim that children should suddenly and abruptly map words beyond 'four' to ANS. Indeed, these mappings are likely formed through counting experience (seeing large sets labeled, or counting large sets themselves). Once children become CP-knowers, they can successfully implement the counting procedure, and thus have the means to form mappings for larger sets. But it is not the case that simply making the transition to CP-knower somehow instantaneously gives the child the mappings beyond 'four'.

With respect to Le Corre and Carey's (2007) second finding, the Merge model does not claim that children do not map the first few words to OTS representations. On the contrary, they likely map them to both systems. Thus, finding that variation around estimates in the 'one' to 'four' range is not scalar does not argue against our model and in fact is consistent with our model. Transitioning to CP-knower does not mean one must abruptly stop using the OTS for small sets. The transition may be gradual as the precision of the ANS increases.

Another important contribution of our study is that the novel multiplicative analysis allowed us to characterize (and quantify) the role of the OTS as it changed from T1 to T2, and in relation to the ANS, providing a new perspective for thinking about how the two systems might interact during this crucial period of development. Although the early contribution of the OTS could manifest in at least three different ways (complementary, compensatory, or amplificatory), we found evidence that it amplifies the impact of the ANS. At the first time point, the influence of the OTS was greater for children with higher ANS ability. This influence diminished by T2, however, such that OTS accuracy no longer had a large influence on the relationship between ANS ability and CP-knower status. Why the shift? What work is the OTS doing early on? One possibility is that the OTS may serve to calibrate or tune the ANS representations. Before the ANS can enumerate a set, the items within that set must be individuated. The common activation of the ANS with the OTS during learning situations may allow the child to capitalize on the precision of the OTS, such that children with high ANS accuracy benefit more from combined activation. Once precision is high enough, the influence of the OTS may then diminish, and the child can then fully realize the structural similarity between the ANS and the verbal count list. That is, the principles inherent in the ANS (stable order, cardinality, one-to-one correspondence) and their similarity to the verbal counting sequence/routine may be made more salient once the OTS becomes less influential. Although the ANS may take on a more dominant role in children's acquisition of cardinal knowledge, the OTS continues to influence other aspects of quantitative reasoning into adulthood (e.g., Balakrishnan & Ashby, 1992; Choo & Franconeri, 2012; Gallistel & Gelman, 2005; Revkin et al., 2008; Trick & Pylyshyn, 1994).

One question raised by these findings is what drives the merging of the two systems? One possibility alluded to earlier is that the OTS might simply reach its capacity limit and become unable to represent larger quantities. For quantities beyond this limit, the ANS would be dominant because it simply would not be in competition with the OTS. Alternatively, the developmental increase in the precision of ANS representations may result in the system reaching some critical threshold of acuity (e.g., being able to reliably discriminate numbers at a 3:4 ratio, Weber fraction = .33) that allows it to be preferred over the OTS in situations requiring cardinal reasoning. Consistent with this possibility, 40% of CP-knowers (28% at T1 and 51% at T2) had a Weber fraction of .33 or better (i.e., smaller), as compared to only 17% of non-CP-knowers (12% at T1 and 21% at T2).

These two possibilities have different implications for early interventions for children at risk for poor long-term mathematics achievement. Although its relative importance is currently debated (De Smedt, Noel, Gilmore & Ansari, 2013), better ANS acuity is associated with higher formal mathematics achievement (Chu et al., 2013; for review see Feigenson, Libertus, & Halberda, 2013), and training adults and children on an ANS task can increase their ANS precision and improve their performance on symbolic arithmetic tasks (Hyde, Khanum, & Spelke, 2014; Park & Brannon, 2013). If the merge is caused by the improvement in ANS precision, then interventions that accelerate developmental change in ANS acuity may result in an earlier transition to CP-knower status that in turn may contribute positively to other aspects of formal mathematical learning (vanMarle et al., 2014). On the other hand, if the merge is primarily the result of reaching the capacity limit of the OTS, then such training may not be an effective intervention.

Our data also speak to other proposals concerning the nature of the ANS–cardinality relationship. Sullivan and Barner (2014), for example, suggest that 5- to 7-year-old children use distinct mapping mechanisms to fix the meaning of number words in the small (up to about 'six') and large number ranges (greater than 'six'). In addition, Gunderson, Spaepen, and Levine (2015) suggest that approximate number word knowledge may not be related to cardinal knowledge, since children who have not yet become CP-knowers have approximate meanings for number words up to 'ten' that are beyond their knower-level. And perhaps even more relevant to the present findings, Mussolin et al. (2014) have suggested that contrary to the present findings, it may be that gains in cardinal knowledge predict increases in ANS accuracy, rather than the reverse as we claim. Although we are unable to address the first two alternatives with our dataset, our failure to find a significant cross-lag correlation for T1 cardinality and T2 ANS accuracy, as well as our regression results showing that becoming a CP-knower is not predictive of gains in ANS accuracy, appear to rule out this last alternative.

In any case, the present study suggests that learning the meanings of the count words, a critical first step in children's development of formal mathematical knowledge, is more nuanced than previously thought. Previous proposals positing that a single mechanism, either the ANS or the OTS, can explain the development of children's cardinal knowledge cannot account for our data. Contrary to the Combine model (Spelke, 2011; Spelke & Tsivkin, 2001a) which predicts the continued influence of both systems on children's cardinal reasoning, we found evidence for the Merge model, in which the ANS and OTS both initially support cardinal knowledge acquisition, but the role of OTS diminishes after the first few words are mapped, allowing children to recognize the corresponding structure in the verbal count list and the ANS, and thus fully grasp an understanding of counting. These findings therefore suggest that children may not undergo genuine novel conceptual change when learning to count. Instead, they build their understanding on two core mechanisms, but come to rely primarily on the ANS, which parallels the verbal counting system in structure and process. Nevertheless, children clearly have the explicit insight that the meaning of successive number words is based on the successor and cardinality principles and this is a conceptual change from a comparative perspective (i.e., relative to other primates) (Beran, Parrish, and Evans, 2015). Our suggestion is that this conceptual change emerges from the explicit recognition of the basic properties inherent in the organization of the ANS. Further research is necessary to better understand the timecourse of this process, and how learning the meaning of the count words sets the stage for the development of formal mathematical knowledge once children enter school.

ACKNOWLEDGEMENTS

This work was supported by grants from the University of Missouri Research Board and DRL-1250359 from the National Science Foundation. We would like to thank the Columbia Public School System, and especially Mary Rook for help facilitating our access to subjects and testing facilities. We are grateful to the teachers, parents, and children who participated in the study. We also thank Mary Hoard and Lara Nugent and the members of the Cognitive Development Lab for their help with data collection and coding, and three anonymous reviewers for helpful comments on earlier versions of this manuscript. The data reported here come from an ongoing longitudinal study and were reported in part in papers presented at the Evolutionary Origins and Early Development of Basic Number Processing conference in Bethesda, MD in May 2013 and the 2015 meeting of the Society for Research in Child Development in Philadelphia, PA.

NOTES

[1] Set-based natural language quantifiers (e.g., singular, dual, and trial markers) also play a role, but because the majority of the conceptual work is carried by the OTS, we will not discuss in detail the contribution of the set-based quantification system, which is described in Carey (2004) and Le Corre and Carey (2007). In fact, recent evidence suggests that natural language quantifiers may help children fix the meanings of the first few number words. Almoammer et al. (2013), for example, showed that children who speak dual-marking languages (Slovenian and Saudi Arabic) learn faster than children who speak singular-marking languages (English), and the delay in learning the meaning of the first unmarked number word occurs at 'two' for English speakers, but 'three' for Slovenian and Arabic speakers (see also Sarnecka, 2014, for review). More generally, having access to a spoken or signed verbal count list may be critical for the development of an exact understanding of sets beyond about 4. Both deaf homesigners who are not exposed to count words in their culture (Spaepen, Coppola, Spelke, Carey, & Goldin-Meadow, 2011) and indigenous peoples whose language does not have specific words for sets beyond about three (Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004) appear to lack exact representations for quantities above four. However, others like Gelman and Butterworth (2005) have argued that although language may affect our numerical cognition, it does not play a causal role in our learning the meanings of the number words. This debate is beyond the scope of this paper, as our sample was primarily English-speaking with a few Spanish-speaking children. We would note, however, that the notion that quantifiers may help children is not incompatible with the Merge model. The mappings between the verbal labels and the ANS still have to happen, and if a language has multiple ways of highlighting sets, then of course, this might facilitate the mappings.

[2] Leslie and colleagues (Leslie et al., 2007; Leslie, Gelman, & Gallistel, 2008) propose a calibration process as a way to transform fuzzy analog magnitudes from the ANS into exact integer representations. They entertain the idea that the verbal count list itself may calibrate the ANS, but dismiss it because it cannot explain how the child comes by the notion of 'exact equality' between different instances of the same integer. The OTS may

provide this link because, at least for small sets, working memory models of two different instances of a set can be put in one-to-one correspondence, which may be considered an instantiation of 'exact equality'. This is a theoretical conjecture, of course, and is in need of empirical confirmation. However, it is plausible and more parsimonious than Leslie et al.'s proposal to build in yet another innate concept (in addition to the ANS and OTS) that embodies the notions of 'exactly one' and 'exact equality'. The OTS may already possess these qualities, but they may only become numerically meaningful once the ANS and OTS become linked via the verbal count list.

[3] Estimation of a stable ω requires a level of accuracy that was not achieved by all children, resulting in developmentally plausible w estimates for only 145 and 149 children at T1 and T2, respectively. In contrast, percent correct was usable for all children.

[4] Even though performance in our sample did not differ depending on whether total area was controlled or left to covary with numerosity, recent studies with adults and children suggest that controlling for non-numerical cues can affect performance (e.g., Clayton, Gilmore, & Inglis, 2015; Szucs, Nobes, Devine, Gabriel, & Gebuis, 2013), raising the possibility that performance on tasks where such controls are employed may reflect more than just ANS acuity, but also visual processing capacities and general cognitive abilities (e.g., inhibition).

[5] We thank an anonymous reviewer for noting that the pattern seen for the OTS (as a significant predictor of CP-knower status at T1, but not T2) could be an artifact of our CP-knowers (but not our non-CP-knowers) being able to encode the hidden sets with verbal number labels ('one', 'two'), effectively reducing the maintenance demands on their OTS. If true, this would lead to a stronger relationship between OTS and CP-knower status at T1 than at T2,when presumably most children can utilize this strategy. In fact, the children in our sample completed an enumeration task as part of their test battery in which they had to verbally count a set of stickers (20 total) affixed to a piece of foamcore board. In our entire sample, only 21 children (20 non-CP-knowers) at T1 and 4 children (all non-CP-knowers) at T2 were unable to enumerate up to at least 'two', which would be the maximum number required to encode the sets in this task. Removing these children from the sample did not change pattern of results. At T1, both ANS ($\beta = .744$, $p = .004$) and OTS ($\beta = .502$, $p = .026$) were significant, along with EF, IQ, and preliteracy. At T2, ANS remained significant ($\beta = .613$, $p = .03$), but OTS did not ($\beta = .284$, $p = .148$). Thus, although it was true that CP-knowers would have been more able to use that strategy, it cannot account for the finding that OTS was a significant predictor of CP-knower status only at T1.

[6] We thank an anonymous reviewer for bringing this alternative explanation to our attention.

[7] Herron (1999) defines ePCP as: $ePCP = \frac{1}{N}[\sum_{y_i=0} \hat{p}_i + \sum_{y_i=1} (1-\hat{p}_i)]$. ePCP is used here to measure fit instead of other measures (e.g., PCP or PRE) because it distinguishes between small and large values of $\hat{p}$ such as .91 and .51 (whereas PCP would treat the two predictions as the same since both are greater than .5). This is the appropriate measure of fit for the multiplicative model since it produces a number between 0 and 1 as its prediction, with greater numbers indicating a higher probability that the child will be a CP-knower.

REFERENCES

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

Balakrishnan, J.D., & Ashby, F.G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research*, *54*, 80–90.

Beck, D.M., Schaefer, C., Pang, K., & Carlson, S.M. (2011). Executive function in preschool children: Test–retest reliability. *Journal of Cognition and Development*, *12*, 169–193.

Beran, M.J., Parrish, A.E., & Evans, T.A. (2015). Numerical cognition and quantitative abilities in nonhuman primates. In D.C. Geary, D.B. Berch & K. Mann Koepke (Eds.), *Evolutionary origins and early development of number processing, Vol. 1, Mathematical cognition and learning* (pp. 91–119). San Diego, CA: Elsevier Academic Press.

Bull, R., Espy, K.A., & Wiebe, S.A. (2008). Short-term memory, working memory, and executive functions in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*, 205–228.

Burr, D.C., Turi, M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision*, *10* (6), 20.

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

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

Cheries, E.W., Wynn, K., & Scholl, B.J. (2006). Interrupting infants' persisting object representations: An object-based limit? *Developmental Science*, *9* (5), F50–F58.

Choo, H., & Franconeri, S.L. (2012). Enumeration of small collections violates Weber's law. *Psychonomic Bulletin & Review*, *21*, 93–99.

Chu, F., vanMarle, K., & Geary, D.C. (2013). Quantitative deficits of preschool children at risk for mathematical learning disability. *Frontiers in Developmental Psychology*, *4*, 195.

Clark, C.A., Sheffield, T.D., Wiebe, S.A., & Espy, K.A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development*, *84*, 662–677.

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica*, *161*, 177–184.

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*, 698–707.

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

Dehaene, S., Spelke, E.S., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974.

De Smedt, B., Noel, M.P., Gilmore, C., & Ansari, D. (2013). The relationship between symbolic and non-symbolic numerical magnitude processing and the typical and atypical development of mathematics: A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*, 48–55.

Drew, T., & Vogel, E.K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *Journal of Neuroscience*, *28*, 4183–4191.

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

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, *97*, 295–313.

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

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

Feigenson, L., Libertus, M.E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, *7*, 74–79.

Gallistel, C.R. (2007). Commentary on Le Corre & Carey. *Cognition*, *105*, 439–445.

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

Gallistel, C.R., & Gelman, R. (2005). Mathematical cognition. In K. Holyoak & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 559–588). New York: Cambridge University Press.

Geary, D.C. (2011). Cognitive predictors of individual differences in achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*, 1539–1552.
Gelman, R. (1972). The nature and development of early number concepts. In H.W. Reese (Ed.), *Advances in child development and behavior* (Vol. 7, pp. 115–167). New York: Academic Press.
Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. In D. Medin (Ed.), *Learning and motivation* (Vol. 30, pp. 61–96). New York: Academic Press.
Gelman, R. (2006). Young natural-number arithmeticians. *Current Directions in Psychological Science*, *15*, 193–197.
Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences*, *9*, 6–10.
Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2004). *Bayesian data analysis* (2nd edn.). London: Chapman and Hall.
Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*, 496–499.
Gunderson, E.A., Spaepen, E., & Levine, S.C. (2015). Approximate number word knowledge before the cardinal principle. *Journal of Experimental Child Psychology*, *130*, 35–55.
Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the 'number sense': The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*, 1457–1465.
Herron, M.C. (1999). Postestimation uncertainty in limited dependent variable models. *Politital Analysis*, *8*, 83–98.
Hyde, D.C., Khanum, S., & Spelke, E.S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*, 92–107.
Hyde, D.C., & Mou, Y. (2015). Neural and behavioral signatures of core numerical abilities and early numerical development. In D.B. Berch, D.C. Geary & K. Mann Koepke (Eds.), *Mathematical cognition and learning* (Vol. 2, pp. 51–77). San Diego, CA: Elsevier.
Hyde, D.C., & Spelke, E.S. (2009). All numbers are not equal: An electrophysiological investigation of large and small number representations. *Journal of Cognitive Neuroscience*, *21*, 1039–1053.
Hyde, D.C., & Spelke, E.S. (2010). Neural signatures of number processing in human infants: Evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*, 360–371.
Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147–155.
Invernizzi, M., Sullivan, A., Meier, J., & Swank, L. (2004). *PreK teacher's manual: Phonological awareness and literacy screening*. Charlottesville, VA: University of Virginia.
Izard, V., Sann, C., Spelke, E.S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy Sciences, USA*, *106*, 10382–10385.
Kahneman, D., Treisman, A., & Gibbs, B.J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 174–219.
Kibbe, M.M., & Leslie, A.M. (2013). What's the object of object working memory in infancy? Unraveling 'what' and 'how many'. *Cognitive Psychology*, *66*, 380–404.
Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*, 395–438.
Le Corre, M., & Carey, S. (2008). Why the verbal counting principles are constructed out of representations of small sets of individuals: A reply to Gallistel. *Cognition*, *107*, 650–662.
LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B.L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, *81*, 1753–1767.
Leslie, A.M., Gallistel, C.R., & Gelman, R. (2007). Where integers come from. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The innate mind, Vol. 3: Foundations and the future* (pp. 109–138). New York: Oxford University Press.
Leslie, A.M., Gelman, R., & Gallistel, C.R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences*, *12*, 213–218.
Leslie, A.M., Xu, F., Tremoulet, P., & Scholl, B. (1998). Indexing and the object concept: Developing 'what' and 'where' systems. *Trends in Cognitive Sciences*, *2*, 10–18.
Lipton, J.S., & Spelke, E.S. (2003). Origins of number sense: Large number discrimination in human infants. *Psychological Science*, *14*, 396–401.
Mou, Y., & vanMarle, K. (2014). Two core systems of numerical representation in infants. *Developmental Review*, *34*, 1–25.
Mussolin, C., Nys, J., Content, A., & Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool children. *PLoS ONE*, *9*, e91839.
Negen, J., & Sarnecka, B.W. (2015). Is there really a link between exact-number knowledge and approximate number system acuity? *British Journal of Developmental Psychology*, *33*, 92–105.
Odic, D., Libertus, M.E., Feigenson, L., & Halberda, J. (2013). Developmental change in the acuity of approximate number and are representations. *Developmental Psychology*, *49*, 1103–1112.
Park, J., & Brannon, E.M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, *24*, 2013–2019.
Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*, 499–503.
Pylyshyn, Z.W., & Storm, R.W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.
Revkin, S.K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, *19*, 607–614.
Sarnecka, B.W. (2014). On the relation between grammatical number and cardinal numbers in development. *Frontiers in Psychology*, *5*, 1132.
Simon, T.J. (1997). Reconceptualizing the origins of number knowledge: A 'non-numerical' account. *Cognitive Development*, *12*, 349–372.
Spaepen, E., Coppola, M., Spelke, E.S., Carey, S.E., & Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences, USA*, *108*, 3163–3168.
Spelke, E.S. (2011). Natural number and natural geometry. In E. Brannon & S. Dehaene (Eds.), *Attention and performance Vol. 24. Space, time and number in the brain: Searching for the foundations of mathematical thought* (pp. 287–317). Oxford: Oxford University Press.
Spelke, E.S., & Tsivkin, S. (2001a). Initial knowledge and conceptual change: Space and number. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 70–100). Cambridge: Cambridge University Press.
Spelke, E.S., & Tsivkin, S. (2001b). Language and number: A bilingual training study. *Cognition*, *78*, 45–88.
Starkey, P. (1992). The early development of numerical reasoning. *Cognition*, *43*, 93–126.
Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.
Sullivan, J., & Barner, D. (2014). Inference and association in children's early numerical estimation. *Child Development*, *85*, 1740–1755.
Szucs, D., Nobes, A., Devine, A., Gabriel, F.C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise measurement of approximate number system acuity and comparison effects between adults and children. *Frontiers in Psychology*, *4*, 444.
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*, 80–102.
vanMarle, K. (2013). Different mechanisms underlie infants' representation of small and large numbers in an ordinal choice task. *Journal of Experimental Child Psychology*, *114*, 102–110.
vanMarle, K., Chu, F., Li, Y., & Geary, D.C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Developmental Science*, *17*, 492–505.

vanMarle, K., & Wynn, K. (2001, April). *Preschoolers' detection of magical events*. Poster presented at the biannual conference of the Society for Research in Child Development (SRCD) in Minneapolis, MN.
Vogel, E.K., McCollough, A.W., & Machizawa, M.G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503.
Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence-III*. San Antonio, TX: The Psychological Corporation, Harcourt Brace & Co.
Williams, E.J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B*, *21*, 396–399.
Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155–193.
Wynn, K. (1992a). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*, 220–251.
Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*, 749–750.
Zosh, J.M., & Feigenson, L. (2009). Beyond 'what' and 'how many': Capacity, complexity, and resolution of infants' object representations. In B. Hood & L. Santos (Eds.), *The origins of object knowledge* (pp. 25–51). New York: Oxford University Press.
Zur, O., & Gelman, R. (2004). Young children can add and subtract by predicting and checking. *Early Childhood Quarterly Review*, *19*, 121–137.

APPENDIX

ANALYSIS OF THE MULTIPLICATIVE MODEL

Let $i = 1, \ldots, I$ denote the child in the study, and let a_i, t_i denote the observed ANS and OTS accuracy scores, respectively. Note that a_i and t_i are bounded between 0 and 1; i.e., they are restricted to the unit interval. Let $c_i = 0, 1$ denote the observed cardinality-knower status of the ith child. We assume there is some probability p_i that the child has learned the cardinality property and that c_i is an outcome from a Bernoulli trial with probability p_i The goal is to understand the influences of ANS and OTS on p_i.

Note that p_i like a_i and t_i is restricted to the unit interval. The most natural operation to combine variables in this context is multiplication because the product of two variables on the unit interval remains on the unit interval. Thus, the unit interval is said to be closed under multiplication. Additionally, any variable on the unit interval can be transformed by a power function and the resultant remains on the unit interval. The unit interval is closed under power-function transformations. Hence a natural model that insures p_i remains on the unit interval is

$$p_i = [a_i]^\alpha [t_i]^\beta, \quad \alpha,\beta \geq 0$$

The exponents α and β serve as weights that denote importance of the respective variables. If one of these weights is 0, for example, the variable has no influence. The greater the weight, the greater the contribution of the variable (Figure A1 shows the influence of exponential parameter α on a hypothetical variable X for various values of α). Hence the goal in analysis is to estimate α and β and assess whether they are substantially different from 0. The multiplicative model is more natural and has fewer parameters than the probit-regression counterparts because in this model there is no need for an intercept.

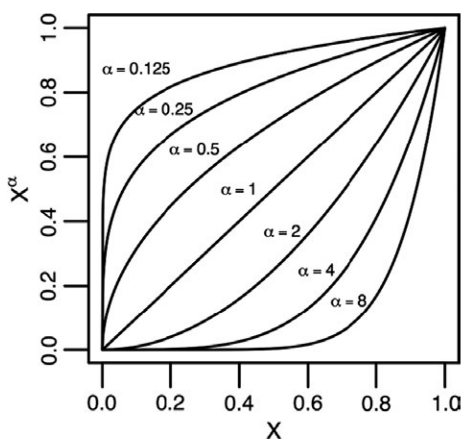

FIGURE A1. Influence of exponential parameter on a hypothetical variable X for various values of α.

The multiplicative model is conveniently analyzed in the Bayesian framework. Prior distributions are needed for α and β. The exponential is a seemingly suitable choice as it places mass on only positve values, favors smaller values, and has a thin tail:

$$\alpha \sim \mathrm{Exp}(\lambda_\alpha)$$

and

$$\beta \sim \mathrm{Exp}(\lambda_\beta)$$

Prior settings on rates λ_α and λ_β are chosen before analysis. We explored the predictions of the model under a number of choices and found that $\lambda_\alpha = \lambda_\beta = .5$ resulted in a wide range of plausible predictions. We discuss robustness checks to this choice subsequently.

The likelihood function for the model is

$$L(\alpha,\beta;a,t,c) = \Pi_i[(a_i^\alpha t_i^\beta)^{c_i}(1-a_i^\alpha t_i^\beta)^{1-c_i}]$$

and the joint posterior distribution over the two parameters is

$$f(\alpha,\beta|a,t,c) \propto \Pi[(a_i^\alpha t_i^\beta)^{c_i}(1-a_i^\alpha t_i^\beta)^{1-c_i}]\lambda_\alpha e^{-\lambda_\alpha}\lambda_\beta e^{-\lambda_\beta}$$

This joint posterior does not follow any form known to us, and so we obtained marginal posterior distributions of the parameters using Markov chain Monte Carlo sampling with Metropolis-Hastings steps (Gelman, Carlin, Stern, & Rubin, 2004). The Metropolis candidate was tuned for acceptance rates between .40 and .45 for all parameters and all runs. Sampling continued for 10,000 iterations with the first 1000 iterations discarded as a burn-in period. The resulting chains showed a modest amount of autocorrelations and were thinned by a factor of 10. Figure A2 shows the outputs for parameter α at T1 after thinning (the outputs for β at T1 as well as α and β at T2 were similar). The relatively small degree of autocorrelation is further evidenced by the autocorrelation plot (Figure A3).

Posterior distributions for α and β at T1 and T2 are provided in Figure 3 in the text. To assess the robustness of the analysis to the

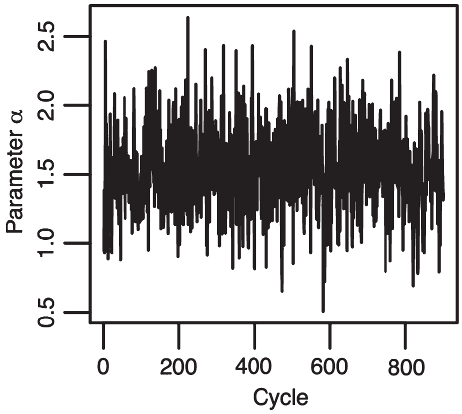

FIGURE A2. Chain output for parameter α after thinning by a factor of 10.

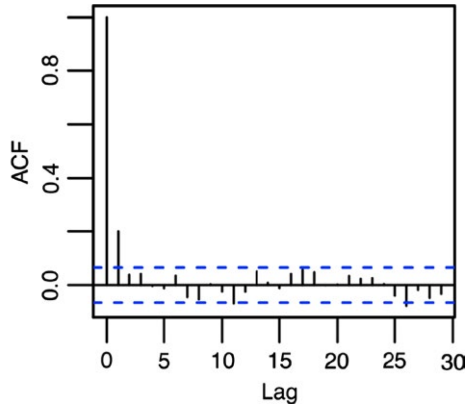

FIGURE A3. Autocorrelation plot for parameter α after thinning by a factor of 10.

choice of prior settings, we reran it with different settings. The rates were increased by a factor of four ($\lambda_\alpha = \lambda_\beta = 2$) and decreased by a factor of four ($\lambda_\alpha = \lambda_\beta = .125$). The effect of this combined factor of 16 was minimal – the posterior means of α and β varied by about 5%. Hence, the findings are not unduly sensitive to prior settings.