Title
Why is number word learning hard? Evidence from bilingual learners

Authors
Wagner, Katie
Kimura, Katherine
Cheung, Pierina
et al.

Publication Date
2015-09-27

DOI
10.1016/j.cogpsych.2015.08.006

Peer reviewed

Cognitive Psychology 83 (2015) 1–21

Contents lists available at ScienceDirect

Cognitive Psychology

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

Why is number word learning hard? Evidence from bilingual learners

Katie Wagner [a,*,1], Katherine Kimura [a,b,1], Pierina Cheung [a,c], David Barner [a,d]

[a] Department of Psychology, University of California, San Diego, United States
[b] Department of Psychology, University of California, Berkeley, United States
[c] Department of Psychology, University of Waterloo, Canada
[d] Department of Linguistics, University of California, San Diego, United States

ARTICLE INFO

Article history:
Accepted 24 August 2015

Keywords:
Number word learning
Counting
Cognitive development
Language acquisition
Bilingualism

ABSTRACT

Young children typically take between 18 months and 2 years to learn the meanings of number words. In the present study, we investigated this developmental trajectory in bilingual preschoolers to examine the relative contributions of two factors in number word learning: (1) the construction of numerical concepts, and (2) the mapping of language specific words onto these concepts. We found that children learn the meanings of small number words (i.e., *one*, *two*, and *three*) independently in each language, indicating that observed delays in learning these words are attributable to difficulties in mapping words to concepts. In contrast, children generally learned to accurately count larger sets (i.e., *five* or greater) simultaneously in their two languages, suggesting that the difficulty in learning to count is not tied to a specific language. We also replicated previous studies that found that children learn the counting procedure before they learn its logic – i.e., that for any natural number, n, the successor of n in the count list denotes the cardinality $n + 1$. Consistent with past studies, we found that children's knowledge of successors is first acquired incrementally. In bilinguals, we found that this knowledge exhibits item-specific transfer between languages, suggesting that the logic of the positive integers may not be stored in a language-specific format. We conclude that delays in learning the meanings of small number words are mainly due to language-specific processes of mapping

* Corresponding author at: Department of Psychology, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0109, United States.
E-mail address: kgwagner@ucsd.edu (K. Wagner).

[1] The first two authors contributed equally to this work.

http://dx.doi.org/10.1016/j.cogpsych.2015.08.006

words to concepts, whereas the logic and procedures of counting appear to be learned in a format that is independent of a particular language and thus transfers rapidly from one language to the other in development.

1. Introduction

Number words like *one*, *two*, and *three*, are learned by young children throughout the world. However, number words are neither universal (Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004) nor easy to acquire. Even among highly trained children in countries such as the U.S., Japan, and China, it takes years to acquire the meanings of the first three or four number words, with learning proceeding in distinct and highly protracted stages that are separated by many months (Barner, Libenson, Cheung, & Takasaki, 2009; Le Corre, Van de Walle, Brannon, & Carey, 2006; Li, Le Corre, Shui, Jia, & Carey, 2003; Sarnecka, Kamenskaya, Yamana, Ogura, & Yudovina, 2007; Wynn, 1990, 1992). For example, after English-speaking children in the U.S. learn the meaning of the word *one*, some 6–9 months pass until they acquire a meaning for *two*, and then several more months elapse before they learn the meaning of *three* (Wynn, 1990, 1992). In the present study, we examined two factors – which are not mutually exclusive – that might contribute to these delays between early stages of number word learning: (1) the construction of numerical concepts like "exactly one" and "exactly two", and (2) the language-specific problem of mapping words like *one* and *two* onto these concepts. To examine the contribution of these two factors, we investigated number word learning in bilingual preschoolers, who solve the linguistic mapping problem twice, once for each of their two languages. We reasoned that if the delays between number word stages are primarily due to the problem of constructing new concepts, then once bilinguals acquire number knowledge in one language they should acquire the same knowledge more easily in their second language – i.e., with little additional time required for language-specific mapping of words in their second language. However, if delays between stages are primarily due to the problem of mapping words onto concepts within a particular language, then prior acquisition of number words in a first language should have little effect on children's learning of corresponding words in a second language.

Across a variety of languages and cultural groups, including Canada, the US, Japan, Russia, Bolivia, Slovenia, Taiwan, and Saudi Arabia, children learn number word meanings in distinct, protracted, stages (Almoammer et al., 2013; Barner, Libenson, et al., 2009; Li et al., 2003; Piantadosi, Jara-Ettinger, & Gibson, 2014; Sarnecka et al., 2007; Schaeffer, Eggleston, & Scott, 1974; Wynn, 1990, 1992). In the first stage, which for middle-class American children begins around the age of 2, children learn to recite a partial count list (e.g., *one*, *two*, *three*, *four*, *five*, etc.), often pointing at objects as they count (see Frye, Briasby, Lowe, Maroudas, & Nicholls, 1989; Fuson, 1988; Gelman & Gallistel, 1978). Despite being able to recite the count list, these children actually have little understanding of how number words represent quantity, and as a result are often classified as "non-knowers." Sometime after the age of 2, they advance to a second stage when they acquire an exact meaning for *one*, at which point they are called "1-knowers" (for discussion of what makes this meaning exact, see Barner, 2012; Barner & Bachrach, 2010; Brooks, Audet, & Barner, 2012; Spelke & Tsivkin, 2001a). After a long delay – sometime between 6 and 9 months – children in the US then learn an exact meaning for *two* to become "2-knowers." They then learn *three* ("3-knowers") and sometimes *four* ("4-knowers"). Up until this point, children are generally referred to as "subset-knowers" since they have only acquired the meanings for a subset of their number words. However, at the next stage children appear to recognize that the counting procedure can be used to enumerate sets and, more specifically, that the last word in a count routine labels the cardinality of the entire set (for discussion of knower levels, see Le Corre & Carey, 2007; Le Corre et al., 2006; Lee & Sarnecka, 2011; Piantadosi, Tenenbaum, & Goodman, 2012; Sarnecka & Carey, 2008; Wynn, 1990, 1992; Davidson, Eng, & Barner, 2012). At this stage, children are generally referred to as Cardinal Principle knowers ("CP-knowers").

Even after becoming CP-knowers, children still appear unsure of how the counting procedure represents number. According to some accounts (Sarnecka & Carey, 2008; Schaeffer et al., 1974; Wynn, 1990, 1992), children become CP-knowers – and understand how counting represents number – by noticing an isomorphism between the structure of the count list and the cardinalities that they represent. In particular, on such accounts, they become CP-knowers when they learn how the count list represents the successor principle – that every natural number n has a successor, defined as $n + 1$. However, recent evidence suggests that many so-called CP-knowers do not have knowledge of this principle and instead initially use a blind procedure for counting and giving correct amounts, with little insight into how or why this procedure works (Davidson et al., 2012). When asked what number comes after a small number like *four* in the count list, most CP-knowers have little difficulty reporting the next word in the list (Davidson et al., 2012; see also Fuson, Richards, & Briars, 1982). However, when told that a box contains 4 objects and asked how many there are when one more object is added, many of these same children randomly guess. Further, even those children who can correctly reply for very small numbers like *four* or *five* are often at chance for larger numbers that are well within their count list, and do not reply correctly for all numbers in their count list until around 6 years of age (Cheung, Rubenson, & Barner, 2015).

Although it remains controversial when children fully understand how counting represents number, the basic developmental trajectory – and its highly protracted nature – is robust. Also, the stage-like nature of children's learning has been corroborated by meta-analyses and computational models (Lee & Sarnecka, 2010, 2011; Sarnecka & Lee, 2009), which can model the delays between stages using a small number of parameters (Piantadosi et al., 2012). However, despite this consensus on the developmental facts, it remains unknown what *causes* transitions between the stages of number word learning and therefore which factors define the value of the parameters that describe delays between them. Consequently, we currently have little understanding of why it takes 1-knowers in the US 6–9 months to learn an exact meaning for *two* or what causes children to become CP-knowers when they do.

What we do know is that these stages – and the delays between them – vary in length between individuals and between groups of children (e.g., Dowker, 2008; Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006). Evidence for the malleability of the knower level transitions is potentially important to understanding what drives learning. Such evidence comes from several sources. First, frequency of input appears to affect learning: children who receive greater exposure to number words in caregiver speech appear to pass through knower level stages more quickly (Almoammer et al., 2013; Gunderson & Levine, 2011; Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010), whereas children who live is semi-numerate cultures where number is discussed infrequently learn number word meanings very late in development, albeit according to the same stages (Piantadosi et al., 2014). Second, children's rate of number word learning appears to be affected by the structure of the language that they learn. Some languages, like English and Russian, have obligatory singular-plural marking (e.g., one box, two boxes), which might help children to identify that *one* refers to singleton sets (since *one* occurs with singular agreement), whereas *two* and larger numbers refer to sets of more than one (since they are used with the plural; Carey, 2004, 2009). Consistent with this, children learning English and Russian are faster to become 1-knowers than children learning languages like Japanese and Mandarin Chinese, which do not have obligatory singular and plural marking (Barner, Libenson, et al., 2009; Li et al., 2003; Sarnecka et al., 2007). Languages like Slovenian and Saudi Arabic not only have singular-plural marking, but also have dual marking, which is used to refer to sets of precisely two (Corbett, 2000). Remarkably, children learning these two languages are faster to learn *one* and *two* (but interestingly not *three*) than children from any other previously studied group, even though these children appear to receive less training with number words than children from other countries (Almoammer et al., 2013).

These facts suggest that the transitions between knower level stages can be accelerated by the frequent and informative use of number words. However, these developmental facts alone leave open *why* frequent and informative input matters and thus why learning number words is so hard. Here we consider two, mutually compatible, possibilities. First, it is possible that the protracted transitions between the stages of number word learning are affected by gradual processes of conceptual change: 1-knowers may take months to become 2-knowers because they do not yet have a concept of "exactly two" and must construct this as part of learning the word (Carey, 2004, 2009; Sarnecka & Carey, 2008; Spelke & Tsivkin, 2001a). Consequently, hearing the word *two* used frequently and in supportive grammatical contexts may be important because this information helps children construct the concept "two". Similarly, becoming a CP-knower may involve a conceptual change that is speeded by frequent exposure to number words and the counting routine. Alternatively, the primary reason for delays between stages may not be conceptual, but may be due to the problem of identifying how the words in a particular language map onto numerical concepts (whether these concepts are innate, are constructed but easy to acquire, or are difficult to acquire but constructed sometime before the onset of number word learning). Consider, for example, a scenario in which the concepts "one", "two", and "three" are innately specified in the child's representational repertoire. In this case, the observed delays between knower level stages might be explained by the difficulty of identifying which words correspond to which set sizes. In the case of counting, a similar gradual, language-specific learning process might take place: learning the cardinal principle in English, for example, might not require conceptual change per se, but might instead involve a mapping of words onto innate concepts of cardinality (Leslie, Gelman, & Gallistel, 2008). This model, like the constructivist alternative, would also predict that frequency of input should affect children's rate of learning.

These two possibilities are by no means mutually exclusive, but they do make predictions that can distinguish the relative role that each factor plays in learning. In particular, they make different predictions for bilingual learners. Specifically, if the primary cause of delays between stages of number word learning is conceptual, then when children acquire knowledge in one language (e.g., English), this knowledge should facilitate learning in their secondary language of instruction (e.g., Spanish) with little additional language-specific experience required. However, if the primary challenge of number word learning is not conceptual – e.g., if these concepts are either very easy to learn or already present in some form before the beginning of number word learning – then the difficult problem may instead take the form of language-specific learning – mapping words onto concepts. In this case, bilingual children may exhibit relatively independent learning trajectories in each of their two languages with little evidence of transfer between them. The best predictor of their number word knowledge, in this case, may be how much input they receive in a particular language, rather than what they know in another language.

Past research on bilingual number knowledge indicates that some forms of knowledge do not readily transfer between languages. Specifically, bilingual speakers generally perform better when making arithmetic computations in their language of math instruction, even if this is not their dominant language otherwise (Dehaene, 1997; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Spelke & Tsivkin, 2001b). Not only are bilinguals faster when doing arithmetic in their language of math instruction, but they also make fewer errors than when tested in their other language (Kolers, 1968; Marsh & Maki, 1976; McClain & Huang, 1982). Finally, one past study of bilingual children suggests that simultaneously learning two dissimilar count lists within a language (like Korean, which has two counting systems) may slow children's memorization of the verbal sequence of each list (Song & Ginsburg, 1988). These facts are generally consistent with the hypothesis that some mathematical knowledge is stored in the form of linguistic rules, procedures, and associations and that transfer of such knowledge between languages is limited. However, these studies leave open the main question of the present study because they focus on arithmetic and language-specific routines, which may bear a fundamentally different relationship to language than number word meanings and the logic that relates them. In particular, because arithmetic performance benefits from the memorization of linguistically coded "math facts" (e.g., $5 \times 7 = 35$), it may be especially linked to a particular language. Currently, it is not known how more basic components of numerical knowledge relate to natural language, and whether there are differences in the role of language when acquiring number word meanings, counting procedures, or the basic logic of counting.

To explore these questions, we tested three groups of bilingual children in the U.S. who spoke either English and French or English and Spanish. Each child was tested with Wynn's (1990, 1992) Give-a-Number task in order to establish their knower level in each language. Also, each child was asked to count as high as they could in each language to identify the language for which they had received the most training with numbers (we called this language their "Primary Number Language" or NL1, and the other their "Secondary Number Language", or NL2). In addition, we tested when bilingual children acquire the successor principle in each language, and whether it exhibits transfer. To do so, we tested a subset of the CP-knowers in both languages with a third task, called the Successor Task (see Sarnecka & Carey, 2008).

2. Method

2.1. Participants

One hundred and forty-seven bilingual speakers of either English and French ($n = 20$; $M = 3; 8$; range = 2; 11–5; 0) or English and Spanish ($n = 127$; $M = 4; 6$; range = 2; 2–5; 6) from the San Diego metropolitan area participated in this study. Eighty-five of these children were from predominately low socioeconomic (SES) Hispanic families and were recruited through letters distributed by teachers at a local preschool that caters to English learners. The remaining 62 children were from primarily non-Hispanic Caucasian, upper-middle class families, from which 20 children were recruited at a preschool where instruction was conducted exclusively in French, while 42 children came from Spanish immersion preschools where children were taught in both English and Spanish. An additional 61 children participated but were excluded from analyses for failure to pass the language proficiency test in English ($n = 14$) or in Spanish ($n = 2$), for failure to complete the tasks in both languages ($n = 30$), for being trilingual speakers ($n = 2$), or for being able to count only as high as their determined knower level in either language ($n = 13$). This last exclusion criterion was important since knowledge can only transfer between two languages if the child first has words available to which this knowledge can be transferred.

As reported by the children's caregivers, 83 children were primarily Spanish speakers, 3 were primarily French speakers, and 44 were primarily English speakers. A subset of the children was equally proficient in both languages: 4 children were equally proficient in English and Spanish and 1 was equally proficient in English and French. Twelve parents did not respond. Although parent report is often predictive of children's verbal fluency in their native language it is not necessarily a reliable indicator of bilingual children's second language, especially when concerning mathematical knowledge, since children may be instructed in a language other than their primary language (i.e., a language that differs from that of their parents). Therefore, in addition to parent report, we also directly measured children's familiarity with numbers by assessing their counting ability in each language. This was then used to determine each child's ''Primary Number Language'' (NL1) and ''Secondary Number Language'' (NL2; for details, see Procedures).

2.2. Procedures

All tasks were carried out by experimenters who were either fluent or native speakers of the test language. Children were tested at a table across from the experimenter in a quiet corner of their classroom. Each session lasted approximately 20 min and included the Give-a-Number task to measure children's comprehension of number words and the Highest Count task to assess children's familiarity with and exposure to number words. Participants in the low SES Spanish–English group also received a language proficiency task to ensure their fluency in both languages. Among this group of low SES Spanish–English bilingual children, CP-knowers also received the Successor Task to test their understanding of the successor function (e.g., that the successor of 5 is equal to 5 + 1).[2] The order of tasks was fixed: (1) Language Proficiency; (2) Give-a-Number; (3) Highest Count; (4) Successor Task (for CP-knowers). All tasks were administered once in English and once in Spanish or French such that each child completed the tasks in one language before completing identical tasks in his or her other language. The order in which the languages were tested was randomized across children.

[2] Subset knowers in the low SES Spanish–English bilingual group also received an exactness task in order to test whether they knew that a change in cardinality required a change in number word label, even for unknown number words (Brooks, Audet, & Barner, 2013; Sarnecka & Gelman, 2004). This task was always after the Highest Count task. These data are not reported here.

2.2.1. Language proficiency test

When we began testing the low SES Spanish–English sample, we observed that some participants were not able to comprehend the task instructions in their second language (often English). Therefore, we introduced a brief language proficiency test that assessed comprehension of words representative of those used in the experimental session. The experimenter began by presenting the child with a plate and six plastic toys that were organized into three pairs and were placed next to the plate (e.g., two fish, two bananas, and two apples). Careful to avoid providing nonverbal cues, such as gestures, the experimenter asked the child to put the fish on the plate, repeating the instructions if necessary. Once the child responded, the experimenter removed the items, returning them to their original positions, and asked the child in the same manner to put the bananas and later the apples on the plate. Children who successfully gave only the fish when asked for fish, only the bananas when asked for bananas, and only the apples when asked for apples in both English and Spanish were included in the study.

2.2.2. Give-a-Number task

This task was adapted from Wynn (1990) and was used to assess children's knowledge of number words in each language. The experimenter began by presenting the child with a plate and ten plastic fish and inviting the child to play a game. For each trial, the experimenter asked the child to place a quantity of fish on the plate, avoiding singular and plural marking by asking, "Can you put *n* on the plate? Put *n* on the plate and tell me when you're all done." Singular-plural marking was avoided since this information could help children identify that *one* refers to a singleton set independent of knowing the meaning of the number word (Carey, 2004). Once the child responded, the experimenter then asked, "Is that *n*? Can you count and make sure?" and encouraged the child to count in the language tested. If the child recognized an error, the experimenter allowed the child to change his or her response. Following the completion of each trial, the objects were returned to their original positions and the next trial was administered until the child's knower-level could be ascertained.

Two versions of Give-a-Number were used. Children who participated in an earlier version of the study completed up to twenty-one trials, consisting of three trials for each of the seven numbers tested (i.e., 1, 2, 3, 4, 5, 8, and 10). The order was quasi-randomized such that each number was tested once before any number was repeated, thus resulting in three sets of seven numbers (see Lee & Sarnecka, 2010, 2011 for discussion of this method). Children who participated in a later version of the study were given a staircased version of the Give-a-Number task in which the experimenter first asked the child to put four fish on the plate (as in Wynn, 1990, Experiment 3). If the child succeeded – that is, by providing exactly four fish when asked for four – the experimenter then asked the child for the next number in the count list, in this case five. If, however, the child failed – that is, the child did not give four fish when asked for four – the experimenter instead asked the child for one fish. For each trial thereafter, if the child succeeded in giving the correct quantity (i.e., *n*), the experimenter proceeded by asking for one more on the subsequent trial (i.e., $n + 1$) up to the number eight. If the child failed to provide the correct quantity, the experimenter then asked for one fewer (i.e., $n - 1$). This pattern of titration continued until the child's knower level could be confidently identified. Children were defined as an *n*-knower (e.g., three-knower) if they correctly provided *n* (e.g., 3 fish) on at least two out of the three trials that *n* was requested and, of those times that the child provided *n*, two-thirds of the times the child did so it was in response to a request for *n*. If *n* was five or higher, the child was classified as a CP-knower.

2.2.3. Highest Count task

After the Give-a-Number task, the experimenter asked each child, "Can you count as high as you can?" in order to measure the child's relative exposure to numbers in each language and, therefore, determine his or her primary and secondary number language. If the child failed to respond or indicated that he or she did not know how to count, the experimenter provided the first number of the count list (e.g., *one*) with rising intonation in an attempt to encourage the child to continue counting. In the event that the child failed to respond after the prompt, the experimenter reassured the child and ended the task.

The child's Highest Count was defined as the largest number counted to before an error. For example, twelve was the highest number recorded for a child who omitted thirteen, whereas ten was the highest number recorded for a child who cyclically repeated the first ten numbers. In cases where children failed to accurately count at the onset of the task yet recited a string of numbers (e.g., "6, 7, 8...."), the highest number recorded was zero.

For each child, "Primary Number Language" (i.e., NL1) was defined as the language in which they counted highest. The other language was labeled their "Secondary Number Language" (i.e., NL2). For example, a child's NL1 was judged to be English if they counted to *twelve* in English and *diez* in Spanish. In cases where the Highest Count matched in both languages (e.g., *ten* and *diez*), the NL1 was defined as the child's primary language as reported by the parent ($n = 6$). We first appealed to Highest Count to define Primary and Secondary Number Languages. This was preferable to classifying children according to their L1 and L2 since many children receive their primary math instruction in their L2. Also, Highest Count was preferable to relying on parental report, since many of the parents in our samples were monolingual, making it difficult or impossible for them to assess their child's knowledge in both languages.

2.2.4. Successor Task

This task was modeled after Sarnecka and Carey (2008) and was designed to assess children's understanding of the successor principle, operationalized here as the knowledge that adding one object to a set (i.e., n) results in an increase of exactly one unit on the count list (i.e., $n + 1$). Children who were identified as CP-knowers were presented with an empty opaque box and a small red container filled with identical plastic toys of one kind (apples, bananas, strawberries, or lemons). To begin each trial, the experimenter directed the child's attention to the box by exclaiming, "Look, there's nothing in the box!" and permitted the child to look inside in order to confirm that the box was empty. The experimenter then held up the container of objects, stating that she had n items, poured the objects inside the box, and closed the lid (e.g., "I have fourteen apples. I'm putting fourteen apples in the box").

To ensure that the child remembered the number of items that the experimenter placed inside the box, they were asked a memory-check question, e.g., "How many apples are in the box?" If the child answered incorrectly or failed to respond, the objects were removed and the trial was repeated until the child responded correctly to the memory-check question.

Once the child correctly answered the memory-check question, the child watched carefully as the experimenter added one identical object to the box (e.g., one apple if the box contained apples, one banana if the box contained bananas, etc.). The experimenter then asked whether there were $n + 1$ or $n + 2$ objects, e.g., "Are there fifteen apples or sixteen apples in the box?" Irrespective of their response, children received neutral feedback following their selection (e.g., "Thank you") and the next trial was administered. The procedure was repeated until all eight trials were complete. Children were tested on two small numbers (5 and 8), three medium numbers (12, 14, and 16), and three large numbers (21, 23, and 27). The order in which the choice alternatives were presented was counterbalanced across trials and trial order was randomized.

3. Results

The focus of the analyses described below was to determine the factors that predict a child's NL2 knower level, and specifically, whether learning can be predicted by a child's knowledge in their NL1. To do this, we first identified each child's NL1 and NL2 using their Highest Count data. We then determined each child's knower level in each of their languages, and asked whether their NL1 knower level played a role in predicting their NL2 knower level, above and beyond effects due to Highest Count and age.

3.1. Highest Count in NL1 and NL2

Highest Count was used to identify children's NL1 and NL2. In the high SES French–English group ($n = 20$), the average Highest Count in French was 13.9 ($SD = 7.0$) and in English was 10.7 ($SD = 4.0$). Sixteen of these children counted higher in French, and 4 counted higher in English. Thus, the NL1 for 16 of the 20 children in this sample was French. In the high SES Spanish–English group ($n = 42$), the average Highest Count in Spanish was 12.3 ($SD = 9.6$) and in English was 20.9 ($SD = 20$). Eight children counted higher in Spanish, and 32 counted higher in English. Two additional children counted equally high in Spanish and English. In these cases we classified their NL1 by their parent report ($n = 1$, English; $n = 1$, Spanish). Overall, the NL1 was English for 33 out of 42 of the children in this group. In the low SES Spanish–English bilinguals group ($n = 85$), the average Highest Count in Spanish was 17.7 ($SD = 11.0$) and in English was 10.4 ($SD = 4.1$). Sixty-eight children counted higher in Spanish and 13 counted higher in English. Four additional children counted equally high in Spanish and English, and in these cases we classified their NL1 by their parent report ($n = 4$, Spanish). Thus, the NL1 was Spanish for 72 of 85 children in this group.

The three groups of bilingual children did not statistically differ in their NL1 counting ability, $F(2,144) = 1.84$, $p = 0.16$, or NL2 counting ability, $F(2,144) = 1.27$, $p = 0.28$. However, the groups did differ in Age, $F(2,144) = 30.77$, $p < 0.001$. The low SES Spanish–English group (M = 4;9, SD = 6 months) was significantly older than the high SES Spanish–English group (M = 4;2, SD = 9 months), $t(125) = 5.13$, $p < 0.001$, who were in turn significantly older than the French–English group (M = 3;8, SD = 6 months), $t(60) = 2.32$, $p = 0.02$. Thus, the low SES Spanish–English children were relatively delayed in learning to recite the count list.

3.2. Give-a-Number task

Table 1 presents the number of children identified as 'N-knowers' in each group and in each language. Across the three groups of bilingual children, there were 33 non-knowers, 29 1-knowers, 34 2-knowers, 24 3-knowers, 18 4-knowers, and 156 CP-knowers (where each child contributed two knower level classifications since each spoke two languages).

Table 2 presents the number of children who were identified as 'N-knowers' in their NL2 as a function of their knower level in NL1, while Fig. 1 illustrates the proportion of children who were 'N-knowers' in their NL2 by their NL1 knower level.

Table 1
Number of children identified as 'N-knowers' in each group (and in each language).

	High SES				Low SES	
	French–English		Spanish–English		Spanish–English	
	French	English	Spanish	English	Spanish	English
Non-knowers	3	2	12	6	3	7
1-knowers	2	3	2	7	3	12
2-knowers	4	4	6	3	11	6
3-knowers	2	2	–	3	12	5
4-knowers	2	–	2	1	7	6
CP-knowers	7	9	20	22	49	49
Total	20	20	42	42	85	85

Table 2
Number of children identified as 'N-knowers' in their NL1 as a function of their NL2 knower level.

NL1 Knower Level	NL2 Knower Level						Total
	Non-knowers	1-knowers	2-knowers	3-knowers	4-knowers	CP-knowers	
Non-knowers	8	1	–	–	1	–	10
1-knowers	8	5	–	1	1	–	15
2-knowers	1	7	10	1	–	1	20
3-knowers	5	1	3	6	1	–	16
4-knowers	1	–	–	–	3	5	9
CP-knowers	–	–	1	–	3	73	77
Total	23	14	14	8	9	79	147

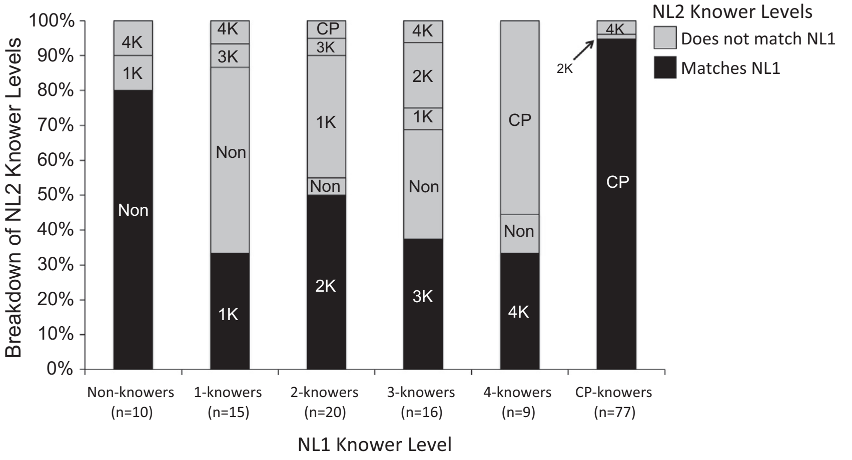

Fig. 1. Distribution of NL2 Knower Level by NL1 Knower Level.

3.3. Predictors of NL1 Knower Level

Before assessing the predictors of NL2 knower level, we first explored the factors that affect NL1 learning. To do this, we conducted an ordinal logistic regression to predict NL1 Knower Level from NL1 Highest Count, Age, and Group (high SES Spanish–English, low SES Spanish–English, and French–English), $R^2 = 0.320$, $\chi^2(9) = 135$, $p < 0.001$. Effect likelihood ratio tests of this model found that NL1 Highest Count, $\chi^2(1) = 55.1$, $p < 0.001$, and Age, $\chi^2(1) = 26.5$, $p < 0.001$, were significant predictors of NL1 Knower Level but Group was not, $\chi^2(2) = 2.09$, $p = 0.35$. Thus, children's NL1 Knower Level increased with age and counting ability, but did not differ significantly across groups.

3.4. Predictors of NL2 Knower Level

At each NL1 knower level, the proportion of children who were at the same knower level in their NL2 was higher than expected by chance, all *ps* < 0.05, see Table 2 and Fig. 1. This rate of concordance was particularly high for CP-knowers (95%). However, concordance rates offer only a weak test of whether NL1 knowledge facilitated NL2 learning. First, because knower levels in both NL1 and NL2 increase as a function of age and exposure to number words, NL1 and NL2 knower levels should be highly correlated as a function of age regardless of whether there is transfer. Second, CP-knowers were highly frequent in our sample (54% of all children were CP-knowers in their NL2). Therefore, if NL1 knower level and NL2 knower level were independent, we would expect approximately 54% of NL1 CP-knowers in our sample to also be NL2 CP-knowers by chance alone. For comparison, only 10% of our participants were NL2 1-knowers (and thus, we would expect only 10% of the NL1 1-knowers in our sample to also be NL2 1-knowers by chance).

Our question was therefore not simply whether NL1 and NL2 knowledge are correlated, but whether the relationship between a child's two knower levels is explained purely by their experience in each language independently, or instead if NL1 knower level can explain additional variance in NL2 knower level above and beyond age and counting ability, as would be predicted by transfer of number word meanings across languages. To test this question, we compared two ordinal regression models. This allowed us not only to consider age and counting ability as factors but also to account for differences in base rates across knower levels. The first model (Give-A-Number Model 1) predicted NL2 knower level from Age, Highest Count, and Group alone, and the second model (Give-A-Number Model 2) added NL1 knower level, and thus asked whether NL1 knower level explained any additional variance.

An effect likelihood ratio test found that Give-A-Number Model 2 is significantly better than a Give-A-Number Model 1 ($\chi^2(5) = 53.6$, $p < 0.001$). Additionally, we compared the models using three different sets of criteria. Misclassification rates and root mean squared error (RMSE) indicate fit whereas the Akaike Information Criterion Value (AICc) measures model parsimony and is used to compare models while balancing goodness of fit with the number of parameters included in the model, with a lower number indicating a better model (Akaike, 1974). The Give-a-Number Model 1, $R^2 = 0.36$, $\chi^2(5) = 147$, $p < 0.001$, (which excluded NL1 Knower Level as a predictive factor) had an overall misclassification rate of 0.33, a RMSE (root mean squared error) of 0.55, and an AICc of 287. Adding NL1 Knower Level in Give-a-Number Model 2, $R^2 = 0.49$, $\chi^2(10) = 200$, $p < 0.001$, reduced the misclassification rate to 0.27, lowered the RMSE to 0.49, and lowered the AICc to 245, suggesting that it was a better fit and a more parsimonious model than Give-a-Number Model 1. These findings confirm that NL1 knower level was a significant predictor of NL2 knower level, overall. Effect likelihood ratio tests also found that NL2 Highest Count, $\chi^2(1) = 7.72$, $p = 0.006$, and Age, $\chi^2(1) = 4.35$, $p = 0.037$, significantly improved the model's ability to predict NL2 Knower Level. However, NL1 Highest Count, $\chi^2(1) = 0.208$, $p = 0.65$ and Group, $\chi^2(2) = 1.74$, $p = 0.42$ did not improve the model.

This first set of analyses indicates that NL1 knower level is generally an important predictor of NL2 knower level, even when controlling for NL2 counting ability, as measured by Highest Count. However, different learning processes may drive children's learning of small (i.e., *one*, *two*, *three*, and *four*) and large (i.e., *five* and greater) number word meanings. Consistent with this possibility, parameter estimates from Give-a-Number Model 2 indicated that transfer occurred at the CP-knower level – and perhaps the 4-knower level – but not at the other knower levels. These estimates, reported in Appendix A, indicate a numerically larger effect size for the 4-knower and CP-knower levels than for the other subset levels.

To verify this conclusion, we next assessed transfer in subset knowers and CP-knowers separately. First, we asked whether there was evidence of transfer when only "subset knowers" – i.e., children who were not CP-knowers in either language – were considered, $n = 64$. Again we compared two ordinal logistic models, Subset Knower Model 1, $R^2 = 0.16$, $\chi^2(5) = 30.3$, $p < 0.001$, which did not include NL1 Knower Level as a predictor, and Subset Knower Model 2, $R^2 = 0.20$, $\chi^2(9) = 37.7$, $p < 0.001$, which did. An effect likelihood ratio test found that Subset Knower Model 2 was not significantly better than Subset Knower Model 1 ($\chi^2(4) = 7.40$, $p = 0.12$). The first model, Subset Knower Model 1, had a misclassification rate of 0.53, a RMSE of 0.69 and an AICc of 184. Although including NL1 Knower Level in Subset Knower Model 2 slightly decreased the misclassification rate, 0.48, and RMSE, 0.67, the AICc was 188, indicating the Subset Knower Model 1 was overall a better explanation of the data. Effect likelihood ratio tests found that only NL2 Highest Count, $\chi^2(1) = 5.61$, $p = 0.018$, and Age, $\chi^2(1) = 3.93$, $p = 0.047$, were significant predictors of NL2 knower-level whereas NL1 Highest Count, $\chi^2(1) = 0.231$, $p = 0.63$, Group, $\chi^2(2) = 0.959$, $p = 0.62$, and NL1 Knower Level were not, as noted above. Appendix A reports parameter estimates for each knower level, none of which were statistically different from 0 for subset stages, suggesting that there was not in fact transfer at the 4-knower stage (see Appendix A for further discussion). These results show that, among subset knowers, NL1 knower level was not a significant predictor of NL2 knower level. Instead, only age and NL2 Highest Count were significant predictors. In sum, we found no evidence of transfer from NL1, suggesting that the effects of transfer reported in Give-a-Number Model 2, above, are likely attributable to learning at the CP-knower stage.

Second, we asked whether there was evidence of transfer when only the transition to the CP-knower stage was considered. We compared two nominal logistic models that coded children as either a CP-knower or subset-knower in each language: CP-knower Model 2, $R^2 = 0.73$, $\chi^2(6) = 149$, $p < 0.001$, which included CP-knower Status in NL1 as a predictor, and CP-knower Model 1, $R^2 = 0.58$, $\chi^2(5) = 115$, $p < 0001$, which did not. An effect likelihood ratio test found that CP-knower Model 2 was a significantly better model of the data than CP-knower Model 1 ($\chi^2(5) = 53.6$, $p < 0.001$). CP-knower Model 1 had a RMSE of 0.31, a misclassification rate of 0.14 and an AICc of 100. In CP-knower Model 2, adding NL1 CP-knower Status not only improved the fit as indicated by a lower RMSE, 0.23, and misclassification rate, 0.07, but also lowered the AICc, 69, indicating a more parsimonious model. Thus, we found converging evidence that cross-linguistic transfer may occur at the CP-knower stage, but not at the subset knower levels. Effect likelihood ratio tests also found that NL1 CP-knower Status, $\chi^2(1) = 33.4$, $p < 0.001$, NL2 Highest Count, $\chi^2(1) = 8.58$, $p = 0.003$, and Age, $\chi^2(1) = 5.51$, $p = 0.019$, significantly predicted NL2 CP-knower Status. There was only a marginal effect of Group, $\chi^2(2) = 5.08$, $p = 0.08$, and no effect of NL1 Highest Count, $\chi^2(1) = 0.83$, $p = 0.36$.

3.5. Performance on the Successor Task

The goal of the Successor Task was to investigate whether learning the successor principle drives the transition from subset knower to CP-knower (Sarnecka & Carey, 2008) or whether successor knowledge is instead acquired after children become CP-knowers, as recent evidence seems to suggest (Davidson et al., 2012). This question is important because although our analyses indicate that transfer occurs at the CP-knower stage, they do not address *what* exactly transfers – knowledge of a counting procedure or conceptual knowledge related to the logic of counting. To investigate this, we first asked whether all CP-knowers understood the successor principle – i.e., that every natural number n has a successor defined as $n + 1$. Evidence that some CP-knowers do not have this knowledge would suggest that the effect of CP-knower transfer reported above does not reflect transfer of this conceptual understanding of the successor principle. Second, we asked whether children who exhibited knowledge of the successor principle in their NL1 also showed such knowledge in their NL2.

Children in the low SES Spanish–English group who were CP-knowers in both languages were tested with the Successor Task. One child did not complete this task in both languages and thus was excluded from analysis. For each child, the response from a particular trial was included in the analyses if the child's Highest Count in the corresponding language was greater than the target number of the trial. For example, data from the "12" trial was only included if a child could count to at least 13 in the tested language. This was done to ensure that the task measured children's conceptual understanding of the successor principle, rather than their knowledge of the count list (since in this case a failure to express knowledge of the successor principle would be poor evidence for lack of transfer). Children who had fewer than two valid trials in either language ($n = 5$) were excluded, resulting in a sample of 41 children ($M = 4$; 11, $SD = 4.6$ months). Among these CP-knowers, the average NL1 Highest Count was 22.5 ($SD = 10.0$), while average NL2 Highest Count was 12.7 ($SD = 3.2$).

3.6. Successor Knowledge in NL1

Replicating Sarnecka and Carey (2008), we found that group performance of CP-knowers on the Successor Task was greater than chance (i.e., 50%) in NL1 ($M = 68\%$, $SD = 23\%$), $t(40) = 5.00$, $p < 0.001$. If the successor principle is the piece of knowledge that allows subset knowers to become CP-knowers then all CP-knowers should be above chance. Thus, the critical question is not whether CP-knowers are above chance as a group but whether each CP-knower is above chance when his/her individual data are examined.

Similar to Davidson et al. (2012), we found that CP-knowers were a heterogeneous group, with performance on the task ranging from 0% to 100% between children. Despite the fact that all participants were CP-knowers, 30% were at or below chance in their NL1. Following Davidson et al., we explored the variability between children by dividing them into three groups using the Highest Count Task: Children whose Highest Count was 16 or less in a given language were classified as "Low Counters" for that language; children whose Highest Count was between 17 and 27 were classified as "Medium Counters"; and children whose Highest Count was higher than 27 were classified as "High Counters." We then explored children's performance in each group for numbers within their respective counting ranges.

Replicating Davidson et al., we found that Low Counters performed quite poorly and many did not differ from chance. As shown in Fig. 2a, 50% of Low Counters performed either at or below chance on Small Numbers. Medium Counters exhibited mixed performance on the Successor Task (see Figs. 2a and 2b). Only 25% of them performed above chance on Small Numbers, though 83% were above chance

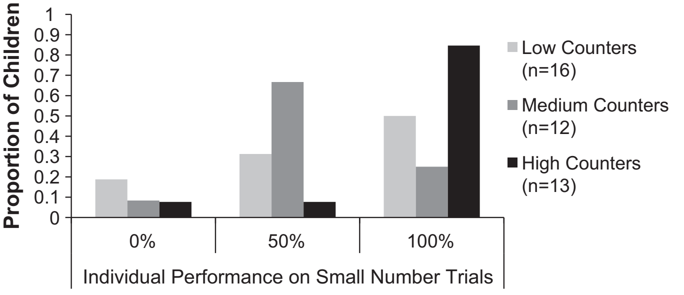

Fig. 2a. Low, Medium, and High counters categorized by their performance (0%, 50%, or 100%) on the small number trials of the Successor Task (i.e., 5 and 8). Chance is 50%.

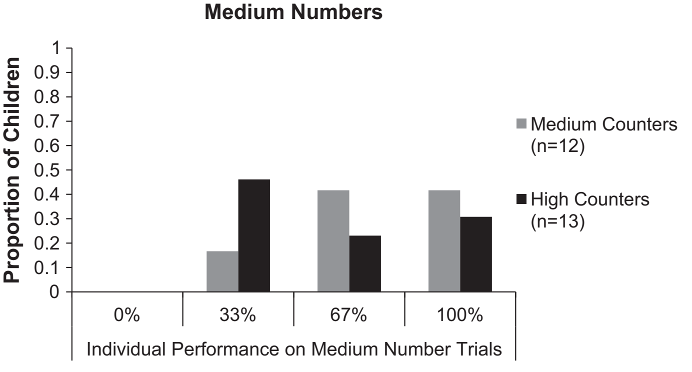

Fig. 2b. Medium and high counters categorized by their performance (0%, 33%, 66% or 100%) on the medium number trials (i.e., 12, 14, and 16) of the Successor Task. Low counters are not depicted because these numbers were out of their counting range. Chance is 50%.

on Medium Numbers.[3] Finally, for High Counters 85% were above chance for small numbers, 50% for medium numbers, and 92% for large numbers (see Figs. 2a–2c). Overall, these results replicate the findings of Davidson et al. (2012) that many CP-knowers (1) cannot deploy the successor principle for all numbers in their count list, and (2) often fail for even very small numbers. These data are consistent with the idea that becoming a CP-knower does not require conceptual understanding of the successor principle. Instead, children may learn the successors of specific numerals incrementally, before abstracting the successor principle as it applies to the entire count list.

3.7. Transfer of successor principle from NL1 to NL2

Although children appear to acquire knowledge of the successor principle *after* they become CP-knowers, it is nevertheless possible that this knowledge transfers when children finally do acquire it after learning the counting procedure. Note, however, that such transfer would need to be item-specific if it exists at all, since even within a language it would appear that there is an absence

[3] This relatively high performance for Medium Counters on Medium numbers is surprising, given (1) their relatively poor performance for small numbers, (2) the poor performance of High Counters on Medium numbers, and (3) the finding by Davidson et al. that children generally perform poorly on Medium numbers. These facts suggest that this particular result may not be reliable. However, it could also be related to the fact that medium counters have only recently become fluent with reciting medium sized numbers; Medium counters may have an easier time accessing the unusual phonological forms of the teens decade than high counters for whom the teens decade are an exception to a general rule that both lower numbers and higher numbers follow.

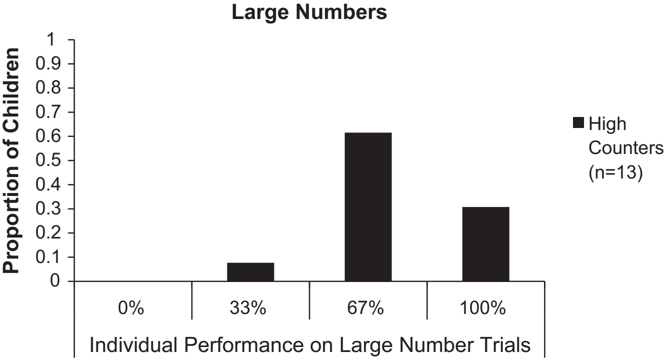

Fig. 2c. High counters categorized by their performance (0%, 33%, 67% or 100%) on the large number trials (i.e., 21, 23, and 27). Low and Medium counters are not depicted because these numbers were out of their counting range. Chance is 50%.

of transfer between numbers – i.e., children do not inductively extend the successor principle to all numbers in their count list until sometime after becoming a CP-knower.

All but one of our participants were Low Counters in their NL2, and it was therefore not possible to evaluate their performance for Medium and Large Numbers. Thus, we restrict our analyses to their performance on Small Numbers (i.e., 5 and 8). The approach we used to analyze the transfer of successor knowledge from NL1 to NL2 was similar to that used in our analysis of knower level transfer. We asked whether children's knowledge of successors in their NL1 could explain variance in their knowledge of successors in their NL2 that could not otherwise be accounted for by age and counting abilities alone. To accomplish this, we compared two standard least square regression models of NL2 Successor Knowledge. Both models included NL1 and NL2 Highest Count, and Age as predictors. However, only Model 2 included NL1 Successor Knowledge as a predictor.

To compare these models, we computed their relative fits to the data using two sets of criteria. Successor Model 1, $F(3,37) = 0.68$, $p = 0.57$, $R^2 = 0.05$, which excluded NL1 Successor Knowledge, had a RMSE (root mean squared error) of 0.35 and an AICc (Akaike Information Criterion value) of 37.2. Successor Model 2, $F(4,36) = 3.3$, $p = 0.02$, $R^2 = 0.27$, which included NL1 Successor Knowledge, lowered the RMSE to 0.31 and lowered the AICc to 29.5, suggesting that Successor Model 2 was a better explanation of the data than Successor Model 1. Furthermore, fixed effect tests indicated that none of the factors that were included in Successor Model 1 (Age, NL1 Highest Count, NL2 Highest Count) were significant predictors of NL2 Successor Knowledge, all $Fs(1,37) < 1.7$, all p's > 0.20.

In Successor Model 2, fixed effect tests revealed that knowledge of the successors of Small Numbers in NL1 was the primary predictor of successor knowledge in NL2, $F(1,36) = 10.4$, $p = 0.003$. However, there was no significant effect of NL2 Highest Count, $F(1,36) = 0.66$, $p = 0.42$, NL1 Highest Count, $F(1,36) = 0.69$, $p = 0.41$, or Age, $F(1,36) = 0.05$, $p = 0.83$. These results are consistent with the idea that knowledge of successors in a child's first language transfers to their second language, at least for some numbers.

To further examine the effect of transfer of successor principle knowledge, we asked whether knowing the successor of a specific number in NL1 (e.g., *cinco*) predicted successor knowledge of the same number in NL2 (e.g., *five*). To test this, we asked which factors predicted children's performance on the "5" and "8" trials on the Successor Task. We chose "5" and "8" because all children contributed data to these two trial types.[4] To predict children's NL2 performance on the "5" trial, we implemented an ordinal logistic regression, $R^2 = 0.09$, $\chi^2(3) = 0.2$, $p = 0.2$, using NL1 performance on "5", NL1 performance on "8", and NL2 performance on "8". Effect likelihood ratio tests showed that NL1 performance on "5", $\chi^2(1) = 4.6$, $p = 0.03$, was the only significant predictor. Neither NL1 perfor-

[4] This analysis was not performed for "12" because there were only 18 children who could count up to at least 13. Similarly, we did not analyze the effect of transfer for "18" because only 1 child could count higher than 18 in NL2.

mance on "8", $\chi^2(1) = 0.18$, $p = 0.67$, nor NL2 performance on "8", $\chi^2(1) = 0.007$, $p = 0.93$, was significant. Likewise, we conducted another ordinal logistic regression to predict NL2 performance on "8", $R^2 = 0.18$, $\chi^2(3) = 10.1$, $p = 0.02$,, and effect likelihood ratio tests found that NL1 performance on "8", $\chi^2(1) = 6.1$, $p = 0.01$, was a significant predictor, but NL1 performance on "5", $\chi^2(1) = 2.8$, $p = 0.10$, and NL2 performance on "5", $\chi^2(1) = 0.18$, $p = 0.67$, were not. Thus, we found evidence that knowing the successors of "5" and "8" in NL1 specifically predicted identical knowledge in NL2. We conducted an additional analysis accounting for chance responding by guessing and found that NL1 knowledge of the successors of "5" and "8" almost perfectly predicted the same knowledge in NL2, see Appendix B.

4. General discussion

We investigated number word learning in 2- to 5-year-old bilingual children. The primary goal of this study was to investigate the long delays that exist between stages of number word learning and examine whether these delays are best explained by processes of gradual conceptual change or instead by language-specific processes of mapping words onto concepts. Our data indicate that early on in the 1-knower, 2-knower, and 3-knower stages of number word learning, knowledge is acquired independently in each language. Neither counting ability nor knower level in children's Primary Number Language (NL1) predicted their NL2 knower level. For example, children who knew *two* in English did not show an advantage for learning *dos* in Spanish. Early NL2 number word knowledge was instead predicted solely by age and experience with NL2 number words (as indicated by counting ability). In contrast, children's classification as Cardinal Principle knowers in their NL2 was strongly predicted by being a CP-knower in their NL1 (in addition to NL2 counting ability, but not NL1 counting ability). This result suggests that the scope of inference at this stage is not restricted to a particular language, but may instead involve a moment of insight that applies to counting in general, independent of any particular language. Finally, we replicated previous findings that not all CP-knowers understand the successor principle, and thus that learning this principle does not likely drive children's ability to use the counting procedure to label and generate sets. Instead, children likely learn about the successor principle gradually in both languages after they become competent users of the counting procedure, with some evidence of transfer.

These results allow us to draw several important conclusions regarding the nature of number word learning, not only as it occurs in bilinguals, but also as it occurs in monolingual learners. First, our results support the intuition described in previous studies (Le Corre & Carey, 2007; Sarnecka & Lee, 2009; Schaeffer et al., 1974; Wynn, 1990, 1992, etc.) that number word learning involves multiple, discontinuous stages of learning, such that small and large number words are acquired via different mechanisms. In our study, the rate at which children learned labels for "one", "two", and "three" was best explained by their exposure to these words in a particular language, and not by whether they had previously learned corresponding words in another language. This result suggests that the delays between subset knower stages are not caused by the problem of constructing new concepts, but instead are due to difficulties identifying which concepts correspond to which words. The situation is different, however, when it comes to how children learn to count. Although learning the counting procedure requires substantial linguistic experience with number words, once children learn the counting procedure, this knowledge becomes available to children in a format that transcends a particular language.

Second, although our findings cannot decide whether concepts like "one", "two", and "three" are innate, they nevertheless suggest that the problem of conceptual change cannot easily explain the protracted transitions between knower level stages. By some accounts, the gradual, stage-like process of acquiring small number words like *one*, *two*, and *three* provides evidence that these concepts must not be innate (see Le Corre & Carey, 2007, for one example). This argument rests on the assumption that delays between stages must be driven by the problem of constructing new concepts. However, the current results bring this assumption into question: number word learning requires not only the availability of relevant concepts but also the ability to identify how these concepts are expressed by a particular language. Our data suggest that this second problem may not be trivial, and may in fact account for most of the delay between subset knower levels. To the extent that conceptual change occurs when learning small number words, our results suggest these changes are likely to be to the meanings of lexical items within a particular language, not to domain general concepts in a ‘‘language of thought”.

A third important conclusion suggested by our study is that becoming a Cardinal Principle knower is likely much more complex than often argued, and may involve several distinct steps as argued by Davidson, Eng, and Barner (2012). In their discussion of the Cardinal Principle, Sarnecka and Carey (2008) note two ways in which we might characterize the knowledge children acquire when they become CP-knowers, using Wynn's (1990, 1992) Give-a-Number criteria:

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

Sarnecka and Carey argued in favor of the second hypothesis on the basis of data which showed that children classified as CP-knowers by Wynn's Give-a-Number task were more likely than subset knowers to exhibit understanding of the successor principle (as measured by what we have called the ‘‘Successor Task”). However, although it may be true that CP-knowers are more likely than subset knowers to understand the successor principle, Davidson et al. (2012) noted that CP-knowers are nevertheless highly heterogeneous in their knowledge, and that the least experienced counters among them often show no evidence of understanding the successor principle. Also, they reported that even the most experienced counters in their sample were not yet able to infer the successors of all numbers in their count list, despite being able to generate the numeral that followed a particular word in the count list (e.g., when asked, ‘‘What comes after four, five or six?”; see Fuson et al., 1982, for a similar result). These findings led them to conclude that becoming a CP-knower cannot be driven by learning the successor principle and generalizing it to all numbers. Instead, Davidson et al. argued in favor of the first hypothesis described by Sarnecka and Carey, that when children become CP-knowers, their knowledge is initially purely procedural in nature. In other words, these children may not actually understand the Cardinal Principle, but may better be understood as ‘‘Counting Procedure” knowers, at least until they show evidence of understanding the logic of counting. Our results are consistent with this conclusion. Like Davidson et al. (2012), we found that children first learn the counting procedure before showing evidence of understanding the successor principle, even for numbers less than ten, and that even our most experienced counters were still far from having generalized the successor principle to all numbers in their count list.

However, our findings add an additional component to the story reported by Davidson et al. First, consistent with the idea that learning the successor principle involves a type of conceptual change, we found that if children were able to infer successors in one language, they were generally able to do so in their second language, too. Also, consistent with Davidson et al.'s hypothesis that successor knowledge is not generalized across all numbers, we found that this transfer was remarkably restricted in scope, and did not extend to all numbers in children's count lists, but instead to specific numbers like the labels for ‘‘five” and ‘‘eight”. Because this effect could be tested on only two numbers and remains to be replicated it remains possible that the degree of specificity we report is an anomaly. Still, it is clear that despite showing some evidence of transfer *across* languages, children did not generalize knowledge of the successor principle *within* languages. Although we are unable to explain precisely why this might be, given our current dataset the result is consistent with two broad alternatives. First, one possibility is that children's failure to generalize to larger numbers may be due to their relatively less fluid knowledge of the count sequence for larger numbers. Although children are sufficiently familiar with the list to allow them to name the successors of large words (Davidson et al., 2012; Fuson et al., 1982), the additional problem of doing this while simultaneously tracking changes to the cardinality of a set may prove especially taxing for larger, less familiar, number sequences. Against this hypothesis, however, children's ability to apply the successor function for a particular number in their NL2 was *not* predicted by their counting ability in this language. Instead, surprisingly, it was strongly predicted by their ability to apply the successor function for the same number in their NL1. This result is difficult to explain on the hypothesis that children's familiarity with the counting procedure mediates their expression of the successor principle.

Another possibility is that successor knowledge is item-based both within a language and across languages – i.e., that transfer is item-based. On this hypothesis, the mechanism that allows children to transfer knowledge of a blind counting procedure from NL1 to NL2 – i.e., the procedure that makes them CP-knowers – may involve forming a type of structure mapping between the count lists of their two languages, such that knowledge about particular sequences within count lists is transferred (for discussion of structure mapping, see Sullivan & Barner, 2012, 2014a, 2014b). Children might know that *five* and *cinco* represent the same quantity, as do *six* and *seis*, such that when they learn that *five* plus *one* equals *six* they can readily infer that *cinco* and *uno* equal *seis*. If we take seriously the specificity of transfer we report, then such a hypothesis may be the most parsimonious explanation of children's behavior. Critically, regardless of how we explain children's difficulty with larger numbers, learning the successor principle could not possibly be what causes children to become CP-knowers, since on this hypothesis children *would* need to reason about both cardinalities and the structure of the count list simultaneously, at least for the very smallest numbers. Currently there is no evidence that they can do so, and substantial evidence that they cannot.

While the focus of this study was to explain delays between knower levels, one additional result is also implicit in the data that we reported. Generally it is assumed that because children move through knower level stages one-by-one in sequence, this sequence must therefore be necessary. Indeed, by some accounts, the sequence of knower level stages may be universal (Piantadosi et al., 2014). Our study suggests that this need not be true, and that stages may in fact be skipped, at least in principle. Although we do not have longitudinal data to directly address this question, we found that when children were identified as CP-knowers in one language, they were very likely to also be CP-knowers in their second language. Because this was *not* true for lower knower levels, it would appear that a child who is a 3-knower in one language but only a 1- or 2-knower in their other might become a CP-knower in both languages at once, thereby skipping several stages in their secondary number language. This is interesting because it suggests that, at least in principle, small number word meanings can be defined from the start by their role in the counting routine, rather than by associations between individual words and set sizes (as is presumably normally the case). This result does not necessarily mean that such a process occurs in monolingual children, but it does raise the possibility that stages could in principle be skipped given the appropriate training.[5] Future research should investigate this question using a longitudinal approach, to explore whether stages are indeed skipped and, if they are, whether the meanings of *one*, *two*, and *three* are measurably different among children who learn them via transfer from the cardinal principle vs. those who learn them one-by-one.

Before concluding, we should note that our study leaves open a central puzzle of number word learning. Specifically, on the assumption that delays between subset knower level stages are driven by the problem of mapping numerals to set sizes, it remains unknown why this mapping problem might be so difficult. Three observations are potentially relevant: (1) object arrays are increasingly difficult to represent as they grow larger, whether they are represented via the approximate number system (Dehaene, 1997) or via systems of object tracking (e.g., Alvarez & Franconeri, 2007), and in each case exhibit reduced noise as children grow older (Halberda & Feigenson, 2008; Trick, Jaspers-Fayer, & Sethi, 2005); (2) numerals decrease exponentially in frequency as they grow larger, such that across languages the label for "one" is vastly more frequent than the label for "two", which is in turn much more frequent than the label for "three" (Dehaene & Mehler, 1992), (3) number words are used in

[5] Consistent with this, computational models of the counting stages suggest that, to the extent that learning the cardinal principle is a process of inference to the best explanation where the successor principle is contained in a child's initial hypothesis space, children may indeed skip steps given sufficient robust training with, e.g., *one* and *two*, but not *three* or *four* (Piantadosi et al., 2012).

heterogeneous linguistic and pedagogical contexts, and may often be used in ways that provide little direct evidence of their meanings to children, such as blind counting routines.

Currently there is no definitive evidence to decide between these possibilities, and it is likely that all three factors play some role. However, there is some reason to believe that (3) – i.e., the way in which numerals are used in caregiver input – may be especially important. Consistent with this, Almoammer et al. provide evidence that cross-linguistic differences in grammatical structure may be more important than counting exposure to learning early number word meanings. In support of this, they found that children learning Slovenian acquire the labels for "one" and "two" up to a year earlier than children learning English, and do so despite receiving much less training with counting (as evidenced by their very minimal counting abilities even at age 5). As Almoammer et al. explain, languages like Slovenian and Saudi Arabic have singular and plural marking, like English, but also dual marking which encodes sets of two, and must be used whenever the label for "two" is used. Consequently, children who comprehend dual marking in these languages may be able to infer the meaning of the label for "two" based on its use in dual contexts. Similarly, as predicted by Bloom and Wynn (1997), children learning English (which has a singular-plural distinction) are faster to learn the meaning of the label for "one" than children learning Japanese and Chinese, which lack obligatory singular-plural marking (Barner, Libenson, et al., 2009; Carey, 2004; Li et al., 2003; Sarnecka et al., 2007). Finally, children who have relatively better understanding of quantifiers within a language (e.g., *all*, *some*, *many*, *most*) are also faster to learn number word meanings (Barner, Chow, & Yang, 2009; Barner, Libenson, et al., 2009), a result also predicted if children use syntactic cues to infer number word meanings (see Bloom & Wynn, 1997). Although such evidence suggests that knower level stages are highly malleable and can be accelerated by informative grammatical cues, they do not rule out a role for frequency or for perceptual noise. Not only might these factors directly affect number word learning, but they could also impact children's acquisition of grammatical structures like singular, plural, and dual morphology, which are highly frequent in language, but which never encode precise quantities greater than three (consistent with a perceptual limit on their meanings; Corbett, 2000). Future studies should explore this question in greater detail, and how frequency, format, and set size interact to affect the rate of number word learning.

To summarize, in a large and diverse sample of bilingual children, we found evidence that language-specific learning, rather than processes of conceptual change, likely explains delays between early knower levels. During these stages, children must learn the numbers independently in each language. However, once a child learns the counting procedure in either language, they are able to transfer this procedural knowledge to their other language. Next, children next learn the successor principle, which also transfers across languages, but in an item-based, incremental, fashion. Overall, these data suggest that in bilinguals and monolinguals alike number word learning is importantly discontinuous and depends on different learning processes at different moments in development, which require fundamentally different types of changes.

Acknowledgments

VIP Village Preschool in Imperial Beach, CA; The Language Academy in San Diego, CA; La Petite École English French Learning Academy in San Diego, CA; Bonita Learning Academy in Chula Visa, CA; Lulus Montessori Preschool in San Diego, CA.

Funded by grants from the NSF (1420249) and the James S. McDonnell Foundation to DB.

Appendix A

A.1. Knower Level transfer analyses

In the *Predictors of NL2 Knower Level* section in the main text, we describe a model, Give-A-Number Model 2, which tests how NL1 Highest Count, NL2 Highest Count, Age, Group, and NL1 Knower Level predict NL2 Knower Level. We describe the effect likelihood ratio tests, which indicate that NL2

Highest Count, Age and NL1 Knower Level but not Group or NL1 Highest Count are statistically significant predictors of NL2 Knower Level.

The parameter estimates of this model provide a way to examine the influence of each NL1 Knower Level transition on NL2 Knower Level (see Table A1). If the β estimate for a particular NL1 Knower Level transition significantly deviates from 0, this indicates that this transition is associated with a change in NL2 Knower Level. In Give-A-Number Model 2, only the β estimates for the 4-knower and CP-knower transitions were statistically greater than 0 (Table A1). The parameter estimates from Give-A-Number Model 2 suggest that if Knower Level transfer occurs, it does so only at the 4-knower and CP-knower stages.

However, a follow-up analysis (Subset Knower Model 2 in the main text) that only included children who were subset knowers in *both* NL1 and NL2 calls into question the idea that knowledge of *four* transfers between languages, see Table A2. Among subset knowers, there is no evidence that 4-knowers in NL1 have a statistically higher NL2 Knower Level than 3-knowers.

This suggests that the statistically significant parameter estimate of 4-knowers in Give-A-Number Model 2 was driven primarily by children that are 4-knowers in their NL1 and CP-knowers in their NL2 ($n = 5$), who were excluded in Model 3, and not by children who were 4-knowers in both NL1 and NL2 ($n = 3$). This is important, because transfer predicts an exact match between a child's NL1 knower level and their NL2 knower level – not merely a higher knower level in the NL2. Consider, for example, transfer of a meaning like "three" from one language to another. This transfer could directly help a child learn *three* in their second language, but only indirectly help them learn *four* – i.e., not directly via transfer. Since by definition a child's NL2 was identified as the language in which they have weaker counting abilities (and thus by extension, less training), we should never predict their NL2 knower level to exceed their NL1 since additional NL2-specific training would be the only way for this to occur. This is borne out by the observation that NL1 KL was higher than NL2 KL in only 7% of participants. Finally, to the extent that transfer occurs from NL1 to NL2, we would also expect it to occur from NL2 to NL1. For these reasons, transfer should always predict an exact match in our study, not simply an accelerated NL2 trajectory. Supporting this analysis, for 1, 2 & 3-knowers there was a strong asymmetry such that children almost always had an NL2 knower level that was lower than or equal to their NL1 knower level (Table A3), whereas no such asymmetry was observed for 4-knowers. In fact, NL1 4-knowers were most likely to be NL2 CP-knowers; children who were NL1 4-knowers and NL2

Table A1
Model 2 parameter estimates for specific NL1 Knower Levels.*

		β	SE	χ^2	p
NL1 KL	1-knower	0.34	0.96	0.13	0.72
	2-knower	0.67	0.71	0.90	0.34
	3-knower	0.50	0.64	0.61	0.44
	4-knower	2.8	0.95	8.6	0.003
	CP-knower	2.1	0.87	5.9	0.015

* A parameter estimate that significantly deviates from 0 indicates that children at that NL1 Knower Level have a statistically higher NL2 Knower Level than children at the previous knower level, when other factors in the model (Age, NL1 Highest Count, NL2 Highest Count and Group) are considered.

Table A2
Model 3 parameter estimates for specific NL1 Knower Levels.

		β	SE	χ^2	p
NL1 KL	1-knower	0.42	0.96	0.20	0.66
	2-knower	0.58	0.72	0.66	0.42
	3-knower	0.92	0.68	1.8	0.18
	4-knower	1.79	1.23	2.2	0.15

Table A3
The number of children with an NL2 knower level that is lower, the same or higher than their NL1 knower level.

NL1 Knower Level	NL2 Knower Levels			Total
	Lower NL2 KL	Same NL2 KL	Higher NL2 KL	
Non-knowers	NA	8	2	10
1,2, & 3-knowers	25	21	5	51
4-knowers	1	3	5	9
CP-knowers	4	73	NA	77

CP-knowers ($n = 5$) were just as common as children who were NL1 CP-knowers and NL2 4-knowers ($n = 3$). See Table 2 in the main text.

Why might some children be CP-knowers in their NL2 but 4-knowers in their NL1? One plausible reason for this is measurement error unique to identifying CP-knowers. Anecdotally, we observed that many children from the low SES group exhibited knowledge of the counting routine, but had particular trouble executing the routine for higher numbers, perhaps due to difficulties with attention and working memory. These difficulties may have prevented a few children who actually understand the counting routine from reaching the criterion required for CP-knower classification. To support this idea, 18% of the low SES subset knowers were 4-knowers, whereas only 7.5% of the high SES subset knowers were 4-knowers in the current study.

Appendix B

B.1. Analysis of exact transfer in the Successor Task

An additional analysis provides evidence that, for "5" and "8", NL1 successor knowledge almost perfectly predicts NL2 successor knowledge, when chance responding is accounted for. First, consider performance for "5". Overall, 73.2% percent of children responded correctly on the NL1 "5" trial. From this, we can infer that 26.8% of children responded incorrectly by guessing. We can also infer that some of the children who responded correctly were also guessing. Since guessing is by definition a random response, we can approximate that an equal percentage of children would guess and respond correctly and the same percentage would guess and respond incorrectly. Given that 26.8% responded incorrectly by guessing, we can infer that approximately 26.8% responded correctly by guessing. This suggests that overall, approximately 46.4% (73.2–26.8%) of all the children had actual knowledge of the successor principle for 5 in their NL1. Thus, approximately only 63.4% of correct responders are likely to have true knowledge of the successor of "5" (46.4% who have actual knowledge over 73.2% who responded correctly). Under the hypothesis of 100% transfer from NL1 to NL2, we would predict that of the 63.4% who have knowledge, 100% should get the trial correct in their NL2. Of the estimated 36.6% of correct responders who performed correctly on the NL1 '5' trial by guessing, approximately half would be expected to perform correctly on the NL2 '5' trial. From this we can predict that approximately 81.7% of the children who performed correctly on the NL1 '5' trial would be expected to perform correctly on the NL2 '5' trial. In fact, 78.5% did. Using the same logic for "8", 76.7% of children who performed correctly on the "8" trial in their NL1 would be expected to perform correctly in their NL2. In fact, 72% did. These results provide strong evidence that knowledge of the successor principle in one language predicts almost identical knowledge in a second language.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.
Almoammer, A., Sullivan, J., Donlan, C., Marušič, F., Žaucer O'Donnell, T., & Barner, D. (2013). Grammatical morphology as a source of early number word meanings. *PNAS, 110*(46), 18448–18453.
Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13), 14.

Barner, D. (2012). Bootstrapping numeral meanings and the origin of exactness. *Language Learning and Development, 8*, 177–185.
Barner, D., & Bachrach, S. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology, 60*, 40–62.
Barner, D., Chow, K., & Yang, S. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology, 58*. 195-21.
Barner, D., Libenson, A., Cheung, P., & Takasaki, M. (2009b). Cross-linguistic relations between quantifiers and numerals in language acquisition: Evidence from Japanese. *Journal of Experimental Child Psychology, 103*, 421–440.
Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language, 24*, 511–533.
Brooks, N., Audet, J., & Barner, D. (2013). Pragmatic inference, not semantic competence, guides 3-year-olds' interpretation of unknown number words. *Developmental Psychology, 49*, 1066–1075.
Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus*, 59–68.
Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.
Cheung, P., Rubenson, M., & Barner, D. (2015). Numbers never end: A study on children's acquisition of the successor principle and its relation to the concept of infinity. In *Proceedings from the society of research in child development*. Philadelphia, PA.
Corbett, G. G. (2000). *Number*. Cambridge, UK: Cambridge University Press.
Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123*, 162–173.
Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43*(1), 1–29.
Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking behavioral and brain-imaging evidence. *Science, 284*, 970–974.
Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Developmental Science, 11*, 650–654.
Frye, D., Briasby, N., Lowe, J., Maroudas, D., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development, 60*, 1158–1171.
Fuson, K. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.
Fuson, K., Richards, J., & Briars, D. (1982). The acquisition and elaboration of the number word sequence. In C. Brainerd (Ed.), *Progress in cognitive development research. Children's logical and mathematical cognition* (Vol. 1, pp. 33–92). New York: Springer Verlag.
Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, Massachusetts: Harvard University Press.
Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*, 496–499.
Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: Relation between parents' input and children's number knowledge. *Developmental Science, 14*(5), 1021–1032.
Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The approximate number system in 3-, 4-, 5-, 6-year-olds and adults. *Developmental Psychology, 44*, 1457–1465.
Klibanoff, R., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. (2006). Preschool children's mathematical knowledge: The effect of teacher "math talk". *Developmental Psychology, 42*, 59–69.
Kolers, P. (1968). Bilinguals and information processing. *Scientific American, 218*, 78–86.
Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*, 395–438.
Le Corre, M., Van de Walle, G., Brannon, E., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology, 52*, 130–169.
Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behavior in number-concept development. *Cognitive Science, 34*, 51–67.
Lee, M. D., & Sarnecka, B. W. (2011). Number-knower levels in young children: Insights from Bayesian modeling. *Cognition, 120*, 391–402.
Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences, 12*, 213–218.
Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge? *Developmental Psychology, 46*(5), 1309–1319.
Li, P., Le Corre, M., Shui, R., Jia, G., & Carey, S. (2003). A crosslinguistic study of the role of singular-plural in number word learning. Paper presented at the 28th annual Boston University conference on language development. Boston, MA.
Marsh, L. M., & Maki, R. H. (1976). Efficiency of arithmetic operations in bilinguals as a function of language. *Memory & Cognition, 4*, 459–464.
McClain, L., & Huang, J. Y. (1982). Speed of simple arithmetic in bilinguals. *Memory & Cognition, 10*, 591–596.
Piantadosi, S. T., Jara-Ettinger, J., & Gibson, E. (2014). Children's development of number in an indigenous farming-foraging group. *Developmental Science, 17*, 553–563.
Piantadosi, S., Tenenbaum, J., & Goodman, N. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition, 123*, 199–217.
Pica, P., Lemer, V., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian Indigene Group. *Science, 206*, 499–503.
Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*, 662–674.
Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition, 92*, 329–352.
Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, J. B. (2007). From grammatical number to exact numbers: Early meanings of "one," "two", and "three" in English, Russian, and Japanese. *Cognitive Psychology, 55*, 136–168.
Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge in early childhood. *Journal of Experimental Child Psychology, 103*, 325–337.
Schaeffer, B., Eggleston, V., & Scott, J. (1974). Number development in young children. *Cognitive Psychology, 6*, 357–379.
Song, M. J., & Ginsburg, H. P. (1988). The effect of the Korean number system on young children's counting: A natural experiment in Numerical Bilingualism. *International Journal of Psychology, 23*, 319–332.

Spelke, E. S., & Tsivkin, S. (2001b). Language and number: A bilingual training study. *Cognition, 78*, 45–88.
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*. Cambridge, UK: Cambridge University Press.
Sullivan, J., & Barner, D. (2012). How are number words mapped to approximate magnitudes? *Quarterly Journal of Experimental Psychology, 66*, 389–402.
Sullivan, J., & Barner, D. (2014a). Inference and association in children's early numerical estimation. *Child Development, 85*, 1740–1755.
Sullivan, J., & Barner, D. (2014b). The development of structural analogy in number-line estimation. *Journal of Experimental Child Psychology, 128*, 171–189.
Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in children: The "Catch the Spies" task. *Cognitive Development, 20*(3), 373–387.
Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*, 155–193.
Wynn, K. (1992). Child's acquisition of the number words and the counting system. *Cognitive Psychology, 24*, 220–251.