

Gesture as a window onto children's number knowledge



Elizabeth A. Gunderson^{a,*}, Elizabeth Spaepen^b, Dominic Gibson^b, Susan Goldin-Meadow^b,
Susan C. Levine^b

^a Temple University, 1701 North 13th St., Philadelphia, PA 19122, United States

^b University of Chicago, 5848 S. University Ave., Chicago, IL 60637, United States

ARTICLE INFO

Article history:

Received 12 March 2014

Revised 2 April 2015

Accepted 13 July 2015

Keywords:

Cardinality

Number words

Gesture

Gesture–speech mismatch

Approximation

ABSTRACT

Before learning the cardinal principle (knowing that the last word reached when counting a set represents the size of the whole set), children do not use number words accurately to label most set sizes. However, it remains unclear whether this difficulty reflects a general inability to conceptualize and communicate about number, or a specific problem with number words. We hypothesized that children's gestures might reflect knowledge of number concepts that they cannot yet express in speech, particularly for numbers they do not use accurately in speech (numbers above their knower-level). Number gestures are iconic in the sense that they are item-based (i.e., each finger maps onto one item in a set) and therefore may be easier to map onto sets of objects than number words, whose forms do not map transparently onto the number of items in a set and, in this sense, are arbitrary. In addition, learners in transition with respect to a concept often produce gestures that convey different information than the accompanying speech. We examined the number words and gestures 3- to 5-year-olds used to label small set sizes exactly (1–4) and larger set sizes approximately (5–10). Children who had not yet learned the cardinal principle were more than twice as accurate when labeling sets of 2 and 3 items with gestures than with words, particularly if the values were above their knower-level. They were also better at approximating set sizes 5–10 with gestures than with words. Further, gesture was more accurate when it differed from the accompanying speech (i.e., a gesture–speech mismatch). These results show that children convey numerical information in gesture that they cannot yet convey in speech, and raise the possibility that number gestures play a functional role in children's development of number concepts.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Children's understanding of number – both exact and approximate – is a critical component of early mathematical development (e.g., Carey, 2009; National Research Council, 2009; Sarnecka & Carey, 2008; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013). Previous research has shown that children learn the cardinal meanings of the first four number words one at a time, in order, and then learn the cardinal principle – that the last number reached when counting a set represents the size of the whole set (e.g., Carey, 2009; Sarnecka & Carey, 2008; Sarnecka & Lee, 2009; Wynn, 1990, 1992). Although this trajectory has been well documented, to understand the mechanism underlying the acquisition of the cardinal principle, it is important to characterize children's skills and limitations during the early stages of the trajectory when they know the cardinal meanings of only a few number words. Are

children at these early stages able to conceptualize, and communicate about, set sizes higher than those they can label with their known number words? We investigated this question by asking children to communicate about set sizes using an alternate modality – gesture – and comparing their ability to respond appropriately using number gestures versus number words. We examine children's ability to use both words and gestures to label small set sizes exactly (1–4), as well as to label larger set sizes approximately (5–10). We argue that assessing children's use of number gestures provides a more fine-grained view of children's numerical cognition than can be obtained using verbal measures alone.

1.1. Cardinal number knowledge

Although children are often able to count a set of objects by age 2½ years, they do not initially understand the relation between the count routine and the cardinal value of the set (Wynn, 1990, 1992). Rather, children appear to go through stages in which they learn the cardinal meanings of each of the first four number words one at a time, in order (e.g., Carey, 2009; Sarnecka & Lee, 2009;

* Corresponding author at: Department of Psychology, Temple University, 1701 North 13th St., Philadelphia, PA 19122, United States.

E-mail address: liz.gunderson@temple.edu (E.A. Gunderson).

Wynn, 1992). A child who has learned the cardinal meaning of “one” (referred to as a “one-knower”) can successfully produce and label sets of one object, but not larger set sizes. Similarly, a child who has learned the cardinal meanings of the words “one” and “two” (referred to as a “two-knower”) can successfully produce and label sets of one or two objects, but not larger set sizes. Children go through parallel stages in which they learn the cardinal meanings of “three” and “four” (referred to as “three-knowers” and “four-knowers”). After learning the cardinal meanings of the first three or four number words, children learn that the last number word reached when counting a set represents the size of the whole set (the cardinal principle) and are referred to as “cardinal-principle-knowers” (Gelman & Gallistel, 1978). This process takes at least 12 months, and may take up to 3 years for some children, from the time they learn the cardinal meaning of “one” to the time they learn the cardinal principle (e.g., Sarnecka, Goldman, & Slusser, 2015; Wynn, 1992).

This developmental trajectory has typically been described as discontinuous (e.g., Sarnecka & Lee, 2009), such that children at a particular “knower-level” have little to no knowledge of the cardinal meanings of numbers above their highest “known” number. However, recent research suggests otherwise. Children at a given knower-level seem to have some knowledge of the next number word above their knower-level (Barner & Bachrach, 2010). For instance, two-knowers tend to say “three” when shown 3 objects more often than when shown 4 objects. Further, at least some children who have not yet learned the cardinal principle (collectively referred to as “subset-knowers”) say larger number words when shown larger sets of objects (Gunderson, Spaepen, & Levine, 2015), indicating that they already have some approximate knowledge of the cardinal meanings of larger number words. These children thus demonstrate partial knowledge of the cardinal meanings of number words above their knower-level.

1.2. Gesture in numerical development

Gestures are a means of communicating numerical knowledge, one that children carry around with them at all times, literally, at their fingertips. A number of researchers have argued that gestures support verbal number knowledge (for recent reviews, see Di Luca & Pesenti, 2011; Goldin-Meadow, Levine, & Jacobs, 2014; but see Crollen, Seron, & Noël, 2011 for a contrasting view). However, most research has focused on counting gestures (i.e., pointing to objects or raising fingers one at a time while counting; Alibali & DiRusso, 1999; Fuson, 1988; Gelman & Gallistel, 1978; Graham, 1999; Potter & Levy, 1968; Saxe, 1977; Saxe & Kaplan, 1981), rather than cardinal number gestures (i.e., holding up a certain number of fingers to represent a set size, a behavior that has been called “finger montring”; Di Luca & Pesenti, 2008). Children begin to produce pointing gestures while counting as early as age 2 years (Gelman & Gallistel, 1978) and, by age 4 years, almost always do so, while at the same time successfully preserving one-to-one correspondence (Saxe, 1977). Counting gestures have been posited to support the counting principles, which are thought, in turn, to support later numerical development (e.g., Gelman & Gallistel, 1978). Consistent with this argument, finger gnosis (the ability to represent one’s fingers mentally) is correlated with math skill (e.g., Fayol, Barrouillet, & Marinthe, 1998; Noël, 2005). These studies suggest that counting gestures are extremely common and potentially important for children’s numerical development.

Cardinal number gestures (referred to simply as “number gestures” in subsequent mentions) have been much less well-studied, despite their ubiquity across cultures (Bender & Beller, 2012). Most of the existing research has focused on adults’ use of these gestures (e.g., Di Luca, Lefèvre, & Pesenti, 2010; Di

Luca & Pesenti, 2008; Spaepen, Coppola, Flaherty, Spelke, & Goldin-Meadow, 2013; Spaepen, Coppola, Spelke, Carey, & Goldin-Meadow, 2011; Suriyakham, 2007). Both adults and 2nd-graders are faster to name the number represented by canonical number gestures (e.g., index and middle finger raised to mean “two”) than by other finger configurations (e.g., index and pinky finger raised to mean “two”) (Di Luca & Pesenti, 2008; Noël, 2005). For adults, canonical number gestures, but not non-canonical ones, prime nearby numbers based on their proximity to the target in the same way that Arabic numerals do (i.e., a place-coding representation) (Di Luca et al., 2010). This finding suggests that adults’ number gestures function symbolically and possess semantic meaning based on their canonical configuration, rather than deriving their meaning merely from the number of fingers (and not the particular fingers) raised.

Interestingly, number gestures may have different meanings when used by numerate versus non-numerate adults (Spaepen et al., 2011, 2013). Profoundly deaf adults who have not learned a sign language or a spoken language and use their own homemade gestures to communicate (known as “homesigners,” Goldin-Meadow, 2003a), and who have therefore never learned a count list or any other formal mathematics, readily use number gestures to communicate about set size (Coppola, Spaepen, & Goldin-Meadow, 2013; Spaepen et al., 2011). When asked to produce a gesture to represent a set of objects, homesigners’ number gestures are accurate up to set sizes of 3, but approximate for set sizes 4 and higher (Spaepen et al., 2011). This pattern suggests that their gestures map onto the same two preverbal systems thought to underlie children’s number word learning – the parallel-individuation system, which represents set sizes 1–3 in an exact manner, and the approximate number system (ANS), which represents larger set sizes in an approximate manner (e.g., Feigenson, Dehaene, & Spelke, 2004).

Importantly, homesigners seem to use all of their number gestures as item-based representations, rather than as summary symbols. When asked to recall sequences of number gestures, homesigners perform more poorly on sequences involving large set sizes (e.g., 4, 5, 4) than on sequences involving small set sizes (e.g., 2, 3, 2) (Spaepen et al., 2013). This pattern suggests that, for the homesigners, the individual fingers in a number display occupy multiple spaces in short-term memory (“one, one, one, one, one”), as opposed to occupying a single space as a summary symbol (“five”); it is harder to remember displays containing many fingers (4 or 5) than displays containing fewer fingers (2 or 3). In contrast, deaf individuals who have learned a conventional sign language (e.g., American Sign Language), including a count list, perform equally well on sequences containing 4’s and 5’s and sequences containing 2’s and 3’s (Spaepen et al., 2013); for these signers, the number handshapes do not appear to be item-based (“one, one, one, one, one”) and thus are likely to be serving as summary symbols (“five”).

Do children, before learning the cardinal principle, view their number gestures as summary symbols, like deaf signers and hearing adults, or as item-based representations, like homesigners? Although the present study was not designed to directly answer this question, we hypothesize that children initially use their gestures as item-based representations, like homesigners, and that gestures serve as an important bridge between preverbal mental representations of number that are item-based (i.e., the parallel-individuation system for small sets) or magnitude-based (i.e., the ANS for large sets) and the arbitrary symbols of speech. We return to this possibility in Section 4. The present study focuses on the critical question that must be answered first: can young children use number gestures to effectively communicate about set sizes that are larger than their known number words?

Although children and their parents spontaneously produce number gestures as early as 3 years of age (Goldin-Meadow et al., 2014; Suriyakham, 2007), there has been relatively little research on how children use their number gestures, which is critical to understanding the role gesture plays in early numerical development (Crollen, Mahe, Collignon, & Seron, 2011; Nicoladis, Pika, & Marentette, 2010). Most relevant to the current research, a study comparing children's performance in speech and gesture concluded that children perform better on verbal than gestural number tasks (Nicoladis et al., 2010). This study used gestural and verbal versions of two tasks. In one task (How Many), 2- to 5-year-olds were shown a picture of a set of objects and were asked to respond using a number gesture or a number word. In a second task (Give-a-Number), children were presented with either a number gesture or a number word and were asked to respond by creating a set of objects corresponding to the number. The results indicated that children were more accurate with number words than with number gestures on both the How Many and Give-a-Number tasks. However, there is reason to believe that these results do not indicate a universal advantage for number words over number gestures in development. First, the observed difference was driven mainly by children's higher accuracy with words than with gestures for the set sizes 6, 7, 8, and 9; these set sizes are problematic because they require children to produce different handshapes with each hand, a difficult task at this age. Second, the advantage of speech over gesture was more robust among older children (4- to 5-year-olds) than among younger children (2- to 3-year-olds). Older children, who are more likely to have already learned the cardinal principle and thus be at or near ceiling on the number word tasks, may show better performance in speech than in gesture because speech is the more practiced modality. But this pattern may not reflect the abilities of subset-knowers, who have not yet learned the cardinal principle.

For subset-knowers, the literature provides reason to believe that gesture may lead speech. In particular, gesture studies examining other concepts show that learners who are in transition with respect to a concept, when asked to talk about the concept, often produce gestures that convey information that differs from the information conveyed in the accompanying speech (Perry, Church, & Goldin-Meadow, 1988; Pine, Lufkin, & Messer, 2004), and that the information conveyed in those gestures is often more correct than the information conveyed in speech (Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; Church & Goldin-Meadow, 1986). Thus, we propose that children's ability to use number gestures may actually precede their ability to use number words, and that this gesture advantage will be observed in subset-knowers for whom cardinal-number knowledge is in a transitional stage, but not in cardinal-principle-knowers for whom cardinal-number knowledge is more fully-developed.

1.3. The present study

The present study uses number gesture as a window onto children's ability to represent and communicate about set sizes, with the goal of addressing two interrelated sets of questions about preschoolers' number knowledge prior to learning the cardinal principle.

Our first question revolves around what children know about small set sizes (1, 2, 3, and possibly 4). Given that even infants can exactly represent and discriminate between set sizes 1 to 3 (and possibly 4, for older children) using the parallel-individuation system (e.g., Feigenson & Carey, 2003; Feigenson et al., 2004; Starkey & Cooper, 1980; Starkey, Spelke, & Gelman, 1990; Strauss & Curtis, 1981), what holds children back from learning the cardinal meanings of the number words "one", "two", and "three"? Do children have difficulty explicitly accessing

and communicating about their mental representations of set sizes 1, 2, and 3, or is it the case that children can access these representations and communicate about them, but have difficulty mapping specific number words to these representations? Examining number gestures in relation to number words allows us to distinguish between these two possibilities. We hypothesize that subset-knowers will be better able to access and communicate about their mental representations of set size using gestures than using words, since number gestures bear the same item-based correspondence to set size as the child's mental representation. Number words are arbitrary symbols, which are likely to be more difficult to map to item-based representations than number gestures.

In the present study, we tested this hypothesis by showing children pictures of objects and asking them to indicate what was on the page using either number words or number gestures. We expected children to be more accurate in gesture than in speech, but only for those set sizes that were small enough to be represented by the parallel-individuation system (i.e., 1, 2, 3, and possibly 4) and for which they had not yet learned an exact number word (i.e., numbers above their knower-level). We expected this to be the case for several reasons. First, gesture precedes speech in other aspects of language development, including spontaneous vocabulary and multi-word utterance production (Goldin-Meadow, 2014; Iverson & Goldin-Meadow, 2005; Özçalışkan & Goldin-Meadow, 2005), so it seemed possible that gesture would precede speech in the context of number word learning as well. Second, we reasoned that children might be able to gesture about cardinal number prior to being able to produce number words correctly because of the item-based nature of number gestures. Children are able to match small sets of objects (1–3) to other small sets of similar objects based on set size around age 2, earlier than they can correctly produce words for those sets of objects (Huttenlocher, Jordan, & Levine, 1994; Mix, 2008). If children see their fingers as sets of objects (as homesigners seem to do; Spaepen et al., 2013), they may be able to match the number of fingers they hold up to the number of objects in a set in a similar way, even before learning a verbal label for the entire set (a summary symbol). Given that number gestures eventually function as summary symbols for numerate adults (Di Luca & Pesenti, 2008), it is possible that number gestures may help children transition from the item-based representations of the parallel-individuation system to the summary symbols necessary for learning the cardinal meanings of number words; we return to this possibility in greater detail in Section 4.

Our second question pertains to children's ability to understand large set sizes (5–10), which are too large to be represented by the parallel-individuation system; the ANS represents them in a noisy fashion (e.g., Feigenson et al., 2004). Many subset-knowers fail to say higher number words to label sets of 10 objects than sets of 5 objects (Le Corre & Carey, 2007). Do subset-knowers know more about large set sizes than their number words suggest? As with smaller sets, children might have approximate mental representations of large set sizes (5–10), and even be able to explicitly access and communicate about those representations, but have difficulty mapping specific number words to them. If so, we would expect children to map number gestures (but not yet number words) to the Approximate Number System (ANS). Although the ANS is not thought to track objects in an item-based fashion as the parallel-individuation system does (e.g., Feigenson et al., 2004), ANS representations are deeply intertwined with spatial extent (e.g., Cantrell & Smith, 2013). Spatial extent is encoded in number gestures simply because they are item-based (i.e., a larger number gesture involves more fingers, which take up more space) but is not encoded in arbitrary number words (i.e., a larger number word does not take up more space or time). Thus, the item-based nature

of number gestures may make them easier to map onto the representations of the ANS.

To determine whether subset-knowers show more knowledge of large set sizes in gesture than speech, we asked children to produce number words and number gestures in response to large set sizes (5–10). We expected that subset-knowers would fail to produce larger number words for larger set sizes, but would succeed with number gestures. Given that even adult homesigners do not create exact matches between their fingers and sets sizes larger than 4 (Spaepen et al., 2011), we did not expect subset-knowers to do so. However, since children are able to match larger set sizes to each other in an approximate fashion starting around 3 years of age (Cantlon, Fink, Safford, & Brannon, 2007), they should be able to match their gestures to larger set sizes in an approximate fashion as well. This prediction is based on the assumption that numerical gestures, because they are item-based with a transparent mapping between number of fingers and number of items in a set, make it easier for children to map a set onto a number of fingers than to a number word, a mapping that may eventually serve as a bridge to arbitrary symbolic representations of number (i.e., to number words).

Since we asked children to respond with number words and number gestures separately, it is possible that any differences found between children's performance on the two tasks may be due to differences in children's interpretation of the task (e.g., whether the experimenter is asking for an exact or approximate response) or to children changing their strategy between the two tasks. Therefore, in addition to examining performance on each task separately, we also examined instances in which children spontaneously produced both gesture and speech in a single response. Since gesture and speech form an integrated system (Goldin-Meadow, 1998, 2003b), we expected children to produce some gesture–speech combinations spontaneously during the numerical tasks. Of particular interest are instances in which children produce a different response in gesture than in speech, referred to as gesture–speech mismatches (Church & Goldin-Meadow, 1986; Goldin-Meadow, 2003b). Examining the numerical values of the gesture and speech components of a mismatch allows us to isolate differences between children's representational capacities in each modality, while equating for strategy use and task interpretation. In line with our predictions for the tasks as a whole, we expected that children, particularly subset-knowers, would display better performance in the gestural component of a mismatch than in the speech component.

2. Method

2.1. Participants

One hundred fifty-five children (85 male, 70 female) participated in the study. The mean age was 4.44 years ($SD = 0.60$, range = 3.11–5.58 years, $N = 154$ since one parent did not provide the child's date of birth). Subjects came from a range of socioeconomic backgrounds. Average family income was \$42,868 ($SD = \$37,406$) with a range from less than \$15,000 to more than \$100,000 per year ($N = 129$). Parents' education was considered to be the maximum education of either parent, and ranged from less than high school to a graduate degree. The average parents' education was 14.2 years (equivalent to 2 years of college; $SD = 2.6$ years; $N = 136$).

An additional 38 children were initially assessed, but were considered ineligible for the study because they did not complete all tasks ($N = 23$), were identified by the teacher as having a hearing disability or developmental delay ($N = 2$), were judged by the experimenters to not have understood the tasks due to a speech

delay ($N = 2$), had unclassifiable knower-level data ($N = 7$),¹ or were classified as “five-knowers” and thus could not be categorized definitively as subset- or CP-knowers ($N = 4$).

2.2. Design

Children completed three tasks as part of a larger battery of math-related tasks. Children were assessed in a quiet area of their preschool. The Give-a-Number task was used to determine children's knower-levels, and always preceded the other two tasks (although it was in some cases separated from the other two tasks by other randomly-ordered tasks in the task battery). Children completed a number gesture elicitation task, What's on this Card-Gesture (WOC-Gesture), and a number word elicitation task, What's on this Card-Speech (WOC-Speech). We chose a fixed order, in which WOC-Gesture always preceded WOC-Speech, as a way to increase the number of gesture–speech combinations that children would produce on the WOC-Speech task. That is, by asking children to gesture first, we expected that children would continue to gesture during the subsequent speech task. We made the decision part-way through data collection that the WOC-Speech task should always *immediately* follow the WOC-Gesture task (earlier in data collection, the two tasks could be separated by other tasks, depending on the random order of tasks in the battery) in the hope of increasing spontaneously produced gesture–speech combinations on the WOC-Speech task. The majority of children (73%) received the WOC-Speech task immediately after the WOC-Gesture task; for the remaining children the tasks were separated by one or more other numerical tasks.²

2.3. Materials

2.3.1. Give-a-Number

The materials for this task consisted of fifteen small multi-colored plastic fish and a clear plastic bowl.

2.3.2. What's on this Card-Gesture (WOC-Gesture)

The materials for the WOC-Gesture task consisted of twenty-four 8.5" × 11" white sheets of paper with color pictures of objects printed on them. The sheets of paper were encased in clear plastic sheet protectors in a three-ring binder. Children were shown four blocks of 6 cards, for a total of 24 cards. Each block depicted one type of object (frogs, birds, flowers, and boats), and each card within the block depicted a different set size (1, 2, 3, 4, 5, and 10 objects). We chose to include 5 and 10 as our large-number set sizes because the gestures for 5 and 10 do not require children to produce different hand-shapes with each hand, a difficult motoric task for preschoolers (the gestures for 6, 7, 8, and

¹ Eleven children included in the sample had knower-level data that was incomplete (i.e., missing one or more trials due to child refusal or experimenter error) but was still considered classifiable. To check the robustness of our results, we re-ran all analyses excluding these children. All results reported as significant at $p < .05$ remained so at $p < .10$ with the following exceptions: subset-knowers' gesture advantage on 3 during mismatches did not reach significance after the Bonferroni correction using the more conservative binomial test ($p = .029$), although it did remain significant using a one-sample t -test ($p = .002$); the correlation between knower-level and large-number slope on the WOC-Gesture task was no longer significant, $r(58) = .19$, $p = .15$; and there were no significant differences in WOC-Gesture versus WOC-Speech for one-knowers ($N = 3$) on set sizes 3 and 9 (Appendix Fig. A1).

² To check the robustness of our results, we re-ran all analyses including only children who received the WOC-Speech task immediately after the WOC-Gesture task ($N = 113$). All results reported as significant at $p < .05$ remained so or were marginally significant at $p < .10$ with the following exceptions: the correlation between knower-level and large-number slope on the WOC-Gesture task was no longer significant, $r(58) = .18$, $p = .18$, and there was no significant difference in four-knowers' accuracy on set size 5 on WOC-Gesture versus set size 6 on WOC-Speech (Appendix Fig. A1).

9 do require coordinating different hand-shapes). In addition, 5 and 10 differ by a 1:2 ratio that is visually discriminable by 3-year-olds in a non-symbolic approximation task (Halberda & Feigenson, 2008).

Within each block, the objects depicted were identical in size, shape, and color. The first block (frogs) was a practice block intended to familiarize children with the gesture response modality, and the set sizes for this block were presented in numerical order (i.e., 1, 2, 3, 4, 5, 10). The set sizes in the three test blocks (birds, flowers, and boats) were presented in a fixed, pseudo-random order, with 6 items per block, for a total of 18 items for analysis.

2.3.3. What's on this Card-Speech (WOC-Speech)

The materials for the WOC-Speech task were nearly identical in structure to those for the WOC-Gesture task. The WOC-Speech task consisted of eighteen 8.5" × 11" sheets of paper presented in clear plastic sheet protectors in a three-ring binder. The cards depicted sets of objects arranged in three blocks of 6 cards per block (18 items total). Within each block, the objects (soccer balls, fish, and bananas) were identical in size, shape, and color. The 6 cards depicted different set sizes (1, 2, 3, 4, 6, and 9 objects). We chose to include 6 and 9 as our large-number set sizes on the WOC-Speech task because we wanted to avoid the possibility that children would answer correctly by guessing the frequently-used number words "five" and "ten." Although the large-number set sizes were not identical on the WOC-Speech (6 and 9) and WOC-Gesture tasks (5 and 10), it is important to note that we also analyze children's simultaneous gesture and speech responses *within a single trial*, mitigating the possibility that these set size differences are responsible for differences in children's response patterns across the tasks.

On the WOC-Speech task, the large-number set sizes, 6 and 9, differ by a 2:3 ratio that is visually discriminable by 3-year-olds in a non-symbolic approximation task (Halberda & Feigenson, 2008). The first item (1 soccer ball) was a practice item, which was not scored or analyzed, resulting in 17 items for analysis. The remaining set sizes in the three test blocks were presented in a fixed, pseudo-random order.

2.4. Procedure

2.4.1. Give-a-Number

In the Give-a-Number task, children were given a pile of 15 fish and were asked to put a certain number of fish in the "pond" (a clear plastic bowl) (Wynn, 1990). Following each incorrect response, the experimenter gave the child the chance to correct his or her answer by saying, "Let's check. Can you count the fish?" After the child counted the fish, the experimenter said, "But I asked for N fish! Can you put N fish in the pond?" The child's second answer was recorded.

The experimenter began by asking for 1 fish. Subsequent trials were determined using the titration method modeled after Wynn (1992). When a child succeeded in giving N fish, the experimenter requested $N + 1$ fish. When a child did not succeed in giving N fish, the experimenter requested $N - 1$ fish. The experimenter ended the session when the child gave N fish correctly at least two out of three times and gave $N + 1$ incorrectly at least two out of three times, or gave all set sizes correctly up to 6, and gave 6 correctly 2 out of 3 times.

2.4.2. WOC-Gesture

The WOC-Gesture task began with a two-part familiarization phase. In the first part, the experimenter asked the child to copy her gestures without reference to any pictures or objects. The goal of this copying procedure was to ensure that the child was able to

produce a unique, numerically-correct gesture for each set size. The experimenter produced the number gestures for 1, 2, 3, 4, 5, and 10, one at a time, in order.³ Between each gesture, the experimenter waited for the child to copy the gesture. If the child did not copy the gesture correctly (typically when the child produced an incorrect number of fingers), the experimenter provided help and correction until the child could produce a unique gesture for each number. In some cases this process involved the child producing a different number gesture from the one the experimenter showed initially. For example, some children who had difficulty producing the originally-presented gesture for 3 (index, middle, and ring fingers) were coached to use a different gesture for 3 (e.g., middle, ring, and pinky fingers). In some cases, the experimenter suggested that the child use his or her other hand to hold down fingers in order to make the correct gesture (e.g., the child could make a gesture for 3 by using the left hand to hold down the thumb and pinky finger of the right hand, leaving the index, middle, and ring fingers up).

In the second part of the familiarization phase, the experimenter modeled the use of each gesture in connection with pictures of the appropriate set size. The experimenter showed the child the first card (a picture of 1 frog) and said, "For this, I would do this [*holds up index finger*]. Can you do that?" The experimenter provided correction and coaching if necessary. This was repeated for each set size in the practice block (set sizes 1, 2, 3, 4, 5, and 10, in order).

To begin the test phase, the experimenter showed the child the first picture in the test block (5 birds) and said, "Now it's your turn. What would you do for this card?" If the child did not respond, the experimenter attempted to elicit a response by saying, "Can you use your fingers to show me what's on this card?" or by referring back to the first practice card (1 frog) and saying, "Remember, for this I would do this [*holds up index finger*]. What would you do for this card?" The experimenter repeated this procedure for each trial in the test phase (3 blocks of six trials each for a total of 18 trials). Children's gesture responses were recorded as the number of fingers they held up, whether or not the particular configuration of fingers was the one that had been modeled for them.

2.4.3. WOC-Speech

In the first trial of the WOC-Speech task, the experimenter showed the child a picture of one soccer ball and said, "What's on the card?" After the child responded, typically by saying "ball" or "a ball", the experimenter said, "That's right, it's ONE ball." Children's responses on this first practice item were not scored. On subsequent items, the experimenter said, "What's on the card?" If the child responded with a gesture but not a number word, the experimenter said, "Can you use your words to tell me what's on here?" If the child counted but did not produce a cardinal response, the experimenter said, "So what's on the card?" If necessary, the experimenter attempted to elicit a cardinal number word response by saying, "What else can you tell me?" "Can you take a guess?", or by referring back to the first card and saying "Remember, this is ONE ball. So what's on this card?" This procedure was repeated for each trial (17 test trials).

³ A potential concern is that the experimenter demonstrated exactly six number gestures, which may have limited children's use of number gestures to 1, 2, 3, 4, 5, and 10, whereas children's use of number words was in principle limited only by their knowledge of the count list (for most children, at least 1–10). To determine whether the limited set of number gestures affected our results, we re-ran our analyses of children's responses to large set sizes (5–10) on WOC-Gesture versus WOC-Speech using this restricted sample of speech data (responses of 1, 2, 3, 4, 5, and 10), parallel to the available number gestures. All results reported as statistically significant remained so.

3. Results

3.1. Knower-levels

We determined children's knower-levels based on their responses on the Give-a-Number task, following the criteria established by Wynn (1992). Specifically, we categorized children as knowing the number N if they responded correctly two out of three times when asked for N objects, and gave N objects no more than half as often (percentage wise) when asked for larger set sizes than when asked for N itself. Each child's knower-level was the highest number for which the child's performance met these criteria. Children who were successful up to the set size 6 were considered cardinal-principle-knowers (CP-knowers). As noted previously, 4 children who were successful up to the set size 5, but not 6, were excluded because they could not definitively be categorized as either subset- or CP-knowers. The sample size and age, by knower-level, is listed in Table 1.

3.2. Small numbers (1–4)

3.2.1. Accuracy on gesture versus speech tasks

We first examined children's accuracy on the WOC-Speech and WOC-Gesture tasks for small numbers, in the range that can be represented by the parallel-individuation system (1–4). Our first prediction was that subset-knowers would be more accurate overall on the WOC-Gesture task than on the WOC-Speech task, but that this would not be the case for CP-knowers. To test this hypothesis, we conducted a 3-way mixed-effects ANOVA on accuracy with the Greenhouse–Geisser correction, with the within-subjects factors of task (WOC-Gesture vs. WOC-Speech) and set size (1, 2, 3, and 4), and the between-subjects factor of CP-knowledge group (subset-knowers vs. CP-knowers). All main effects and interactions were statistically significant ($p < .01$), including the 3-way interaction of task, set size, and CP-knowledge group, $F(2.3, 357.4) = 4.47$, $p < .01$, $\eta_p^2 = .03$. To interpret this three-way interaction, we next examined the two-way interactions between task and set size within subset-knowers and within CP-knowers separately.

Accuracy by task and set size for subset-knowers is depicted in Fig. 1, Panel A (additional graphs of accuracy, average response, and histograms of response patterns by knower-level are presented in Appendix Figs. A1–A3). A 2 (task: WOC-Gesture vs. WOC-Speech) \times 4 (set size: 1, 2, 3, 4) repeated-measures ANOVA with the Greenhouse–Geisser correction revealed significant main effects of task, $F(1, 72) = 21.2$, $p < .001$, $\eta_p^2 = .23$, and set size, $F(2.6, 185.9) = 135.6$, $p < .001$, $\eta_p^2 = .65$, and a significant task \times set size interaction, $F(2.5, 177.4) = 9.87$, $p < .001$, $\eta_p^2 = .12$. Subset-knowers were significantly more accurate on the WOC-Gesture task than on the WOC-Speech task for set sizes 2 ($t(72) = 5.03$, $p < .001$, $d = 0.59$; sign test: $Z = -4.20$, $p < .001$) and 3 ($t(72) = 4.12$, $p < .001$, $d = 0.48$; sign test: $Z = -4.08$, $p < .001$),

Table 1

Age of children by knower-level ($N = 155$).

Knower-level	N	Age	
		Mean (SD)	Range
Pre-knowers	9 ^a	4.2 (0.9)	3.3–5.5
One-knowers	9	3.9 (0.3)	3.4–4.4
Two-knowers	25	4.3 (0.6)	3.2–5.6
Three-knowers	17	4.1 (0.5)	3.1–4.9
Four-knowers	13	4.4 (0.6)	3.5–5.3
Cardinal-principle-knowers	82	4.6 (0.5)	3.4–5.6
Total	155	4.4 (0.6)	3.1–5.6

^a There were 9 pre-knowers in our sample, but age data was only available for 8 of them.

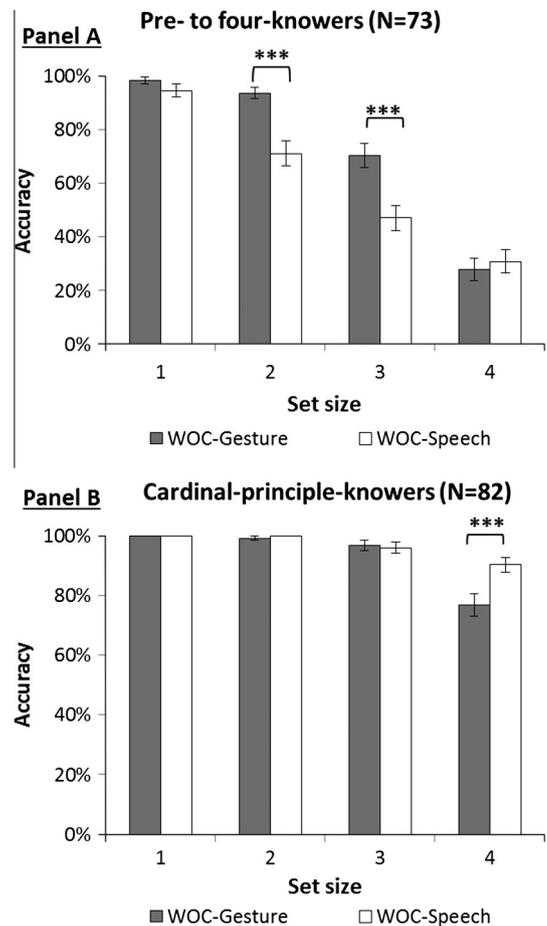


Fig. 1. Accuracy on the WOC-Gesture and WOC-Speech tasks by set size. *** $p < .001$.

but not for set size 1 where they were near ceiling on both the gesture and speech tasks ($t(72) = 1.39$, $p = .17$, $d = 0.16$; sign test: exact $p = 0.69$), nor for set size 4, ($t(72) = -.66$, $p = .51$, $d = 0.08$; sign test: $Z = -0.39$, $p = 0.74$).

For all CP-knowers, accuracy by task and set size is depicted in Fig. 1, Panel B. A 2 (task: WOC-Gesture vs. WOC-Speech) \times 4 (set size: 1, 2, 3, 4) repeated-measures ANOVA with the Greenhouse–Geisser correction revealed significant main effects of task, $F(1, 81) = 8.84$, $p < .01$, $\eta_p^2 = .10$, and set size, $F(1.6, 127.2) = 29.5$, $p < .001$, $\eta_p^2 = .27$, and a significant task \times set size interaction, $F(1.4, 116.9) = 11.1$, $p < .001$, $\eta_p^2 = .12$. CP-knowers were significantly more accurate on the WOC-Speech task than on the WOC-Gesture task for set size 4 ($t(81) = 3.77$, $p < .001$, $d = .42$; sign test: $Z = -3.95$, $p < .001$) but not set sizes 1, 2, or 3 (t -tests: all $ps > .30$; sign tests: all $ps > 0.30$), where CP-knowers were near ceiling on both the gesture and speech tasks.

Our second prediction was that subset-knowers' advantage in gesture would be driven by performance on set sizes above each child's knower-level. Therefore, we examined performance on each set size for children for whom that set size was above, at, or below the child's knower-level (Fig. 2). It was not possible to conduct an overall ANOVA across set sizes since different children were represented for each set size. Instead, within each grouping (above, at, or below a child's knower-level), we conducted four paired-samples t -tests comparing speech vs. gesture for each set size. Using a Bonferroni correction for four comparisons, p -values less than .0125 were considered significant. We first examined numbers above a child's knower-level (Fig. 2, Panel A). The set size 1 is above a child's knower-level only for pre-knowers; although the difference in accuracy on 1 between the WOC-Gesture ($M = 85.2\%$,

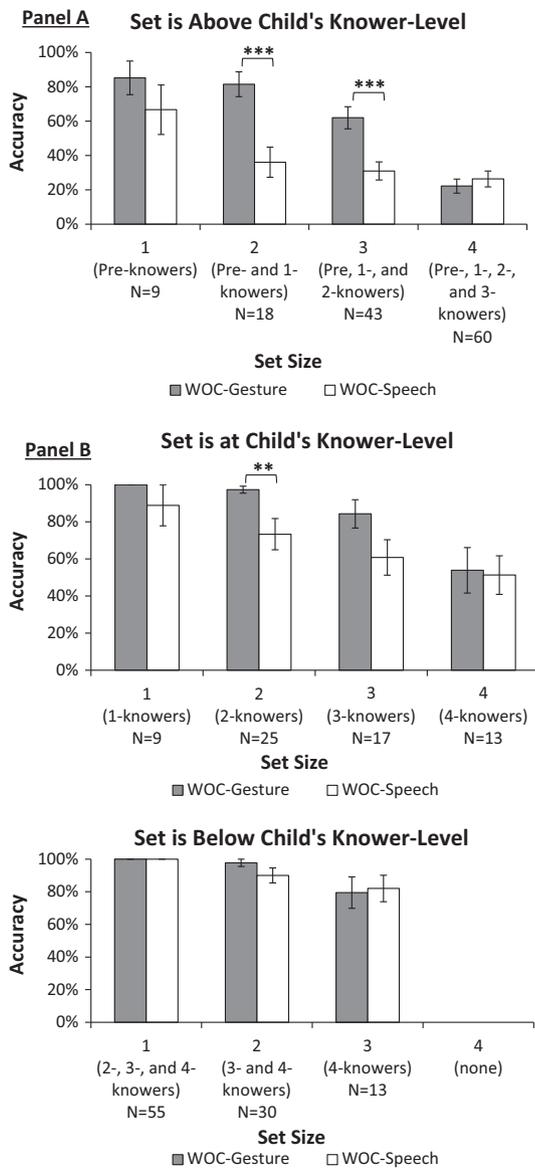


Fig. 2. Accuracy on WOC-Gesture and WOC-Speech, by knower-level and set size. $**p < .0125$ (Bonferroni correction for four comparisons), $***p < .001$.

SD = 29.4%) and WOC-Speech tasks ($M = 66.7\%$, $SD = 43.3\%$) was in the predicted direction, it did not reach statistical significance ($t(8) = 1.01$, $p = .34$, $d = .34$; sign test: exact $p = 1.0$). The set size 2 is above the knower-level of both pre- and one-knowers; these children showed a marked advantage on 2 for WOC-Gesture ($M = 81.5\%$, $SD = 30.7\%$) over WOC-Speech ($M = 36.1\%$, $SD = 37.2\%$; $t(17) = 4.03$, $p < .001$, $d = 0.95$; sign test: exact $p = .004$). For the set size 3, pre-, one-, and two-knowers showed a large advantage on WOC-Gesture ($M = 62.0\%$, $SD = 42.2\%$) over WOC-Speech ($M = 31.0\%$, $SD = 34.4\%$; $t(42) = 4.06$, $p < .001$, $d = .62$; sign test: $Z = -3.60$, $p < .001$). In other words, when the set sizes 2 and 3 were above a child's knower-level, children were more than twice as accurate on the WOC-Gesture task than on the WOC-Speech task. Finally, for the set size 4, pre-, one-, two-, and three-knowers were analyzed, and showed no significant difference in accuracy between the WOC-Gesture ($M = 22.2\%$, $SD = 31.7\%$) and WOC-Speech ($M = 26.4\%$, $SD = 36.0\%$) tasks ($t(59) = 0.88$, $p = .38$, $d = .11$; sign test: $Z = -0.39$, $p = .70$).

We next examined children's performance on set sizes that were at their knower-level (Fig. 2, Panel B), again using a Bonferroni correction for four comparisons so that p -values less

than .0125 were considered significant. Two-knowers performed significantly better on set size 2 on the WOC-Gesture task ($M = 97.3\%$, $SD = 9.2\%$) than the WOC-Speech task ($M = 73.3\%$, $SD = 41.9\%$; $t(24) = 2.90$, $p < .01$, $d = .58$; sign test: exact $p = .039$; note that the paired-samples t -test reached significance while the sign test did not). One-knowers, three-knowers, and four-knowers did not significantly differ in their accuracy on the WOC-Gesture versus WOC-Speech tasks at their knower-level (t -tests: all p 's $> .03$; sign tests: all p 's $> .03$), although WOC-Gesture accuracy was directionally higher than WOC-Speech accuracy in each group (by 11, 24, and 3 percentage points in one-, three-, and four-knowers, respectively).

Finally, we examined children's performance on set sizes that were below their knower-level (Fig. 2, Panel C). Only the set sizes 1, 2, and 3 were analyzed, since the set size 4 was not below the knower-level of any subset-knowers. Using a Bonferroni correction for three comparisons, p -values less than .016 were considered significant. Performance on the set sizes 1, 2, and 3 was near ceiling for both WOC-Gesture and WOC-Speech tasks (between 79% and 100% accuracy), and the difference between WOC-Gesture and WOC-Speech tasks was not significant for any set size (t -tests: all p 's $> .03$; sign tests: all p 's $> .06$).

3.2.2. Gesture–speech mismatches within a response

One potential concern with analyzing the WOC-Gesture and WOC-Speech tasks separately is that differences between the tasks could be due to children interpreting the tasks differently, employing different strategies on each task, or to children becoming fatigued (since the WOC-Speech task always followed the WOC-Gesture task). However, since gesture and speech are produced in different modalities, children can produce both types of responses simultaneously within a single response (referred to as “gesture–speech combinations”). Gesture–speech combinations on set sizes 1–4 were quite frequent; 81% of children produced at least one combination, and the average number of combinations per child was 6.2 ($SD = 5.5$, out of 23 trials). The majority of gesture–speech combinations occurred on the WOC-Speech task ($M = 63.9\%$, $SD = 39.2\%$, $N = 119$), which was expected because the WOC-Gesture task occurred before the WOC-Speech task in order to prime children to gesture during the WOC-Speech task. In addition, it was often the case that the number conveyed in gesture was different from the number conveyed in the accompanying speech (referred to as a “mismatch”). In fact, among subset-knowers, 35.5% ($SD = 33.1\%$, $N = 68$) of all gesture–speech combinations on set sizes 1–4 were mismatches (on the remaining 64.5% of gesture–speech combinations, the gesture and speech conveyed the same number). Among CP-knowers, however, only 3.2% ($SD = 14.5\%$, $N = 58$) of combinations on set sizes 1–4 were mismatches.

To address potential concerns about differences between the WOC-Speech and WOC-Gesture tasks, we examined whether the same gesture advantage occurred within a single response, that is, on gesture–speech mismatches. Fifty-four participants (34.8%) produced at least one gesture–speech mismatch on set sizes 1–4. More than half of subset-knowers mismatched (68.5%): 100% of pre-knowers, 88.9% of one-knowers, 68.0% of two-knowers, 52.9% of three-knowers, and 53.8% of four-knowers produced at least one gesture–speech mismatch. In contrast, only 4.9% of CP-knowers ever mismatched on set sizes 1–4. Mismatches could occur on both the WOC-Gesture and WOC-Speech tasks, although the majority of mismatches occurred on the WOC-Speech task ($M = 74.4\%$, $SD = 38.3\%$, $N = 54$). We aggregated children's production of gesture–speech mismatches across the WOC-Speech and WOC-Gesture tasks in all subsequent analyses.

Subset-knowers' accuracy in the gestural and spoken components of mismatches is displayed in Fig. 3 (only participants who

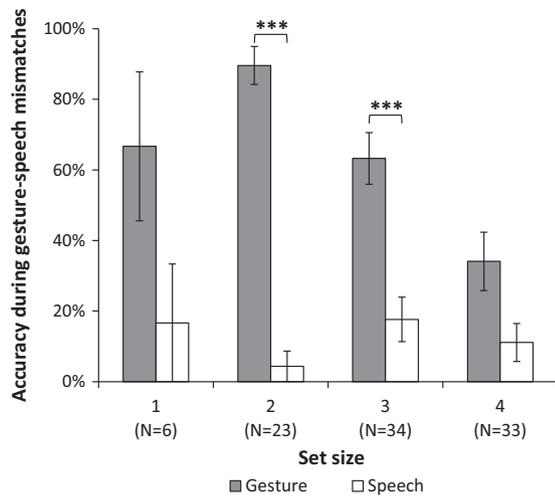


Fig. 3. Subset-knowers' accuracy in gesture versus speech during simultaneous gesture–speech mismatches. Statistical tests are based on comparison of accuracy in gesture versus speech when the child was accurate in either gesture or speech. *** $p < .001$.

mismatched at least once on a given set size are represented in each bar). Since accuracy in gesture and speech were not independent, we could not compare them using paired-samples tests. Instead, we calculated the percent of mismatches that were accurate in gesture out of all mismatches that were accurate in either gesture or speech. We then compared this percentage to chance (50%) for each set size using one-sample t -tests and non-parametric binomial tests. (Note that mismatches that were not correct in either gesture or speech were excluded from these comparisons; they accounted for 17% of mismatches on set size 1, 6% of mismatches on set size 2, 19% of mismatches on set size 3, and 55% of mismatches on set size 4.) We used a Bonferroni correction for 4 comparisons, with $p < .0125$ considered statistically significant. Of all mismatches on the set size 1 that were accurate in either gesture or speech, 80.0% ($SD = 44.7\%$, $N = 5$) were accurate in gesture (and not speech); however, this did not significantly differ from chance (t -test: $p = 0.21$; binomial test: $p = 0.38$). Of all mismatches on the set size 2 that were accurate in either gesture or speech, 95.7% ($SD = 20.9\%$, $N = 23$) were accurate in gesture (and not speech), significantly above chance ($t(22) = 10.5$, $p < .001$, $d = 2.19$; binomial test: $p < .001$). For the set size 3, 80.0% ($SD = 38.5\%$, $N = 30$) of mismatches that were accurate in gesture or speech were accurate in gesture (and not speech), significantly above chance ($t(29) = 4.27$, $p < .001$, $d = .78$; binomial test: $p < .01$). For the set size 4, 75.0% ($SD = 44.7\%$, $N = 16$) of mismatches that were accurate in gesture or speech were accurate in gesture (and not speech), however, this difference was not statistically above chance after the Bonferroni correction ($t(15) = 2.24$, $p = .04$, $d = .56$; binomial test: $p = .08$).

As noted earlier, the majority of mismatches (79%) occurred on set sizes above a child's knower-level; only 15% of mismatches occurred at a child's knower-level and 5% occurred on set sizes below a child's knower-level. Given the small number of mismatches at or below a child's knower-level, we did not do separate analyses of mismatches above, at, and below children's knower-levels.

3.3. Large numbers (5–10)

3.3.1. Approximation on gesture versus speech tasks

We next examined children's responses to larger set sizes (5–10) on the WOC-Speech and WOC-Gesture tasks, considered separately. Since our hypotheses for the larger set sizes involved

children's approximation ability, we restricted our analyses to subset-knowers, who rarely count to reach an exact answer on WOC tasks (Le Corre, Van de Walle, Brannon, & Carey, 2006); CP-knowers typically count to determine the exact answer for larger set sizes, thus preventing the WOC tasks from measuring approximation ability (e.g., Le Corre et al., 2006). (See Appendix Figs. A1–A3 for accuracies, average responses, and histograms of response data for all knower-levels and set sizes.) In all analyses, speech responses greater than “30” were excluded from our analyses as has been done in prior studies (Le Corre & Carey, 2007). Gesture responses were never greater than 10. Three subset-knowers were excluded from these analyses because they did not have at least one valid response on each set size for each task, leaving 70 subset-knowers for analysis.

We examined children's responses in two ways. First, we measured the regression slope relating the set size requested to children's average response on each set size. This value served as a measure of whether children's responses (in speech or gesture) were numerically greater when shown larger sets of objects. Second, to measure how far children's answers were from the correct response, we examined children's Percent Absolute Error (PAE) for each set size requested, where $PAE = |Response - Target|/Target$. Since we did not expect children to respond exactly, we did not analyze accuracy data for these set sizes (see Appendix Fig. A1 for accuracy data).

Subset-knowers' average responses on the WOC-Speech and WOC-Gesture tasks are displayed in Fig. 4, Panel A. On the WOC-Speech task, subset-knowers' average responses to set sizes 6 ($M = 5.15$, $SD = 2.09$) and 9 ($M = 5.09$, $SD = 2.65$) did not significantly differ ($t(69) = 0.30$, $p = .76$; sign test: $Z = -0.39$, $p = .70$). In contrast, on the WOC-Gesture task, subset-knowers' average responses to set sizes 5 ($M = 4.59$, $SD = 1.72$) and 10 ($M = 6.67$, $SD = 2.34$) were significantly different ($t(69) = 8.42$, $p < .001$, $d = 1.0$; sign test: $Z = -6.88$, $p < .001$). Subset-knowers' slopes from 5 to 10 on the WOC-Gesture task ($M = 0.42$, $SD = 0.41$) were significantly higher than their slopes from 6 to 9 on the WOC-Speech task ($M = -0.02$, $SD = 0.59$) ($t(69) = 5.38$, $p < .001$, $d = 0.64$; sign test: $Z = -4.64$, $p < .001$). Note that although the distance between the numbers on the Gesture task (5 and 10) is larger than the distance between the numbers on the Speech task (6 and 9), our slope measure adjusts for this difference. Children's slopes on the WOC-Gesture task were significantly correlated with their knower-levels, $r(68) = .24$, $p < .05$, but their slopes on the WOC-Speech task were not, $r(68) = .11$, $p = .37$. However, we interpret the correlation between knower-levels and WOC-Gesture slopes with caution as this correlation was not robust to alternative analyses (see Footnotes 1 and 2: the correlation was not significant when analyzing only children with complete knower-levels or when examining only children who completed the WOC-Speech task immediately after WOC-Gesture task). Children's slopes on the WOC-Speech and WOC-Gesture task also were not significantly correlated with each other, $r(68) = .12$, $p = .33$.

In order to compare children's ability to approximate across tasks more directly, we calculated the PAE for each set size. Since the PAE is proportional to the target set size, we were able to directly compare responses to medium (5 or 6) and large (9 or 10) responses even though we used different target set sizes in the WOC-Speech (6 and 9) and WOC-Gesture (5 and 10) tasks. A 2 (task: WOC-Speech, WOC-Gesture) \times 2 (set size: medium [5 or 6], large [9 or 10]) ANOVA revealed a significant main effect of task, $F(1, 69) = 9.70$, $p < 0.01$, $\eta_p^2 = .12$, a significant main effect of set size, $F(1, 69) = 4.47$, $p < 0.05$, $\eta_p^2 = .06$, and a significant task \times set size interaction, $F(1, 69) = 7.26$, $p < 0.01$, $\eta_p^2 = .10$. For the large set sizes (9 and 10), subset-knowers' PAE was significantly higher on the WOC-Speech task ($M = 48.4\%$, $SD = 21.4\%$) than on the

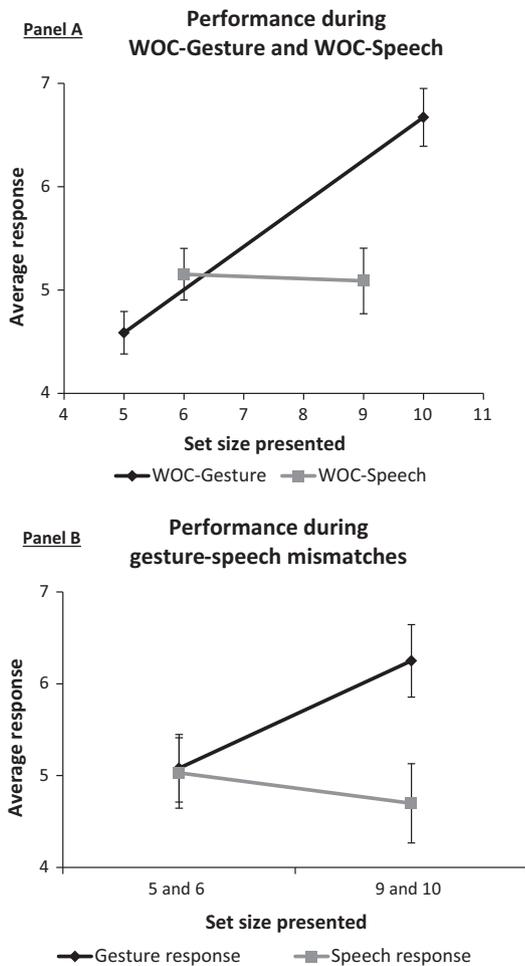


Fig. 4. Subset-knowers' average responses to large set sizes (5–10) on the WOC-Gesture task and on the WOC-Speech task (Panel A) and on the gesture and speech components of the gesture–speech mismatches children produced on both tasks (Panel B). (In Panel B, average responses for each child included only those trials on which the child produced gesture–speech mismatches. Children mismatched most during the WOC-Speech task (trials 6 and 9), and so did not mismatch on equal numbers of trials of 5, 6, 9, and 10. Therefore, in order to compare the graphed data to accurate responding, we calculated the mean response per child if children had responded perfectly accurately on the trials on which they mismatched. If response accuracy had been perfect on these trials, the mean response to trials of 5 and 6 would have been 5.7 (i.e., there were more mismatches on 6 than 5), and the mean response to trials of 9 and 10 would have been 9.2 (i.e., there were more mismatches on 9 than 10).)

WOC-Gesture task ($M = 33.3\%$, $SD = 23.4\%$) ($t(69) = 4.44$, $p < 0.001$, $d = 0.53$; sign test: $Z = -3.13$, $p < .01$). In contrast, for the medium set sizes (5 and 6), subset-knowers' PAE did not significantly differ for the WOC-Speech task ($M = 36.4\%$, $SD = 26.1\%$) versus the WOC-Gesture task ($M = 34.2\%$, $SD = 22.8\%$) ($t(69) = .55$, $p = 0.59$; sign test: $Z = -0.36$, $p = 0.72$). In other words, subset-knowers' responses were farther from the target when they were asked to label a large set size (9 or 10) with a number word than with a number gesture. In contrast, their responses were closer to the target when asked to label a medium set size (5 or 6) with either a number word or gesture.

3.3.2. Gesture–speech mismatches within a response

For the reasons outlined earlier, we also examined children's responses in the gestural and spoken components of a mismatch. For the set sizes 5–10, 77.1% of subset-knowers produced at least one gesture–speech combination (whether a match or mismatch), and the average number of gesture–speech combinations per child

was 3.6 ($SD = 3.1$, out of 12 trials). Among subset-knowers, 63.9% ($SD = 39.1\%$, $N = 54$) of all gesture–speech combinations on set sizes 5–10 were mismatches. Forty-three subjects (61% of subset-knowers) produced at least one gesture–speech mismatch on the set sizes 5–10, with 88.9% of pre-knowers, 50.0% of one-knowers, 60.9% of two-knowers, 47.1% of three-knowers, and 69.2% of four-knowers producing at least one gesture–speech mismatch. As in the smaller numbers, the majority of mismatches on set sizes 5–10 occurred on the WOC-Speech task ($M = 76.0\%$, $SD = 35.8\%$, $N = 43$). We combined gesture–speech mismatches on both tasks in the subsequent analyses.

We first examined gesture and speech responses to set sizes 5–10 among the 34 subset-knowers who produced at least one gesture–speech mismatch for both a medium number (5 or 6) and a large number (9 or 10) (Fig. 4, Panel B). On these mismatches, the slope of children's responses was significantly higher in gesture ($M = 0.35$, $SD = 0.77$) than in speech ($M = -0.09$, $SD = 0.72$) ($t(33) = 2.22$, $p < .05$, $d = 0.38$; sign test: $Z = -2.37$, $p < .05$). The slope in gesture (but not in speech) was positive, reflecting the fact that gesture responses were larger to the large set sizes (9 or 10) than to the medium set sizes (5 or 6).

We also examined whether subset-knowers showed more error (as measured by PAE) during mismatches in their speech responses or in their gesture responses for each set size. A 2 (response modality: gesture vs. speech) \times 2 (set size: medium [5 or 6], large [9 or 10]) ANOVA revealed a significant main effect of response modality, $F(1,33) = 5.10$, $p < .05$, $\eta_p^2 = .13$, a significant main effect of set size, $F(1,33) = 6.30$, $p < .05$, $\eta_p^2 = .16$, and no significant response modality \times set size interaction, $F(1,33) = 1.45$, $p = .24$, $\eta_p^2 = .04$. Although the modality \times set size interaction did not reach significance, perhaps due to the relatively small number of children who produced mismatches for these numbers, we examined whether the pattern of results was in the same direction as when we analyzed all responses on the WOC-Speech and WOC-Gesture tasks. Paralleling the results based on all responses, children's responses during mismatches showed that, for the large set sizes (9 or 10), children's PAE was significantly greater in speech ($M = 52.4\%$, $SD = 21.3\%$) than in gesture ($M = 37.6\%$, $SD = 17.5\%$) ($t(33) = 3.37$, $p < .01$, $d = 0.58$; sign test: $Z = -3.26$, $p = .001$), whereas for the medium set sizes (5 or 6), children's PAE did not significantly differ in speech ($M = 39.3\%$, $SD = 31.5\%$) versus gesture ($M = 33.8\%$, $SD = 20.5\%$) ($t(33) = 0.76$, $p = .45$; sign test: $Z = -1.04$, $p = 0.30$).

4. Discussion

As predicted, we found that subset-knowers were better at labeling both small and large set sizes using number gestures than number words. When shown 2 or 3 objects, subset-knowers were significantly more accurate when giving a gesture response than a speech response. Moreover, this difference was strongest among children who had not yet learned the number words for those set sizes (i.e., the numbers were above their knower-level). For these children, accuracy in gesture was more than twice as high as accuracy in speech. This striking pattern of results shows that, before children learn the cardinal meanings of the number words “two” and “three”, they are able to access non-verbal representations of those set sizes and communicate about them using gesture. Thus, subset-knowers are limited not by their inability to conceptualize or communicate about these set sizes, but by their inability to map number words onto these numerical concepts.

Interestingly, we did not find a difference between subset-knowers' performance in speech and gesture for set sizes 1 or 4. For set size 1, only pre-knowers (who had not yet mastered the word “one”) were expected to show a difference. For the 9

pre-knowers in our sample, responses to set size 1 were, in fact, more accurate in gesture (85%) than in speech (67%). Although this difference was not statistically significant, the direction of the difference was the same as for set sizes 2 and 3, suggesting that the lack of a significant effect may have been due to low power.

For set size 4, in contrast, subset-knowers performed poorly in both gesture and speech, with no significant difference in accuracy between the two tasks. Subset-knowers' low gesture accuracy for set size 4 (28%) compared to set size 3 (70%) suggests that children were not using a simple one-to-one matching strategy (i.e., matching one finger to each item) to complete the task. If they had used such a strategy, one would expect them to be similarly successful in representing set sizes 3 and 4 in gesture, since there is no theoretical upper limit to one-to-one matching, and the gestures for 3 and 4 are similarly easy to produce (if anything, 3 is physically more difficult to gesture than 4). However, our data showed a sharp drop-off in gesture accuracy for the set size 4 versus 3, and subset-knowers in our sample were slightly (but not significantly) more accurate in speech over gesture for sets of 4 objects.

Importantly, 4 was not only the set size at which subset-knowers' performance dropped off in both gesture and speech, but it was also the set size at which CP-knowers showed a significant *speech* advantage over gesture. This pattern suggests that 4 may be a critical value – accurately representing quantities 4 and greater may require the ability to produce summary symbols for sets. We argue that CP-knowers' speech advantage on sets of 4 items reflects the relative ease of producing summary symbols in speech, which children are likely to use more often than gesture to represent cardinal number. In contrast, subset-knowers' poor performance on set size 4 reflects their inability to create a summary symbol for exactly 4 items in either speech or gesture, as well as the failure of the parallel-individuation system to encode an item-based representation (i.e., “one-one-one-one”). These findings are consistent with previous work showing that the parallel-individuation system cannot represent set sizes above 3 (e.g., Feigenson & Carey, 2003, 2005). Note also that subset-knowers show the same pattern of gesture accuracy in representing set size as homesigners (Spaepen et al., 2011) – they produce correct number gestures for sets of 1–3 but not for sets of 4 or more. Homesigners' number gestures have been shown to function in working memory as item-based representations rather than summary symbols (Spaepen et al., 2013). We suggest that subset-knowers may also recruit the parallel-individuation system to create item-based gestural representations of small sets (1–3), without having a summary symbol for these sets.

Subset-knowers' ability to label larger set sizes (5–10) approximately was also better in gesture than in speech. We measured subset-knowers' approximation ability, rather than accuracy, since we did not expect children who had not yet learned the cardinal principle to label large set sizes exactly. Subset-knowers used larger number gestures in response to 9 or 10 objects than in response to 5 or 6 objects, but did not distinguish these set sizes using number words. Subset-knowers' failure to distinguish these set sizes in the verbal modality is consistent with previous work (Gunderson et al., 2015; Le Corre & Carey, 2007; Odic, Le Corre, & Halberda, 2015). Researchers have argued that subset-knowers have difficulty producing large enough number words to label these large set sizes (Gunderson et al., 2015; Wagner & Johnson, 2011). The present study provides corroborating evidence for this position by showing that, when allowed to use gestures (where the set size 10 is readily available simply by showing both hands), children were able to produce larger values for larger set sizes even though they could not do so in speech.

We note several limitations to the present study. First, it is possible that children's success in responding with number gestures

was attributable not only to the affordances of the gesture modality per se, but also to the fact that the experimenter modeled appropriate use of number gestures for each set size while introducing the gesture task; recall that the experimenter only modeled the word “one” while introducing the speech task. We included experimenter modeling in the gesture task because we anticipated that some children would be unfamiliar with using number gestures to represent set sizes, and we wanted to ensure that all children understood the task. We did not expect such brief modeling (one example per set size) to affect children's ability to respond to the task, given that attempts to increase children's knower levels that involved more structured input than ours have yielded only modest improvements (e.g., Huang, Spelke, & Snedeker, 2010). In addition, the fact that children did not show a gesture advantage for set size 4 indicates that modeling does not fully explain our results, since the experimenter modeled all set sizes. However, it is possible that modeling may have helped children on some items, for example, their large-number approximation in gesture. An important direction for future research will be to assess the effects of experimenter modeling by examining children's performance with and without modeling in both gesture and speech tasks.

Second, it is possible that children succeeded in gesture in part because they have more prior experience using number gestures in a cardinal context than in a counting context, whereas they have more prior experience using number words in a counting context than in a cardinal context (Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010; Suriyakham, 2007). If so, children's experience using number words to count may have interfered with their ability to use number words as summary symbols, that is, as cardinal number words. However, it is important to note that greater use of counting on the WOC-Speech task is typically associated with better, not worse, performance in labeling the cardinal value (e.g., Le Corre et al., 2006).

Our results – an advantage of gesture over speech for both small and large set sizes – appear, on the surface, to contradict a previous study by Nicoladis et al. (2010), which used similar methods. However, we believe that this apparent contradiction can be explained by important differences between the two studies. Specifically, we hypothesized that children would express their knowledge better in gesture than in speech *only* in those circumstances where they had not yet acquired a stable verbal representation of the numerical quantity in question. If children already knew the number word for a given set size, or if they already knew the cardinal principle (and thus were able to accurately label all set sizes in their count list), then we expected them to use their highly-practiced verbal labels during both the gesture and speech tasks; performance in gesture would then be (at best) the same as speech, and potentially worse than speech because it is less well-practiced.

Based on these predictions, we categorized children by their level of verbal number knowledge and focused our analyses on children who had not yet learned the cardinal principle. In contrast, Nicoladis et al. (2010) categorized children based on age rather than verbal number knowledge. Indeed, our results showed that cardinal-principle knowers, who already had stable verbal representations of the numbers 1–4, did not show an advantage for gesture over speech, and actually showed an advantage for speech over gesture for set size 4. The high levels of accuracy overall in children in the Nicoladis et al. (2010) study suggest that most participants in their sample were cardinal-principle-knowers, which may explain their finding a speech advantage over gesture. In contrast, in our data, subset-knowers showed a strong advantage for gesture over speech for set sizes 2 and 3 (and a non-significant gesture advantage in the same direction for set size 1), especially when those sets were above their knower-level.

A second major difference between the present study and the study by Nicoladis et al. (2010) is that we measured subset-knowers' responses to large set sizes (5–10) in terms of approximate values rather than exactly accurate values, since these children are unable to use counting to report set sizes above 4 accurately (e.g., Le Corre et al., 2006). We examined slope of responses and percent error, and found that subset-knowers were better able to express the approximate value of large set sizes with gesture than with speech. We excluded CP-knowers from the analysis of large numbers since our tasks were untimed and therefore could not be considered approximation tasks for children who typically use counting to determine set size. Large-number approximation ability does, however, continue to develop among at least some CP-knowers (e.g., Le Corre & Carey, 2007), suggesting that examining CP-knowers' gesture and speech approximation using tasks that prevent counting may be a productive direction for future research.

Finally, we purposely did not test set sizes 6, 7, 8, and 9 because they are especially difficult for children to produce in gesture as they require coordinating two different handshapes simultaneously. A large part of the speech advantage found by Nicoladis et al. (2010) was driven by children's responses to 6, 7, 8, and 9, which may be attributable to manual coordination problems rather than conceptual difficulties with gesture.

Importantly, children's better performance in gesture than in speech was apparent not only when examining separate gesture- and speech-eliciting tasks, but also when examining instances in which children produced different numbers in gesture and speech within the same response, referred to as gesture–speech mismatches. These mismatches were quite frequent among subset-knowers, 69% of whom produced at least one mismatch for set sizes 1–4, and 61% of whom produced at least one mismatch for set sizes 5–10. Further, children's response patterns during gesture–speech mismatches paralleled their overall responses such that gestures were more accurate (or closer to correct, in the case of large numbers) than speech. This result reduces the possibility that differences in children's performance on the gesture and speech tasks can be attributed to task demands, strategy use, fatigue, or to differences in the large-number set sizes presented in each task (speech: 6 and 9 versus gesture: 5 and 10), since the same pattern of results was found when gesture and speech were produced within a single response.

The finding that children frequently produced gesture–speech mismatches, and that their gestures were better than their speech in representing set sizes on these mismatches, raises several exciting new research questions. Previous work indicates that producing gesture–speech mismatches when talking about a concept reflects readiness to learn that concept (Alibali & Goldin-Meadow, 1993; Church & Goldin-Meadow, 1986; Goldin-Meadow, 2003b; Perry et al., 1988). The gesture–speech mismatches we observed may be an index of important individual differences between children as they progress through the developmental trajectory of number word learning. For example, there may be critical developmental stages that occur *between* knower-levels, where children who produce gesture–speech mismatches on numbers above their knower-level may be on the verge of learning the cardinal meanings of these higher number words. These children may be especially susceptible to instruction, a prediction that can be tested in future research.

The present study also raises a more general question about the role of gesture in children's numerical representations: do number gestures facilitate number word learning? One way that gestures could promote number word learning is by shaping the input children receive from others around them, input that is crucial for developing verbal number knowledge (Gunderson & Levine,

2011; Levine et al., 2010). That is, children's own number gestures may prompt adults (such as parents or teachers) to provide the number word (i.e., a summary symbol) that corresponds to the child's gesture (see Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007, who find this tailoring of input in parent–child interactions with respect to non-number words and sentences). This tailor-made input may be especially powerful since it takes advantage of the child's attention to the set size in question, expressed through his or her own gesture, and aligns it with the verbal label (a summary symbol) for that set size.

Children's cardinal number gestures may also play a more direct role in facilitating their number word learning by serving as a bridge between the item-based representations of the parallel-individuation system (for small sets) or the magnitude-based representations of the ANS (for large sets), and the arbitrary summary symbols of the adult verbal number system (i.e., number words). Given that preschool-aged children are able to match dissimilar objects based on set size (Huttenlocher et al., 1994; Mix, 2008), fingers may serve, at least initially, as simply another set of objects (e.g., “finger, finger, finger”) rather than as summary symbols for sets (e.g., “three”). Unlike objects, however, cardinal number gestures are a form of communication, and producing cardinal number gestures may give children practice at explicitly accessing their mental representations of set size (via the parallel-individuation system or the ANS) in a communicative context. This practice may help them to access these mental representations in the context of number word learning as well. Another possibility is that children initially view their cardinal number gestures as sets of objects (“finger, finger, finger”), but that the physical connection between the fingers (i.e., for small numbers, all fingers are connected to one hand) helps children view these fingers as a set, scaffolding children toward a summary-symbol understanding of a given set size. Viewing cardinal number gestures as summary symbols may, in turn, improve acquisition of number words. A third possibility is that children may initially map sets of objects to number gestures, taking advantage of their shared item-based nature to master the first step in number learning, and then gradually map number gestures to number words, taking advantage of their shared representational and communicative properties to eventually learn the cardinal principle.

Alternatively, children may learn number gestures as arbitrary summary symbols even before they learn the cardinal principle; in this case, the gesture advantage we have found would not be attributable to the fact that number gestures can be item-based. In future work, we could examine young hearing children's ability to use number gestures as summary symbols versus item-based representations by asking them to complete a number gesture recall task similar to the memory span test used with homesigners (Spaepen et al., 2013). Another potentially fruitful direction would be to compare children's ability to represent set size by producing number gestures versus matching a set of objects to a target set (e.g., pennies), and then examine which of these abilities (number gesture or object matching) better predicts later number word knowledge. These studies would help determine the point in development when number gestures become summary symbols, and whether those gestural summary symbols play a unique role in children's number word learning.

In summary, the present work shows that children can represent and communicate about number in gesture in ways that are not revealed in their use of number words, particularly for numbers that they have not yet learned in speech. Gesture represents an under-utilized window onto children's numerical representations and holds the potential to improve our

understanding of the typical trajectory preschoolers follow in their numerical development, as well as individual differences in that trajectory.

In addition to these theoretical implications, our findings have potential practical implications for parents and teachers. Asking children to gesture about particular numbers may facilitate the frequency of the instructional input they receive about those numbers (cf. [Broaders et al., 2007](#); [LeBarton, Goldin-Meadow, & Raudenbush, 2015](#)). Moreover, observing these number gestures may help educators understand a child's numerical knowledge at a finer-grained level and thus enable them to provide the child with number-word input that is appropriate to their level of understanding. Future research is needed to determine whether number gestures signal a readiness-to-learn and, if so, how adults can best capitalize on this finer-grained view of children's number understanding. Using the knowledge provided by the child's number gesture as well as their number words has the potential to improve numeracy instruction and learning for all children.

Acknowledgments

This research was supported by the McCormick Foundation and CME Group Foundation grant, "Improving Children's Early Math Skills: From the Classroom to the Lab and Back" to Susan Levine and Susan Goldin-Meadow; by the National Science Foundation Science of Learning Center Grant SBE-1041707, the Spatial Intelligence and Learning Center (SILC), to Susan Levine and Susan Goldin-Meadow; by the National Center for Education Research (NCER) Grant R305C050076 to Elizabeth Gunderson; and by NCER Grant R305B090025 and National Science Foundation Graduate Research Fellowship DGE-1144082 to Dominic Gibson. We thank Jodi Khan for administrative and technical support, and Caroline Ganesin and Emalie Norton for help in data collection.

Appendix A

See [Figs. A1–A3](#).

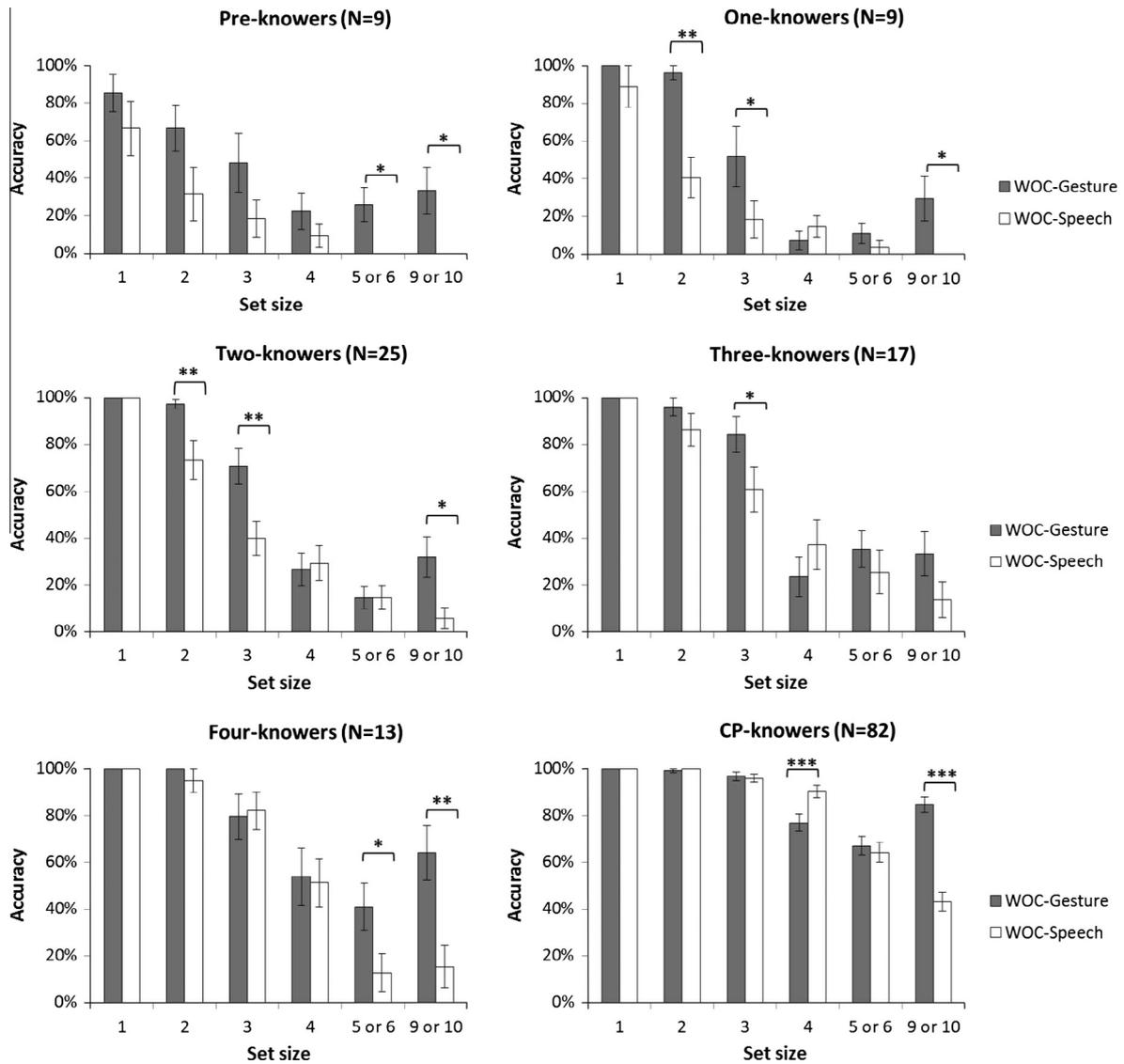


Fig. A1. Accuracy on WOC-Gesture and WOC-Speech, by knower-level and set size. * $p < .05$, ** $p < .01$, *** $p < .001$.

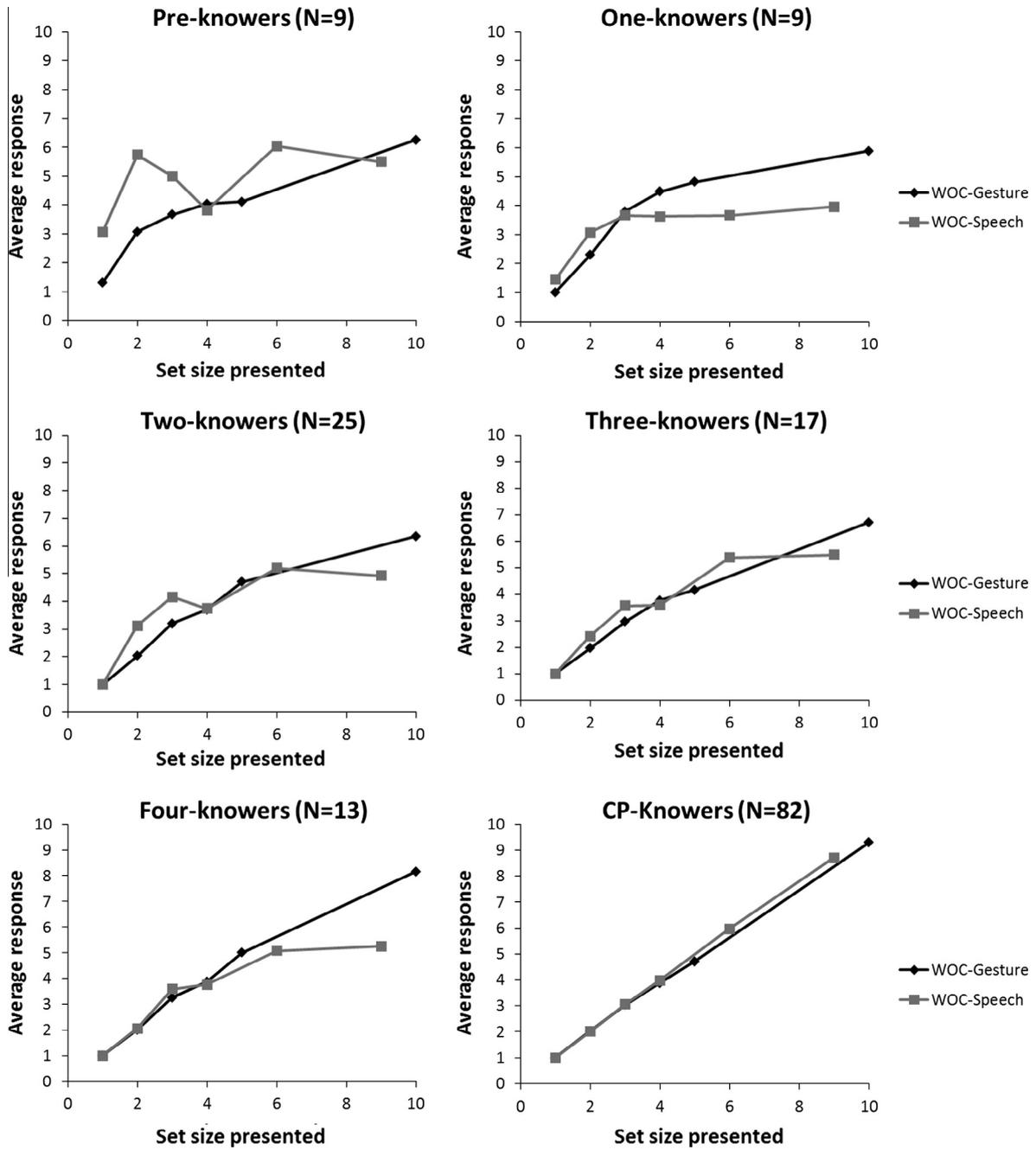


Fig. A2. Average response on WOC-Gesture and WOC-Speech, by knower-level and set size.

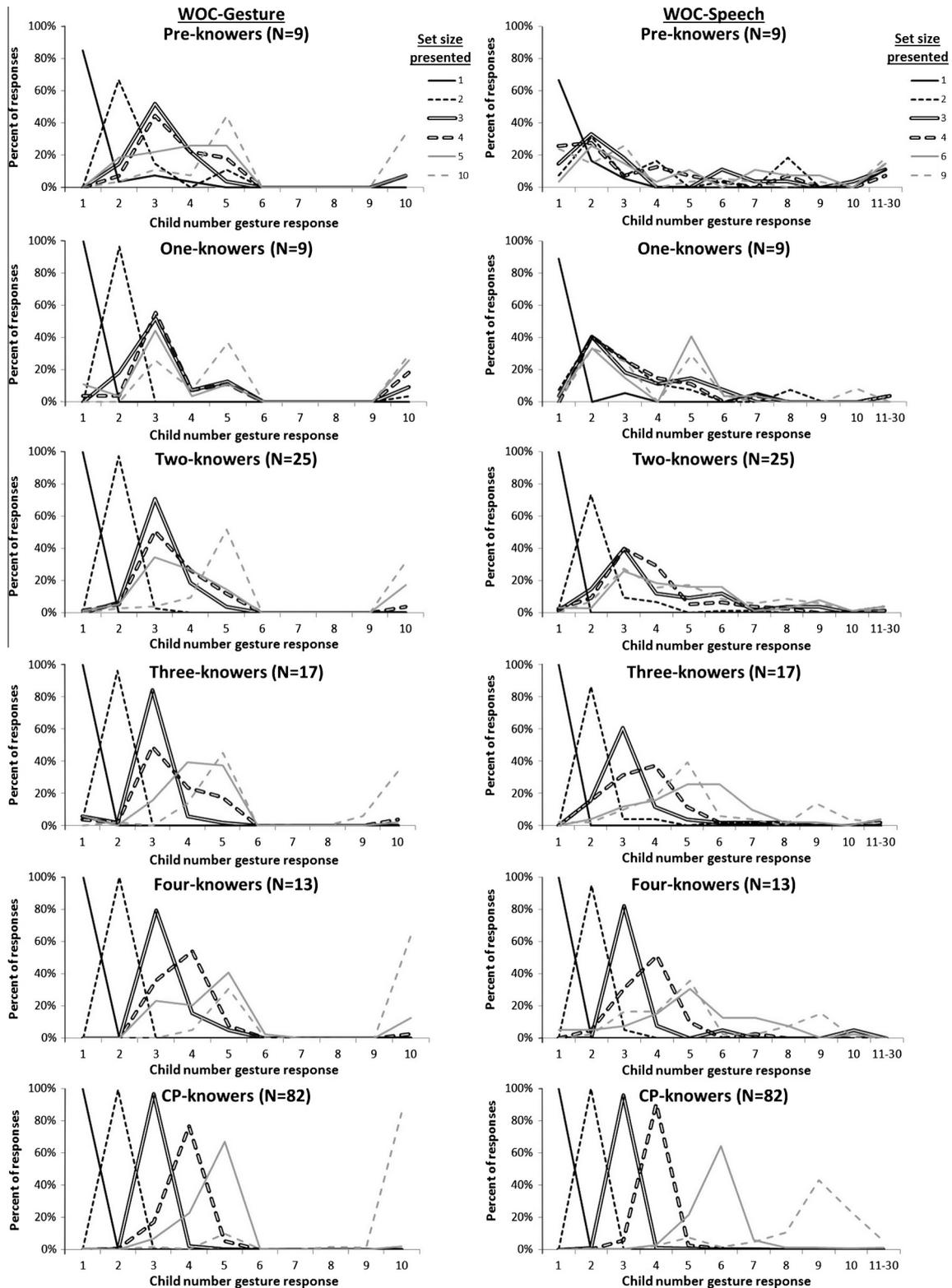


Fig. A3. Histogram of responses on WOC-Gesture (left) and WOC-Speech (right), by knower-level and set size.

References

Alibali, M. W., & DiRusso, A. A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development*, 14(1), 37–56. [http://dx.doi.org/10.1016/S0885-2014\(99\)80017-3](http://dx.doi.org/10.1016/S0885-2014(99)80017-3).
 Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology*, 25(4), 468–523. <http://dx.doi.org/10.1006/cogp.1993.1012>.

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, 60(1), 40–62. <http://dx.doi.org/10.1016/j.cogpsych.2009.06.002>.
 Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition*, 124(2), 156–182. <http://dx.doi.org/10.1016/j.cognition.2012.05.005>.
 Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of*

- Experimental Psychology: General*, 136(4), 539–550. <http://dx.doi.org/10.1037/0096-3445.136.4.539>.
- Cantlon, J., Fink, R., Safford, K., & Brannon, E. M. (2007). Heterogeneity impairs numerical matching but not numerical ordering in preschool children. *Developmental Science*, 431–440.
- Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition*, 128(3), 331–352. <http://dx.doi.org/10.1016/j.cognition.2013.04.008>.
- Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.
- Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23(1), 43–71. [http://dx.doi.org/10.1016/0010-0277\(86\)90053-3](http://dx.doi.org/10.1016/0010-0277(86)90053-3).
- Coppola, M., Spaepen, E., & Goldin-Meadow, S. (2013). Communicating about quantity without a language model: Number devices in homesign grammar. *Cognitive Psychology*, 67, 1–25. <http://dx.doi.org/10.1016/j.cogpsych.2013.05.003>.
- Crollen, V., Mahe, R., Collignon, O., & Seron, X. (2011). The role of vision in the development of finger–number interactions: Finger-counting and finger-monitoring in blind children. *Journal of Experimental Child Psychology*, 109(4), 525–539. <http://dx.doi.org/10.1016/j.jecp.2011.03.011>.
- Crollen, V., Seron, X., & Noël, M.-P. (2011). Is finger-counting necessary for the development of arithmetic abilities? *Frontiers in Psychology*, 2. <http://dx.doi.org/10.3389/fpsyg.2011.00242>.
- Di Luca, S., Lefèvre, N., & Pesenti, M. (2010). Place and summation coding for canonical and non-canonical finger numeral representations. *Cognition*, 117(1), 95–100. <http://dx.doi.org/10.1016/j.cognition.2010.06.008>.
- Di Luca, S., & Pesenti, M. (2008). Masked priming effect with canonical finger numeral configurations. *Experimental Brain Research*, 185(1), 27–39. <http://dx.doi.org/10.1007/s00221-007-1132-8>.
- Di Luca, S., & Pesenti, M. (2011). Finger numeral representations: More than just another symbolic code. *Frontiers in Psychology*, 2(272), 1–3. <http://dx.doi.org/10.3389/fpsyg.2011.00272>.
- Fayol, M., Barrouillet, P., & Marinthe, C. (1998). Predicting arithmetical achievement from neuro-psychological performance: A longitudinal study. *Cognition*, 68(2), B63–B70. [http://dx.doi.org/10.1016/S0010-0277\(98\)00046-8](http://dx.doi.org/10.1016/S0010-0277(98)00046-8).
- Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, 6, 568–584. <http://dx.doi.org/10.1111/1467-7687.00313>.
- Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, 97(3), 295–313. <http://dx.doi.org/10.1016/j.cognition.2004.09.010>.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 307–314. <http://dx.doi.org/10.1016/j.tics.2004.05.002>.
- Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.
- Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Goldin-Meadow, S. (1998). The development of gesture and speech as an integrated system. *New Directions for Child Development*, 79, 29–42.
- Goldin-Meadow, S. (2003a). *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. NY: Psychology Press.
- Goldin-Meadow, S. (2003b). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.
- Goldin-Meadow, S. (2014). Widening the lens: What the manual modality reveals about language, learning, and cognition. *Philosophical Transactions of the Royal Society, Series B*, 369, 20130295. <http://dx.doi.org/10.1098/rstb.2013.0295>.
- Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science*, 10(6), 778–785. <http://dx.doi.org/10.1111/j.1467-7687.2007.00636.x>.
- Goldin-Meadow, S., Levine, S., & Jacobs, S. (2014). Gesture's role in learning arithmetic. In L. Edwards, F. Ferrara, & D. Moore-Russo (Eds.), *Emerging perspectives on gesture and embodiment in mathematics*. Charlotte, NC: Information Age Publishing.
- Graham, T. A. (1999). The role of gesture in children's learning to count. *Journal of Experimental Child Psychology*, 74(4), 333–355. <http://dx.doi.org/10.1006/jecp.1999.2520>.
- Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: Relations between parents' input and children's cardinal-number knowledge. *Developmental Science*, 14(5), 1021–1032. <http://dx.doi.org/10.1111/j.1467-7687.2011.01050.x>.
- 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. <http://dx.doi.org/10.1016/j.jecp.2014.09.008>.
- Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, 44(5), 1457–1465. <http://dx.doi.org/10.1037/a0012682>.
- Huang, Y. T., Spelke, E., & Snedeker, J. (2010). When is four far more than three? *Psychological Science*, 21(4), 600–606. <http://dx.doi.org/10.1177/0956797610363552>.
- Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General*, 123(3), 284–296. <http://dx.doi.org/10.1037/0096-3445.123.3.284>.
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 367–371. <http://dx.doi.org/10.1111/j.0956-7976.2005.01542>.
- 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. <http://dx.doi.org/10.1016/j.cognition.2006.10.005>.
- Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology*, 130–169. <http://dx.doi.org/10.1016/j.cogpsych.2005.07.002>.
- LeBarton, E. S., Goldin-Meadow, S., & Raudenbush, S. (2015). Experimentally-induced increases in early gesture lead to increases in spoken vocabulary. *Journal of Cognition and Development*, 16(2), 199–220. <http://dx.doi.org/10.1080/15248372.2013.858041>.
- 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. <http://dx.doi.org/10.1037/a0019671>.
- Mix, K. S. (2008). Surface similarity and label knowledge impact early numerical comparisons. *British Journal of Developmental Psychology*, 26, 13–32. <http://dx.doi.org/10.1348/026151007X189109>.
- National Research Council (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Nicoladis, E., Pika, S., & Marentette, P. (2010). Are number gestures easier than number words for preschoolers? *Cognitive Development*, 25(3), 247–261. <http://dx.doi.org/10.1016/j.cogdev.2010.04.001>.
- Noël, M.-P. (2005). Finger gnosis: A predictor of numerical abilities in children? *Child Neuropsychology*, 11(5), 413–430. <http://dx.doi.org/10.1080/09297040590951550>.
- Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition*, 138, 102–121. <http://dx.doi.org/10.1016/j.cognition.2015.01.008>.
- Özçalışkan, S., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, B101–B113.
- Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, 3(4), 359–400. [http://dx.doi.org/10.1016/0885-2014\(88\)90021-4](http://dx.doi.org/10.1016/0885-2014(88)90021-4).
- Pine, K. J., Lufkin, N., & Messer, D. (2004). More gestures than answers: Children learning about balance. *Developmental Psychology*, 40(6), 1059–1067. <http://dx.doi.org/10.1037/0012-1649.40.6.1059>.
- Potter, M. C., & Levy, E. I. (1968). Spatial enumeration without counting. *Child Development*, 39(1), 265–272. <http://dx.doi.org/10.2307/1127377>.
- Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, 108, 662–674. <http://dx.doi.org/10.1016/j.cognition.2008.05.007>.
- Sarnecka, B. W., Goldman, M. C., & Slusser, E. B. (2015). How counting leads to children's first representations of exact, large numbers. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford handbook of numerical cognition*. New York: Oxford University Press.
- Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology*, 325–337. <http://dx.doi.org/10.1016/j.jecp.2009.02.007>.
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, 114(3), 418–431. <http://dx.doi.org/10.1016/j.jecp.2012.10.012>.
- Saxe, G. B. (1977). A developmental analysis of notational counting. *Child Development*, 48(4), 1512–1520. <http://dx.doi.org/10.2307/1128514>.
- Saxe, G. B., & Kaplan, R. (1981). Gesture in early counting: A developmental analysis. *Perceptual and Motor Skills*, 53(3), 851–854. <http://dx.doi.org/10.2466/pms.1981.53.3.851>.
- Spaepen, E., Coppola, M., Flaherty, M., Spelke, E., & Goldin-Meadow, S. (2013). Generating a lexicon without a language model: Do words for number count? *Journal of Memory and Language*, 69, 496–505. <http://dx.doi.org/10.1016/j.jml.2013.05.004>.
- 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*, 3163–3168. <http://dx.doi.org/10.1073/pnas.1015975108>.
- Starkey, P., & Cooper, R. G. J. (1980). Perception of numbers by human infants. *Science*, 210, 1033–1035. <http://dx.doi.org/10.1126/science.7434014>.
- Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, 36(2), 97–127. [http://dx.doi.org/10.1016/0010-0277\(90\)90001-z](http://dx.doi.org/10.1016/0010-0277(90)90001-z).
- Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerosity. *Child Development*, 52(4), 1146–1152. <http://dx.doi.org/10.2307/1129500>.
- Suriyakham, L. W. (2007). *Input effects on the development of the cardinality principle: Does gesture count?* Doctoral dissertation, University of Chicago.
- Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, 10–22. <http://dx.doi.org/10.1016/j.cognition.2010.11.014>.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155–193. [http://dx.doi.org/10.1016/0010-0277\(90\)90003-3](http://dx.doi.org/10.1016/0010-0277(90)90003-3).
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, 24, 220–251. [http://dx.doi.org/10.1016/0010-0285\(92\)90008-P](http://dx.doi.org/10.1016/0010-0285(92)90008-P).