Cognition 180 (2018) 59–81

Contents lists available at ScienceDirect

Cognition

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

Meaning before order: Cardinal principle knowledge predicts improvement in understanding the successor principle and exact ordering

Elizabet Spaepen[a,§,*], Elizabeth A. Gunderson[b,§], Dominic Gibson[a], Susan Goldin-Meadow[a], Susan C. Levine[a]

[a] University of Chicago, United States
[b] Temple University, United States

ARTICLE INFO

Keywords:
Number
Cardinal principle
Successor function
Successor principle
Exact ordering

ABSTRACT

Learning the *cardinal principle* (the last word reached when counting a set represents the size of the whole set) is a major milestone in early mathematics. But researchers disagree about the relationship between cardinal principle knowledge and other concepts, including *how counting implements the successor function* (for each number word N representing a cardinal value, the next word in the count list represents the cardinal value N + 1) and *exact ordering* (cardinal values can be ordered such that each is one more than the value before it and one less than the value after it). No studies have investigated acquisition of the successor principle and exact ordering over time, and in relation to cardinal principle knowledge. An open question thus remains: Is the cardinal principle a "gatekeeper" concept children must acquire before learning about succession and exact ordering, or can these concepts develop separately? Preschoolers ($N = 127$) who knew the cardinal principle (CP-knowers) or who knew the cardinal meanings of number words up to "three" or "four" (3–4-knowers) completed succession and exact ordering tasks at pretest and posttest. In between, children completed one of two trainings: counting only versus counting, cardinal labeling, and comparison. CP-knowers started out better than 3–4-knowers on succession and exact ordering. Controlling for this disparity, we found that CP-knowers improved over time on succession and exact ordering; 3–4-knowers did not. Improvement did not differ between the two training conditions. We conclude that children can learn the cardinal principle without understanding succession or exact ordering and hypothesize that children must understand the cardinal principle before learning these concepts.

1. Introduction

Preschool numeracy lays a critical foundation for later mathematics. Indeed, kindergarten-entry math predicts math achievement through high school (Watts, Duncan, Siegler, & Davis-Kean, 2014), and is a better predictor of later achievement in math and reading than kindergarten-entry reading, attention, or socio-emotional skills (Duncan et al., 2007). Understanding the development of early mathematical skills is therefore critical for supporting long-term achievement in this domain. One noteworthy step in a preschool child's development is acquisition of the *cardinal principle*: that the last word reached when counting a set represents the size of the whole set (Gelman & Gallistel, 1978). Although this accomplishment may seem quite easy to adults, knowing the cardinal principle is a major milestone for a preschooler, leading to many new numerical competencies (e.g., Le Corre, 2014; Le Corre, Van de Walle, Brannon, & Carey, 2006; Mix, 2008; Sarnecka & Wright, 2013; Wynn, 1992).

1.1. Developmental trajectory of cardinal number knowledge

A large body of research has established that children can count a set of objects correctly by age 3 (Fuson, 1988; Gelman & Gallistel, 1978; Wynn, 1990). However, children do not initially understand the meaning of the count list, nor the quantities represented by each number word. Children learn the cardinal meanings of the number words (e.g., that "two" means a set of two objects) one at a time, and in order (e.g., Le Corre & Carey, 2007; Sarnecka & Lee, 2009; Wynn, 1992). First, children can comprehend and produce sets of "one" object on request, but fail to accurately produce larger sets of objects ("one-knowers"). A few months later, children can comprehend and produce sets of "two", but not higher numbers ("two-knowers"). Children go through the same stages for "three" and "four" ("three-knowers" and

* Corresponding author at: UChicago STEM Education, 1427 East 60th St., Chicago, IL 60637, United States.
E-mail address: liesje@uchicago.edu (E. Spaepen).
§ Both authors contributed equally to this work.

https://doi.org/10.1016/j.cognition.2018.06.012
Received 1 July 2016; Received in revised form 7 June 2018; Accepted 18 June 2018
0010-0277/

"four-knowers"), and then learn the cardinal principle (Wynn, 1992). This developmental sequence is protracted, taking between 1 and 2 years from when a child becomes a one-knower to when the child learns the cardinal principle (Sarnecka, Goldman, & Slusser, 2015; Wynn, 1992). Learning the cardinal principle is a major achievement at this age. After learning the cardinal principle, children can typically produce sets of any size that they can count to (Wynn, 1992); correct errors made when producing a set by spontaneously counting (Le Corre et al., 2006); match dissimilar sets based on set size (Mix, 2008); and recognize equi-numerosity (that only sets that can be placed in one-to-one correspondence have the same number) (Sarnecka & Wright, 2013).

But how do children conceptualize the relationships *between* numbers as they learn the values of individual numbers and number words over this protracted period? That is, as children learn the cardinal meanings of the numbers "one," "two," and "three," do they learn them as largely separate pieces of knowledge, or do they also consider what makes "two" different from "one" and "three"? This concept, which we will call "exact numerical relations," has at least two related, but separable parts: (1) knowledge of *how counting implements the successor function*—for each number word N representing a cardinal value, the next word in the count list represents the cardinal value N + 1 (for brevity we will call this the "the successor principle" or simply "succession"); (2) knowledge of *exact numerical order*—that cardinal values themselves (not just the words to describe them) can be ordered, such that each is exactly one more than the previous value, and one less than the value after it.

Here we consider not only children's developing knowledge of the successor principle (for numbers less than 10), but also their knowledge of exact ordering. We consider each concept separately, and we also consider how these pieces of knowledge relate to each other in children's developing knowledge of natural number.

1.2. Relationship between cardinal principle and successor principle knowledge

As a group, children who understand the cardinal principle also typically perform above chance on a task tapping knowledge of the *successor principle* (Sarnecka & Carey, 2008). Nevertheless, little is known about *how* learning the cardinal principle relates to the acquisition of the concept of succession. One theoretical proposal posits that learning the cardinal principle entails learning the successor principle (Sarnecka & Carey, 2008). To explain the transition from four-knower to cardinal-principle-knower, Carey (2009) proposed a bootstrapping theory in which children learn the count list initially as a sequence of meaningless placeholders. They then learn the individual meanings of "one", "two", "three", and "four" by mapping these words onto the enriched parallel-individuation system in which long-term memory representations of particular sets are mapped onto verbal number words (Le Corre & Carey, 2007). These representations can then be used to create one-to-one correspondence with other small groups, such that, for example, any set that maps onto the representation in long-term memory representation of "Thing A-Thing B-Thing C" gets assigned the verbal label "three." Because the enriched parallel-individuation system has an upper limit of 4 items, children are unable to map the number words above "four" onto states of this system directly. Instead, Carey and colleagues theorized that children begin to notice the correspondence between the number words whose cardinal meanings they have learned individually ("one", "two", "three", and "four") and the order of the count list. Specifically, they argued that children notice that the next number in the count list corresponds to a state of the enriched parallel-individuation system that has one more item than the previous number in the count list. In other words, a four-knower would notice that "two" refers to sets that are exactly one more than "one", that "three" refers to sets that are exactly one more than "two, and that "four" refers to sets that are exactly one more than "three." The child then makes a logical induction that the successor principle holds for all numbers in their count list—in other words, that the next number in their count list refers to a set size exactly one more than the previous number in their count list. According to this theory, inducing the successor principle is what propels children to become cardinal-principle-knowers (CP-knowers).

To assess this theory, Carey and colleagues developed a task to measure children's understanding of the successor principle (Sarnecka & Carey, 2008). In this task, called the Unit task, the experimenter puts a certain number of objects into a box while labeling them (e.g., "I'm putting 5 buttons in the box"), adds either 1 or 2 objects to the box, and asks the child whether there are now 6 or 7 buttons in the box. A child who understands the successor principle should know that adding one object to a set of N (in this case, 5 buttons) means that the set now contains N + 1 (in this case, 6 buttons, the next number in the count list). Consistent with the bootstrapping theory, subset-knowers, including 3- and 4-knowers, performed at chance on this task, whereas CP-knowers performed above chance.

Although this result is consistent with the bootstrapping theory, it does not rule out the possibility that CP-knowers learn the successor principle *after* becoming CP-knowers, rather than *as* they learn the cardinal principle. Indeed, several studies have shown that children can understand the cardinal principle without succeeding on tasks that require understanding the successor principle (Davidson, Eng, & Barner, 2012; Wagner, Kimura, Cheung, & Barner, 2015). Many CP-knowers showed little or no evidence of understanding the successor principle, performing at or below chance on the Unit task. This is notable because the successor principle is implicitly represented in correct implementation of the counting procedure (i.e., counting that satisfies the counting principles of 1-to-1 correspondence, stable order, and the cardinal principle); however, this implicit representation is insufficient for CP-knowers to succeed on the Unit task, which requires applying successor principle knowledge in a different context than counting. In addition, among CP-knowers in these studies, performance on the Unit task was associated with performance on other numerical tasks, including counting fluency and the ability to estimate set sizes without counting, suggesting that the individual differences reflect true differences in number knowledge rather than artifacts of noise in the data. As a result, Davidson et al. (2012) have argued that children may *only* be able to learn succession once they grasp the cardinal principle.

Because the available studies have used only cross-sectional data, they have not been able to answer the question of whether learning the cardinal principle is *necessary* for children to learn the successor principle. That is, is the cardinal principle a "gatekeeper" skill that is needed before children can learn about succession? This question requires looking at change over time.

Two learning trajectories relating succession and cardinal principle knowledge are plausible, given current research. One possibility is that children learn the successor principle only *after* the cardinal principle, as Davidson and colleagues have argued. A second possibility is that children learn the successor principle and cardinal principle independently, and can learn the two concepts in either order. To distinguish between these possibilities, we need to compare learning over time among children who are not yet CP-knowers (3- and 4-knowers) to children who have learned the cardinal principle on the same successor principle tasks.

The two possible learning trajectories make divergent, testable predictions about who should improve over time in their understanding of the successor principle. Our prediction is that children *must* learn the cardinal principle prior to learning the successor principle, because it requires children to think not only about the cardinal meanings of number words, but also about the meanings of the words and values in relation to each other. If this is the case, then *only* children who are cardinal-principle-knowers should improve their understanding of the successor principle over time. Alternatively, if knowledge of the successor principle and cardinal principle develop independently, then cardinal principle knowledge should not relate to improvement on the successor principle over time: Children who do, and children who do not, yet understand the cardinal principle could both improve their understanding of the successor principle over time.

1.3. Relationship between cardinal principle and exact ordering knowledge

Less is known about how children learn *exact ordering*; that is, that cardinal values can be ordered such that each value is one more than the number before it, and one less than the number after it. Most prior work on numerical ordering has focused on ordering number symbols rather than ordering sets of objects (Lyons & Beilock, 2009, 2011; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Sella, Berteletti, Lucangeli, & Zorzi, 2017; Xu & LeFevre, 2016). However, it is likely that sequencing arbitrary symbols is not the same skill, or related in the same way to learning the cardinal principle, as understanding the more conceptual idea that cardinal values themselves *can* be ordered, where each set contains exactly one more item than the previous set and one less than the next set.

To our knowledge, little work has investigated exact ordering of sets based on their cardinal values. However, related work has established that even pre-linguistic infants have some sense of approximate numerical ordering of sets. By 11 months, infants represent basic ordinal relations between numbers—when repeatedly shown sets of dots that increase in value, 11-month-old infants notice when the values appear in decreasing order (Brannon, 2002). In addition, children who know very little about verbal number words (pre-knowers) are unable to make ordinal judgments about small set sizes, but children with even a small amount of knowledge of the verbal count list are able to make simple ordinal judgments (choosing which of two area-controlled images has more) (Brannon & Van de Walle, 2001). Importantly, however, neither of these experiments required children to demonstrate knowledge of *exact* numerical ordering; that is, to find the correct numerical order of values that differed by one object beyond set sizes that can be represented by parallel individuation.

Orders based on cardinal value satisfy a "later-is-greater" rule—-cardinal values represented by number words later in the count list are larger than cardinal values represented by numbers earlier in the count list. Children appear to learn that number words in the count list follow the "later-is-greater" rule after they learn the cardinal principle (Le Corre, 2014), consistent with our hypothesis that understanding the relations between set sizes develops only after the cardinal principle. However, it is possible to apply a type of non-symbolic later-is-greater rule to sequences of cardinal values (without connecting those values to number words or symbols) using only pre-linguistic number systems (parallel individuation and the approximate number system). But note that children who use this kind of system will not arrive at a fully correct order when numbers go higher than 4 or 5, as there will not be any way to distinguish between sets of 6 and 7, for example.

We know very little about how children acquire the ability to order sets by cardinal value, and even less about how that learning is related to learning the cardinal principle. However, as previously noted, ordering by cardinal value requires children to think relationally about those values (as opposed to thinking only about individual set sizes), just as solving successor principle tasks does. As a result, we expect cardinal principle knowledge to be a necessary precursor to exact ordering knowledge; if this is the case, then only children who are cardinal-principle-knowers should improve their understanding of exact ordering over time. In contrast, if exact ordering and the cardinal principle can develop independently, then CP knowledge should not relate to improvement on exact ordering tasks over time—children who do, and do not, yet understand the cardinal principle could improve their exact ordering ability over time.

We also know little about how exact ordering and knowledge of the successor principle are related. As noted earlier, both exact ordering and succession require children to consider relations between cardinal values, and this may only be possible for children to do consistently once they have a firm understanding of those values themselves (i.e., after they learn the cardinal principle). We therefore hypothesize that exact ordering and succession will have similar relationships to cardinal principle knowledge. In addition, considering the two concepts together will allow us to examine how succession and ordering are related during acquisition.

1.4. The present study

This study was designed to address two separate questions. First, *who* can learn the successor principle and exact ordering over time—-only CP-knowers, or also 3- and 4-knowers? Second, can explicit teaching help children learn the successor principle and exact ordering (compared to practice counting objects)? To address these questions, we designed an experimental training regimen intended to accelerate children's learning of the successor principle and exact ordering when compared to a control condition. The experimental training involved counting objects, labeling cardinal values, explicit comparisons between simultaneously presented sets, and explicit teaching of the "plus one" relation between consecutive set sizes. We compared this training to a control training condition that only involved counting set sizes with 1–6 objects. We administered these training conditions, along with a pretest and posttest assessing knowledge of exact numerical relations, to preschool children who were on the verge of learning the cardinal principle (3- and 4-knowers) and to preschool children who had already learned the cardinal principle (CP-knowers).

Because previous research suggests that moving children along this developmental trajectory in short-term experimental studies is very difficult (e.g., Huang, Spelke, & Snedeker, 2010; Mix, Sandhofer, Moore, & Russell, 2012), we considered the possibility that the two conditions would not produce significantly different learning gains. In that case, we reasoned that the pretest and posttest results could still provide valuable information about our first question—*who* can learn the successor principle and exact ordering over time. Specifically, if children can learn these concepts either before or after they learn the cardinal principle, then both CP-knowers *and* 3–4-knowers ought to benefit from training and/or the passage of time, consistent with the independence of these developmental milestones. In contrast, if children only gain an understanding of exact numerical relations after they understand the cardinal principle, then CP-knowers (but *not* 3- and 4-knowers) across both conditions ought to benefit from training and/or the passage of time between pre- and posttest. This outcome would provide the strongest evidence to date that children must know the cardinal principle to be ready to learn concepts about exact relations between numbers, such as exact ordering and the successor principle.

It is important to note that children who know the cardinal principle will vary in how long they have known it. Additionally, as Davidson et al. (2012) showed, many CP-knowers have already learned the successor principle (and presumably are able to place cardinal values in exact numerical order), whereas others have not. Because we were interested in understanding whether knowing the cardinal principle is a "gatekeeper" to learning about exact numerical relations, we were particularly interested in studying CP-knowers whose pretest succession and ordering scores were in the same range as those of 3- and 4-knowers (low-successor-knowledge CP-knowers and low-ordering-knowledge CP-knowers)—likely those children who only recently transitioned to being CP-knowers. By restricting our CP-knower group to those children with comparable scores on exact numerical relations tasks to 3- and 4-knowers, we can test whether being a CP-knower is key to being *ready* to learn exact numerical relations.

The task we used to assess exact ordering ability was adapted from a task used with Nicaraguan homesigners: Ordering sets of fingers and dots (Spaepen, 2008). In this task, children put cards representing the set sizes 1–7 in order. The cards depict set sizes using Dots and Gestures (images of hands holding up each number of fingers). We chose to have children order up to 7 so that children could not succeed using only the nonverbal approximate number system (ANS) because discrimination of the largest sets (6:7 ratio) is beyond the sensitivity of the ANS at this age (Halberda & Feigenson, 2008). Instead, to succeed, children need to understand that the next card must depict a set that is exactly one more than the previous card. This task allows us to look at a strict measure of children's exact ordering ability—that is, their ability to order all 7 cards correctly—but also allows us to look at their partial understanding of ordering using more fine-grained analysis, such as the highest cardinal value they can place correctly.

We included several additional tasks to understand children's numerical knowledge better. As a point of comparison for the Unit task, we tested children's performance on the What Comes After task (Davidson et al., 2012). This task requires children to reason about the sequence of numbers in the count list not starting at "one," but does not require them to understand the numbers' relations to set size and the successor principle. Because the task is structurally similar to the Unit task but does not require successor principle knowledge, it allowed us to determine whether improvements on the Unit task were specific to tasks requiring successor principle knowledge, or general to other numerical tasks.

Finally, we assessed children's ability to order sets sizes 1–3. These data provide an explicit test of 3- and 4-knowers' ability to order sets within their known number range. Failure on this task would be particularly striking given that children could use either exact or approximate numerical strategies to succeed on this task.

2. Method

2.1. Participants

A total of 127 children (61 females) participated. Children were on average 4.6 years old ($SD = 0.57$, range = 3.1 to 5.6 years; see Table 1 for ages within each knower-level). Children were socioeconomically diverse. Family income ranged from less than \$15,000 per year to more than \$100,000 ($M = \$47,378$, $SD = \$37,122$, $n = 102$). Parents' maximum years of education ranged from less than high school to a graduate degree ($M = 14.6$ years, $SD = 2.5$, $n = 110$, where 14 years is a 2-year associate's degree).

Initially, 275 children provided parental consent for the study. However, in order to be included in the study, children had to complete the Give-N task at pretest and be categorized as a three-knower, a four-knower, or a CP-knower, and also complete the Give-N task at posttest. As a result, 148 children were ineligible for the study: 18 who did not complete the Give-N task at pretest; 116 who completed Give-N and were categorized as pre-, one-, two-, or five-knowers; 4 who had uncodable Give-N data at pretest; and 10 who did not complete the Give-N task at posttest. Pre-, one-, and two-knowers were excluded because we expected that they were not far enough along in the developmental trajectory of number knowledge to benefit from successor principle training. Five-knowers were excluded because they are uncommon ($n = 6$ in our sample) and cannot be definitively categorized as either CP-knowers or subset-knowers. One child who failed to complete the Unit task at pretest was excluded from analyses involving the Unit task, but was included in the other analyses (i.e., the child was included in the total sample of 127).

Table 1
Age of children by knower-level.

Knower-level	N	Age	
		Mean (SD)	Range
Three-knowers	20	4.1 (0.5)	3.1–4.9
Four-knowers	12	4.5 (0.6)	3.5–5.3
Cardinal-principle-knowers	95	4.7 (0.5)	3.4–5.6
Total	127	4.6 (0.6)	3.1–5.6

2.2. Pretest and posttest measures

Knower-level. Knower-level was determined using the Give-N task (Wynn, 1990). In this task, children place a certain number of toy fish into a pond. The experimenter first asked the child to put one fish in the pond. If the child responded incorrectly, the experimenter provided one chance for the child to correct him/herself by saying, "But I asked for N fish! Let's check. [Experimenter and child count fish.] Can you put N fish in the pond?" The child's answer after the correction prompt was recorded. On subsequent trials, if the child responded correctly, the experimenter asked the child to place N + 1 fish in the pond. If the child responded incorrectly, the experimenter asked the child to place N-1 fish in the pond. If the child placed the correct number of fish in the pond up to 6, and placed 6 correctly a second time, he or she was considered a cardinal-principle-knower (CP-knower). Otherwise, the child was categorized as an N-knower if s/he met the following criteria established by Wynn (1992): N was the highest number for which (1) the child gave the correct response to N on at least 2 out of 3 trials, and (2) the child produced N fish no more than half as often, percentage-wise, when asked for other numbers (above N) than when asked for N itself. As noted previously, only three-knowers, four-knowers, and CP-knowers were included in this study.

Unit task. Successor principle knowledge was assessed using the Unit Task (Sarnecka & Carey, 2008). On each trial, the experimenter placed N fish into an opaque box and covered the lid, saying, "Watch, I'm putting N fish in the box." The experimenter checked that the child was paying attention by saying, "How many fish?" and providing feedback if the child responded incorrectly. Then the experimenter added either 1 or 2 fish, depending on the trial, saying, "Now watch." Finally, the experimenter asked, "Now, is it N + 1 or N + 2?" The responses were always presented in ascending order; on half of the trials the correct response was N + 1 and on the other half the correct response was N + 2. Testing began with two practice trials of 1 + 1 and 2 + 1. On these trials, the child was allowed to look into the box after responding. The experimenter then conducted six test trials, instantiating the equations 4 + 1, 4 + 2, 5 + 1, 5 + 2, 6 + 1, and 6 + 2. No feedback was given on the test trials. Trials were presented in a single random order. The child's overall score was the percent correct of 6 items.

Ordering tasks. We assessed ordering knowledge using several Ordering tasks. In these tasks, the experimenter presented the child with a set of 4″ × 6″ note cards, each representing a different set size (see Appendix Fig. A1 for images of the stimuli). The experimenter arrayed the cards on the table in a pseudo-random arrangement so that the child could see all of the cards simultaneously. The experimenter asked the child to put the cards in order. The experimenter indicated that the first card should be placed on the child's left, then the next card should be placed to the right of the first card, etc. If needed, the experimenter prompted the child by saying, "Which one goes next?"

Ordering was assessed using two number representations: Dots and Gestures.[1] For each representation, there were two tasks. The child was first tested with 3 cards (1, 2, and 3). The child was asked to place the three cards in order. If the child did not put the card representing 1 first, the experimenter told the child that "this card goes first" and asked the child to place the remaining cards again. Only the child's first response (before feedback) was analyzed. The child was then tested with 7 cards (1–7). The order of the Dot and Gesture tasks was counterbalanced

[1] We administered parallel Ordering tasks involving Arabic numerals (Ordering 1–7 Arabic Numerals and Ordering 1–3 Arabic Numerals), which were always administered after the Ordering Dots and Gestures tasks. However, we also administered an Arabic numeral identification task, and found that only 20% of 3-knowers, 17% of 4-knowers, and 65% of CP-knowers could correctly label all of the Arabic numerals from 1 to 7. The children's lack of familiarity with the Arabic numerals, especially among 3–4-knowers, makes the Ordering Arabic Numerals tasks difficult to interpret. Nevertheless, we provide descriptive statistics for these tasks in Appendix F.

across children. To assess our primary dependent measure—exact ordering—we used children's accuracy on each Ordering task; that is, whether the child put all of the cards in the correct order. In addition, to explore children's performance on these tasks in a more fine-grained manner, we created a highest correct score on each task.[2] The highest correct score was the highest number for which the child had put all numbers in the correct order up to that point. For example, if a child created an order on an Ordering 1–7 task of [1, 2, 3, 4, 6, 7, 5], her Highest Correct score would be 4. The minimum possible Highest Correct score was zero (i.e., the child's order began with a number other than 1). The maximum possible highest correct score on the Ordering 1–7 tasks was 6, rather than 7, because ordering the numbers 1–6 correctly necessitates ordering the number 7 correctly, by process of elimination. For the same reason, the maximum possible highest correct score on the Ordering 1–3 tasks was 2.

What Comes After. This task assessed children's knowledge of the next number in the count sequence (Davidson et al., 2012). The experimenter asked, "What comes after N? N + 1 or N + 2?" On half of the trials, the order of presentation of the responses was reversed, i.e., "N + 2 or N + 1?" There were 4 trials: two trials each of "What comes after 4?" and "What comes after 5?"

2.3. Procedure

Each child participated in six one-on-one sessions with an experimenter, consisting of one pretest session, four training sessions, and one posttest session. Sessions took place in a quiet area of the child's preschool. All sessions took place on separate days (with the exception of 2 children who received Training 3 and Training 4 on the same day due to time constraints). The average time from pretest to posttest was 19.0 days (SD = 5.8, range = 10–39 days). The number of days from pretest to posttest did not differ between conditions, $t(125) = 0.06$, $p = 0.95$, nor did it differ for 3–4-knowers versus CP-knowers, $t(125) = 0.55$, $p = 0.59$. On average, the first training session took place 5.9 days after the pretest (SD = 4.6, range = 1–20 days). Subsequent training sessions were spaced 3–4 days apart, on average. The posttest took place an average of 3.5 days after the final training (SD = 2.4 days, range = 1–12 days).[3]

Pretest and Posttest. The What Comes After and Give-N tasks were presented first, in that order (except in rare cases when the child was unwilling to participate, other tasks were presented first and then the experimenter returned to these tasks). The order of the Unit and Ordering tasks was randomized across children. Several additional numerical tasks were assessed at pretest and posttest as part of a larger study of numerical development (Gunderson, Spaepen, Gibson, Goldin-Meadow, & Levine, 2015; Gunderson, Spaepen, & Levine, 2015).

Training. Children were randomly assigned, within knower-level, to the Experimental or Control condition. The pretest and posttest were identical in each condition; the four training sessions differed between conditions. Scripts for both conditions are provided in Appendix B. Both conditions involved exposure to the numbers 1, 2, 3, 4, 5, and 6. In both conditions, the child and experimenter practiced counting dots and beads, and the experimenter provided feedback, counting each set size (set sizes 1 to 6) three times per session. In addition, in the Experimental condition, the child and experimenter labeled cardinal values, the different set sizes of beads were available simultaneously and were spatially aligned from lowest to highest, and the experimenter provided explicit instruction that the next number is one more than the previous number.

3. Results

3.1. Data analysis plan

Our primary outcome measures were Unit task performance, which measures explicit successor principle knowledge, and accuracy of Ordering 1–7 Dots and Ordering 1–7 Gestures, which test exact ordering skill. In both cases, we predicted that CP-knowers, but not 3–4-knowers, would improve over time.

Our data analyses proceed in four steps. (1) We explore whether Condition affected children's performance on our outcome measures. (2) We examine *successor principle knowledge* based on Unit task performance, looking at pretest performance by knower-level, and change from pretest to posttest as a function of CP-knowledge. We also compare Unit task performance to a related task that measures knowledge of the order of the count list, but not succession (What Comes After). (3) We examine *exact ordering knowledge*, starting with our primary measures (accuracy on Ordering 1–7 Dots and Gestures), then moving on to examine partially correct performance on these tasks (highest correct on Ordering 1–7 Dots and Gestures) and performance on Ordering 1–3 Dots and Gestures. As with successor principle knowledge, we examine pretest knowledge across knower-levels, as well as change from pretest to posttest as a function of CP-knowledge. (4) We explore relations between improvement on succession and exact ordering over time.

We focus our analyses of change over time on the subgroups of CP-knowers whose pretest scores on the successor principle and exact ordering tasks were the same as the 3–4-knowers' scores, providing a strong test of the importance of cardinal principle knowledge *per se* in learning the successor principle.

The data for this project are available on Open Science Framework at osf.io/xuq9h (Spaepen et al., 2018).

3.2. Preliminary analyses: relation of training condition to learning

We first examined whether training condition (Experimental vs. Control) was related to change over time in our key DVs (successor principle knowledge and exact ordering), or on any other DVs (What Comes After, additional measures of ordering knowledge, or knower-level). As detailed in Appendix C, we found no evidence that the training Conditions differed in their effect on any measure. Because of the lack of differences between training Conditions, we excluded Condition from our subsequent analyses to reduce model complexity.

3.3. Main analyses: successor principle knowledge

We next examined successor principle knowledge as measured by performance on the Unit task. We examined pretest knowledge among

[2] We also asked whether participants might show some knowledge of "later-is-greater" on the Ordering 1–7 task, beyond the knowledge they displayed on our measure of Highest Correct. For example, if a child's order was [1, 3, 2, 4, 6, 5, 7], their Highest Correct score of 1 might not capture use of an approximate later-is-greater strategy, with some error. To assess this possibility, we examined the slope relating participants' actual response to the correct response, for those numbers beyond their Highest Correct number (see Appendix G). In this example, the slope relating numbers above their Highest Correct to the correct sequence (i.e., relating [3, 2, 4, 6, 5, 7] to [2, 3, 4, 5, 6, 7]) would be positive (slope = 0.89). We did not find any evidence that children showed use of "later-is-greater" ordering strategies beyond their Highest Correct number at either pretest or posttest (i.e., their slopes were not significantly greater than zero). We therefore use Highest Correct as our fine-grained measure of ordering performance (in addition to our strict measure, Accuracy).

[3] Because we found no effects of Condition on change from pretest to posttest (see Appendix A), we explored the possibility that time-related development (unrelated to the training) was responsible for the changes from pretest to posttest among low-knowledge CP-knowers. If so, we might expect that the number of days between pretest and posttest (which ranged from 10 to 39 days) would predict children's improvement from pretest to posttest. On the Unit task, a mixed-effects ANCOVA with Time (Pretest, Posttest) as a factor and Age and Number of Days as covariates found no significant effect of Number of Days ($F(1, 66) = 0.15$, $p = .700$, $\eta_p^2 = .002$) and no Time x Number of Days interaction ($F(1, 66) = 1.29$, $p = .261$, $\eta_p^2 = .019$). For the Ordering 1–7 Dots and Gestures tasks, we conducted logistic regressions predicting posttest scores from Age and Number of Days among low-ordering-knowledge CP-knowers (note that pretest scores were all zero). Number of Days was not a significant predictor of posttest score on Ordering 1–7 Dots (B = −0.06, SE = 0.07, $p = 0.383$) or Ordering 1–7 Gestures (B = −0.08, SE = 0.06, $p = 0.190$). Thus, we found no evidence that the length of time between pretest and posttest accounted for the improvement in low- knowledge CP-knowers' performance.

Table 2
Pretest task performance by knower-level.

	Three-knowers (n = 20) M[a] (SD)	Four-knowers (n = 12) M[a] (SD)	CP-knowers (n = 95) M[a] (SD)	Test of difference between all knower-levels[b]	3-knowers vs. 4-knowers[c]	3-knowers vs. CP-knowers[c]	4-knowers vs. CP-knowers[c]
Unit task accuracy	0.48 (0.16)	0.50 (0.12)	0.60*** (0.24)	*n.s*	*n.s.*	*	*
What Comes After accuracy	0.53 (0.20)	0.48 (0.27)	0.77*** (0.27)	***	*n.s.*	***	**
Ordering 1–7							
Dots Accuracy	0.05**	0.00	0.37***	**	*n.s.*	**	*
Gestures Accuracy	0.00	0.00	0.44***	***	*NA*	***	**
Dots Highest Correct (max. 6)	0.80 (1.64)	0.75 (0.97)	3.77 (2.12)	***	*n.s.*	***	***
Gestures Highest Correct (max. 6)	0.55 (0.83)	0.25 (0.87)	4.08 (2.31)	***	*n.s.*	***	***
Ordering 1–3							
Dots Accuracy	0.25	0.58**	0.86***	***	*n.s.*	***	*
Gestures Accuracy	0.20	0.42*	0.84***	***	*n.s.*	***	***
Dots Highest Correct (max. 2)	0.95 (0.76)	1.33 (0.89)	1.78 (0.59)	***	*n.s.*	***	*
Gestures Highest Correct (max. 2)	0.65 (0.81)	1.00 (0.95)	1.75 (1.62)	***	*n.s.*	***	***

Notes. For tasks with a binary outcome (i.e., Ordering accuracy), the percent of participants who were correct is reported.
[a] Asterisks indicate significance of comparison to chance (note: not calculated for Ordering Highest Correct). Comparisons to chance for the Unit and What Comes After tasks were one-sample t-tests versus 50%. Comparisons to chance for Ordering 1–7 Dots & Gestures accuracy were binomial tests versus 0.02%. Comparisons to chance for Ordering 1–3 Dots & Gestures accuracy were binomial tests versus 16.7%.
[b] Unit task and What Comes After task compared using ANOVA. Ordering Accuracy compared using Pearson's χ^2. Ordering Highest Correct was compared non-parametrically using Kruskal-Wallis H test.
[c] Unit task and What Comes After task compared using independent-samples t-tests. Ordering Accuracy compared using Pearson's χ^2. Ordering Highest Correct was compared non-parametrically using Mann-Whitney U test. $p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

all participants as a function of knower-level, and examined change over time among low-successor-knowledge participants as a function of CP knowledge. We also examined the What Comes After task, given that it has a similar response format, and requires some of the same knowledge, as the Unit task (i.e., the ability to identify the next number in the count sequence without starting from "one"), but does not require conceptual understanding of the successor principle.

3.3.1. Successor principle: Unit task

Pretest performance. Pretest performance on the Unit task was consistent with prior work (see Table 2). Only CP-knowers' scores were significantly above chance, replicating Sarnecka and Carey (2008). CP-knowers performed significantly better than 3- and 4-knowers, who did not differ significantly from each other. However, there was variability in performance among CP-knowers (for histograms, see Appendix Fig. A2). Very few CP-knowers (16%, n = 15) were at ceiling on the Unit task, and a large majority (73%, n = 69) had scores that were in the same range as the 3–4-knowers (i.e., 17% to 67% correct on the Unit task), also replicating prior work (Davidson et al., 2012).

Effect of knower-group on learning. We compared 3–4-knowers (n = 32) to those low-successor-knowledge CP-knowers whose pretest scores on the Unit task matched the range of 3–4-knowers' scores (n = 69) (Fig. 1); after matching the groups on the range of pretest Unit task scores, the mean pretest Unit task scores did not significantly differ, $t(99) = .280$, $p = .780$. A 2 (Time: Pretest, Posttest) × 2 (Knower-Group: 3–4-knowers, low-successor-knowledge CP-knowers) mixed-effects ANCOVA, with Age as a covariate,[4] showed a significant Time × Knower-Group interaction ($F(1, 99) = 4.33$, $p = 0.040$, $\eta_p^2 = 0.042$). The main effects of Time, Knower-Group, and Age did not reach statistical significance ($ps > 0.05$). Three- and 4-knowers did not significantly improve from pretest ($M = 0.49$, $SD = 0.15$) to posttest

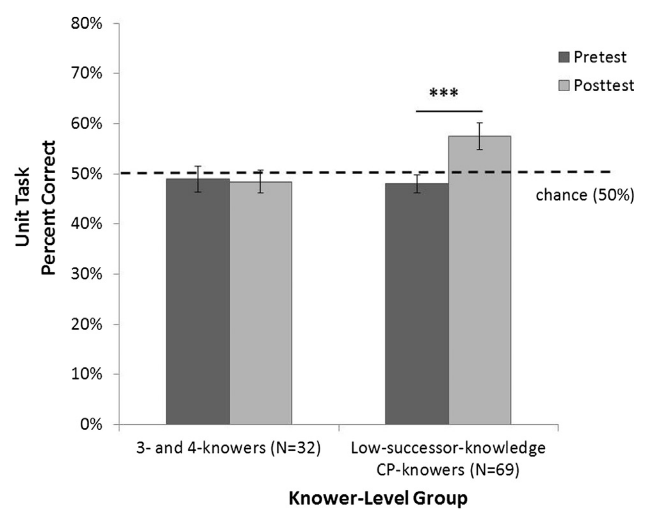

Fig. 1. Performance on the Unit task at Pretest and Posttest among 3–4-knowers and low-successor-knowledge CP-knowers.

($M = 0.48$, $SD = 0.13$), $t(31) = 0.14$, $p = 0.893$, $d = -.02$. In contrast, low-successor-knowledge-CP-knowers significantly improved in successor principle knowledge from pretest ($M = 0.48$, $SD = 0.15$) to posttest ($M = 0.57$, $SD = 0.22$), $t(68) = 3.46$, $p < .001$, $d = 0.42$.[5] In other words, only children who already understood the cardinal principle improved over time in successor principle knowledge.

Effects of other numerical skills on learning. We examined whether other measured numerical skills were a better predictor of growth on

[4] We conducted preliminary analyses for all ANCOVAs and GEEs involving Time (Unit, What Comes After, and Ordering tasks) with the Age x Time interaction included. The Age x Time interaction was not statistically significant in any model. To reduce the complexity of the models, we excluded Age x Time interactions.

[5] The "low-successor-knowledge CP-knowers" in this analysis consisted of all 69 CP-knowers who scored in the 3–4-knowers' range on the Unit task at pretest. Forty of these children also got zero correct on the Ordering 1–7 tasks and thus arguably have very low levels of prior successor principle knowledge. Importantly, these 40 CP-knowers still improved from pretest to posttest on the Unit task, $t(39) = 2.07$, $p = .045$, $d = 0.33$, whereas the 3–4 knowers did not.

Table 3
Pretest and Posttest Performance on Ordering Measures, by Knower-Group.

	Three- and four-knowers (n = 31)		Low-ordering-knowledge CP-knowers (n = 46)		Pretest to posttest change		
	M[a] (SD)	Median (range)	M[a] (SD)	Median (range)	Group diff.[b]	3–4-knowers	Low-ordering-knowledge CP-knowers
Pretest Ordering 1–7							
Dots Accuracy	0.00	NA	0.00	NA	*NA*		
Gestures Accuracy	0.00	NA	0.00	NA	*NA*		
Dots Highest Correct (max. 6)	0.61 (1.05)	0.00 (0–4)	2.11 (1.48)	2.00 (0–5)	***		
Gestures Highest Correct (max. 6)	0.35 (0.71)	0.00 (0–3)	2.30 (2.00)	2.00 (0–5)	***		
Pretest Ordering 1–3							
Dots Accuracy	0.35**	NA	0.72***	NA	**		
Gestures Accuracy	0.29	NA	0.72***	NA	***		
Dots Highest Correct (max. 2)	1.06 (0.81)	1.00 (0–2)	1.54 (0.78)	2.00 (0–2)	**		
Gestures Highest Correct (max. 2)	0.77 (0.88)	0.00 (0–2)	1.54 (0.78)	2.00 (0–2)	***		
Posttest Ordering 1–7							
Dots Accuracy	0.00	NA	0.24	NA	**	*NA*	***
Gestures Accuracy	0.03**	NA	0.37	NA	***	*n.s.*	***
Dots Highest Correct (max. 6)	1.03 (1.30)	1.00 (0–5)	3.83 (1.69)	4.00 (0–6)	***	*	***
Gestures Highest Correct (max. 6)	0.90 (1.62)	0.00 (0–6)	4.04 (2.18)	5.00 (0–6)	***	*	***
Posttest Ordering 1–3							
Dots Accuracy	0.45***	NA	0.89***	NA	***	*n.s.*	*
Gestures Accuracy	0.19	NA	0.85***	NA	***	*n.s.*	*n.s.*
Dots Highest Correct (max. 2)	1.03 (0.95)	1.00 (0–2)	1.85 (0.47)	2.00 (0–2)	***	*n.s.*	*
Gestures Highest Correct (max. 2)	0.65 (0.80)	0.00 (0–2)	1.74 (0.65)	2.00 (0–2)	***	*n.s.*	*n.s.*

Notes. Analyses include 3–4-knowers (n = 31) and low-ordering-knowledge CP-knowers (n = 46, matched to 3–4-knowers based on pretest Ordering 1–7 Dots and Gestures accuracy). *p < .05, **p < .01, ***p < .001.

[a] Significance of comparison to chance (Ordering 1–7 Accuracy: binomial tests versus 0.02%; Ordering 1–3 Accuracy: binomial tests for Ordering 1–3 versus 16.7%, for Ordering 1–7 versus .02%; not calculated for Highest Correct).

[b] Accuracy: Pearson's χ^2; Highest Correct: non-parametric Mann-Whitney U test. [c] Accuracy: McNemar test; Highest Correct: non-parametric Wilcoxon Signed-Rank test.

the Unit task than CP knowledge (for 3–4-knowers and matched, low-knowledge CP-knowers). To test this, we repeated our ANCOVAs, with Unit task performance as the DV. We included Knower-Group, Age, Time, and Time × Knower-Group as predictors, and added each Pretest Task and Time × Pretest Task interaction in separate models. In these analyses, a significant Time × Pretest Task interaction would indicate that higher pretest task performance was related to greater Unit task growth over time, over and above the demonstrated impact of Knower-Group on Unit task growth. However, none of the Time × Pretest Task interactions were significant (ps > .05) in separate models for the following Pretest tasks: What Comes After, Ordering 1–7 Accuracy (averaged across Dot and Gesture items), and Ordering 1–3 Accuracy (averaged across Dot and Gesture items).

In sum, no other measured numerical skills significantly predicted growth in successor principle knowledge, after cardinal principle knowledge was accounted for.

3.3.2. Comparison task: What Comes After

Pretest performance. The pattern of pretest performance on the What Comes After task was similar to the Unit task (see Table 2): CP-knowers performed above chance, whereas 3- and 4-knowers did not. CP-knowers performed significantly better than both 3- and 4-knowers, who did not significantly differ from each other. At pretest, CP-knowers performed significantly better on the What Comes After task than on the Unit task ($t(93) = 6.14$, p < .001, $d = 0.63$). In contrast, 3- and 4-knowers' performance was not significantly different between the two tasks (ps > .50) and their performance did not differ from chance on either task.

Effects of knower-group on learning. We compared growth on the What Comes After task among 3–4-knowers versus low-successor-knowledge CP-knowers to determine whether CP-knowers' growth was specific to successor principle knowledge, or extended to other similar numerical tasks. A Time × Knower-Group mixed-effects ANCOVA with age as a covariate revealed a significant main effect of Knower-Group ($F(1, 98) = 19.86$, p < .001, $\eta_p^2 = 0.169$), but no significant Time × Knower-Group interaction ($F(1, 99) = .003$, $p = .958$, $\eta_p^2 = 0.000$). Three- and 4-knowers did not significantly improve on What Comes After from pretest ($M = 0.51$, $SD = 0.22$) to posttest ($M = 0.52$, $SD = 0.18$), $t(31) = 0.17$, $p = 0.869$, $d = 0.03$. Similarly, low-successor-knowledge-CP-knowers also did not improve from pretest ($M = 0.72$, $SD = 0.27$) to posttest ($M = 0.74$, $SD = 0.28$) on What Comes After, $t(68) = 0.331$, $p = .742$, $d = 0.04$. Thus, in contrast to the Unit task, where low-successor-knowledge CP-knowers significantly improved over time, on the structurally similar What Comes After task, neither low-successor-knowledge CP-knowers nor 3–4-knowers improved over time. This finding suggests that the CP-knowers' growth over time was specific to conceptual understanding of the successor principle, and could not be attributed to an increase in procedural fluency with the count list.

3.4. Main analyses: Exact Ordering

We next examined exact ordering performance—measured by accuracy on the Ordering 1–7 Dots and Gestures tasks—at pretest (comparing all knower-levels) and in terms of change from pretest to posttest (comparing 3–4-knowers to CP-knowers with similar levels of pretest ordering knowledge).

Pretest performance. Because each task had only one trial, accuracy scores for each child were either 0 (incorrect) or 1 (correctly ordering all sets up to 7); we report the percent of children who were correct at pretest by knower-level in Table 2. We conducted a binomial Generalized Estimating Equation (GEE) on pretest Ordering 1–7 Accuracy, with Task (Dots, Gestures) as a within-subjects factor, Knower-Group (3–4-knower versus CP-knower) as a between-subjects factor (it was not possible to analyze 3- and 4-knowers separately due to floor performance by 4-knowers), and Age as a covariate. Knower-Group was a significant predictor of pretest accuracy ($B = 3.53$, $SE = 1.03$, $p = .001$), with CP-knowers outperforming 3- and 4-knowers; neither Task ($B = 0.26$, $SE = 0.20$, $p = .200$) nor Age ($B = 0.58$, $SE = 0.37$, $p = .120$) were significant predictors.

Although performing better than 3–4-knowers, CP-knowers again were a diverse group. Only 30% of CP-knowers ($n = 28$) were able to correctly order both 1–7 Dots and 1–7 Gestures, whereas 48% ($n = 46$) were incorrect on both, performing like the 3–4-knowers who were almost uniformly incorrect on both. Thus, at pretest, 3–4-knowers showed no evidence of exact ordering, whereas nearly half of CP-knowers were able to exactly order quantities up to 7.

Effect of knower-group on learning. Thirty-one of the 32 3–4-knowers got neither item correct on the Ordering 1–7 Dot and Gesture tasks at pretest. We compared these 31 low-ordering-knowledge 3–4-knowers to the 46 low-ordering-knowledge CP-knowers who also got neither Ordering 1–7 task correct at pretest. (Data for all CP-knowers are provided in Appendix D.) See Table 3 for raw pretest and posttest scores by task and knower-group.

We conducted a binomial GEE predicting posttest accuracy from Task, Knower-Group, and Age (it was not possible to include Time as a within-subjects factor due to lack of variance at pretest). Knower-Group was a significant predictor of posttest Ordering 1–7 accuracy ($B = 3.28$, $SE = 1.05$, $p = .002$), with low-ordering-knowledge CP-knowers outperforming 3- and 4-knowers; neither Task ($B = 0.70$, $SE = 0.43$, $p = .104$) nor Age (B = 0.11, SE = 0.38, $p = .774$) were significant predictors (see Fig. 2).

Although matched to 3–4-knowers in terms of Ordering 1–7 Accuracy, subsequent analyses (see Table 3 and below) suggested that

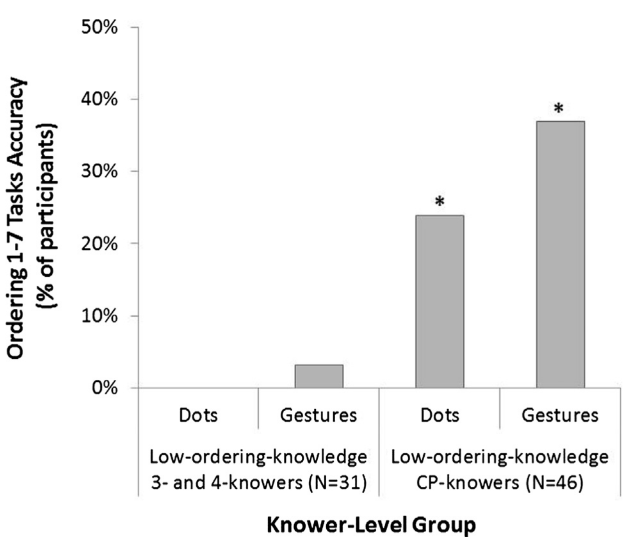

Fig. 2. Percent of children who were accurate on the Posttest Ordering 1–7 Dots and Ordering 1–7 Gestures tasks among low-ordering-knowledge 3–4-knowers and low-ordering-knowledge CP-knowers. Note: Asterisks indicate significance of change from pretest accuracy (0.0% correct for all groups) to posttest accuracy. * = $p < .001$.

the $n = 46$ low-ordering-knowledge CP-knowers were more likely to order some items correctly at pretest (median Highest Correct score = 2) than 3–4-knowers (median Highest Correct score = 0). To assess whether these CP-knowers' partial knowledge of ordering could explain our results, we examined those CP-knowers who were comparable at pretest to 3–4-knowers in terms of Ordering 1–7 Accuracy *and* Highest Correct scores (i.e., Accuracy = 0 and Highest Correct ≤ 1 on both Ordering 1–7 Dots and Gestures). Fifteen CP-knowers (of $n = 95$, 15.8%) met these criteria and were considered "very-low-ordering-knowledge CP-knowers" at pretest. Twenty-seven 3–4-knowers (of $n = 32$, 84.4%) met these criteria. Importantly, very-low-ordering-knowledge CP-knowers significantly outperformed very-low-ordering-knowledge 3–4-knowers on posttest Ordering 1–7 Dots Accuracy (CP-knowers: 20.0% correct; 3–4-knowers: 0.0% correct, $\chi^2(1) = 5.82$, $p = .016$) and posttest Ordering 1–7 Gestures Accuracy (CP-knowers: 33.3% correct; 3–4-knowers: 0.0% correct, $\chi^2(1) = 10.22$, $p = .001$). (Floor performance by 3–4-knowers at posttest precluded use of binomial GEE for these analyses.) Further, very-low-ordering-knowledge CP-knowers ($n = 15$) did not significantly differ from moderately-low-ordering-knowledge CP-knowers ($n = 31$ who at pretest were not accurate on either task, but had a Highest Correct score greater than 1 on at least one task) at posttest on Ordering 1–7 Dots (moderately low: 25.8% correct; very low: 20.0% correct; $\chi^2(1) = 0.19$, $p = .665$) or Ordering 1–7 Gestures (moderately low: 38.7% correct; very low: 33.3% correct; $\chi^2(1) = 0.13$, $p = .723$). These results suggest that moderately-low-ordering-knowledge CP-knowers' partial ordering knowledge, as reflected in Highest Correct scores greater than 1, was not driving their improvement in ordering at posttest.

In summary, the CP-knowers who started out with the same pretest performance on exact ordering as 3–4-knowers (i.e., failure to order exactly at pretest) significantly outperformed 3–4-knowers at posttest.

Effects of other numerical skills on learning. We assessed whether other measured numerical skills at pretest predicted greater improvement over time on Ordering 1–7 Accuracy, after Knower-Group was accounted for. We conducted binomial GEEs predicting Posttest Ordering 1–7 Accuracy, with Task as a within-subjects factor, Knower-Group as a between-subjects factor, and Age as a covariate, and added each pretest task in separate models. None of the pretest task scores were significant predictors of posttest Ordering 1–7 Dots Accuracy ($ps > .05$) in separate models for the following pretest tasks: Unit task, What Comes After, and Ordering 1–3 Accuracy (averaged across Dot and Gesture items).

As for results on the Unit task, no other measured numerical skills significantly predicted gains in Ordering 1–7 Accuracy, after Knower-Group was accounted for.

3.5. Additional analyses: ordering tasks

The preceding analyses were conducted on accuracy on the Ordering 1–7 tasks (i.e., percent of children who ordered all items correctly up to 7). To explore children's ability to order quantities further, we examined three additional measures, all of which capture some aspects of ordering skill, but are less strict than our main measure of exact ordering (Ordering 1–7 accuracy). These additional measures are highest correct scores on Ordering 1–7 (i.e., the highest set size before the child's first error); accuracy on Ordering 1–3; and highest correct scores on Ordering 1–3.

3.5.1. Ordering 1–7 highest correct

Pretest performance. We used the highest correct score to gain a more fine-grained view of ordering performance (see Appendix A Fig. A3 for histograms). We treated highest correct scores as ordinal in all analyses. The possible range for highest correct scores on the Ordering 1–7 tasks was 0 (no items correctly ordered) to 6 (all items correctly ordered). See Table 2 for unadjusted pretest scores by knower-level.

We conducted an ordinal GEE on Pretest Ordering 1–7 Highest

Correct, with Task (Dots, Gestures) as a within-subjects factor, Knower-Level (3-knower, 4-knower, CP-knower) as a between-subjects factor, and Age as a covariate. CP-knowers had significantly higher scores than 3-knowers ($B = 2.57$, $SE = 0.48$, $p < .001$) and 4-knowers ($B = 3.01$, $SE = 0.55$, $p < .001$), who did not significantly differ from one another ($B = -0.44$, $SE = 0.64$, $p = .492$). Age was also a significant predictor ($B = 0.70$, $SE = 0.32$, $p = .027$), whereas Task was not ($B = 0.12$, $SE = 0.14$, $p = .392$).

Descriptively, for both tasks at pretest, the majority of 3- and 4-knowers had a Highest Correct score of zero, indicating that their order began with a quantity other than one (62.5% for Ordering 1–7 Dots; 71.9% for Ordering 1–7 Gestures). In contrast, < 15% of CP-knowers had a Highest Correct score of zero at pretest on each task (8.4% on Ordering 1–7 Dots; 14.7% on Ordering 1–7 Gestures). In sum, at pretest on the Ordering 1–7 tasks, most 3- and 4-knowers failed to order quantities correctly within their known range, and instead often started their order with a quantity other than one, or put only the first quantity in the correct position. In contrast, more than half of CP-knowers at pretest ordered the quantities 1–4 correctly before making a mistake.

Effect of knower-group on learning. Ordering 1–7 Highest Correct scores at pretest and posttest by knower-level and task is displayed in Table 3. We conducted an ordinal GEE with Time (pretest, posttest) and Task (Dots, Gestures) as within-subjects factors and Knower-Group (3–4-knower, low-ordering-knowledge CP-knower) as a between-subjects factor, and Age as a covariate (see Table 4, Model 1). In these analyses, a significant Time × Knower-Group interaction was expected to show that low-ordering-knowledge CP-knowers improved more than 3–4-knowers from pretest to posttest. Indeed, we found a significant Time × Knower-Group interaction ($B = 0.88$, $SE = 0.37$, $p = .017$; Table 4, Model 1). Follow-up GEEs within each Knower-Group indicated that the effect of Time was significant within both 3–4-knowers ($B = 0.80$, $SE = 0.30$, $p = .007$; Table 4, Model 2) and low-ordering-knowledge CP-knowers ($B = 1.62$, $SE = 0.26$, $p < .001$; Table 4, Model 3). Thus, both knower-groups improved over time, and the significant Time × Knower-Group interaction indicates that low-ordering-knowledge CP-knowers improved significantly more than 3–4-knowers.

It is important to note that even though both groups improved, the improvement in each group was qualitatively different. Three- and four-knowers' median highest correct scores at pretest were zero for both Dots and Gestures, and by posttest changed to a median of 1 for Dots and remained at 0 for Gestures. Thus, on Ordering 1–7 Dots they improved from putting none of the items in order, to putting the first item in the correct place (arguably not truly ordering). In contrast, low-ordering-knowledge CP-knowers improved from putting a median of 2 quantities in order at pretest on both Ordering 1–7 Dots and Gestures, to a median of 4 quantities in order for Dots and 5 quantities in order for Gestures at posttest.

3.5.2. Ordering 1–3 accuracy

Pretest performance. Accuracy on Ordering 1–3 at pretest, by task and knower-level, is shown in Table 2. We conducted a binomial GEE on pretest Ordering 1–3 accuracy with Task (Dots, Gestures) as a within-subjects factor, knower-level (3-knower, 4-knower, CP-knower) as a between-subjects factor), and Age as a covariate. Knower-level significantly related to pretest performance, with CP-knowers outperforming 3-knowers ($B = 2.83$, $SE = 0.53$, $p < .001$) and 4-knowers ($B = 1.70$, $SE = 0.50$, $p < .001$), who did not significantly differ from one another ($B = 1.14$, $SE = 0.60$, $p = .060$). Neither age ($B = 0.30$, $SE = 0.42$, $p = .471$) nor task ($B = -0.27$, $SE = 0.26$, $p = .297$) significantly predicted pretest Ordering 1–3 accuracy.

We also compared performance to chance (chance = 16.7%) using binomial tests. CP-knowers (Dots: 86.3% correct, $p < .001$; Gestures: 84.2% correct, $p < .001$) and 4-knowers (Dots: 58.3% correct, $p = .001$; Gestures: 41.7% correct, $p = .036$) were significantly above chance at pretest, whereas 3-knowers were not (Dots: 25.0% correct, $p = .231$; Gestures: 20.0% correct, $p = .433$). Thus, at pretest, 3-knowers showed no evidence of being able to order quantities within the range of their known numbers, even when those three quantities were the only ones presented. Four-knowers showed some ability to order 1–3 quantities (performing above chance), but still only about half of 4-knowers were able to order these quantities correctly. In contrast, CP-knowers were approaching ceiling on this task.

Effect of knower-group on learning. Performance at pretest and posttest among 3–4-knowers and low-ordering-knowledge CP-knowers (the same matched groups as for the Ordering 1–7 tasks) is shown in Table 3. We conducted a binomial GEE on Ordering 1–3 accuracy with Time (Pretest, Posttest) and Task (Dots, Gestures) as within-subjects factors, Knower-Group (3–4-knowers, low-ordering-knowledge CP-knowers) as a between-subjects factor, and Age as a covariate. The Time × Knower-Group interaction was not significant ($B = 0.99$, $SE = 0.52$, $p = .054$), indicating that 3–4-knowers did not significantly differ from low-ordering-knowledge CP-knowers in their change from pretest to posttest. There were significant main effects of Knower-Group ($B = 1.49$, $SE = 0.43$, $p < .001$), indicating that low-ordering-knowledge CP-knowers outperformed 3–4-knowers, and of Task ($B = -0.44$, $SE = 0.20$, $p = .029$), indicating that accuracy was higher for Ordering 1–3 Dots than Gestures. There were no significant main effects of Time ($B = 0.00$, $SE = 0.38$, $p = 1.00$) or Age ($B = 0.68$, $SE = 0.37$, $p = .064$). Thus, in terms of accuracy on Ordering 1–3, the Knower-Groups differed in performance at both time points, but neither group significantly improved over time.

3.5.3. Ordering 1–3 highest correct

Pretest performance. Descriptive analyses of pretest performance by knower-level and task are in Table 2; see Appendix A for histograms. We conducted an ordinal GEE on pretest Ordering 1–3 Highest Correct with Task (Dots, Gestures) as a within-subjects factor, Knower-Level (3-

Table 4
Effects of Time and Knower-Group on Ordering Highest Correct (HC): Ordinal GEE results.

	Ordering 1–7 (HC)			Ordering 1–3 (HC)		
	Model 1 All (N = 77) B (SE)	Model 2 3-4-knowers (n = 31) B (SE)	Model 3 Low-ordering-knowledge CP-knowers (n = 46) B (SE)	Model 4 All (N = 77) B (SE)	Model 5 3-4-knowers (n = 31) B (SE)	Model 6 Low-ordering-knowledge CP-knowers (n = 46) B (SE)
Age	0.54 (0.30)	0.73 (0.51)	0.41 (0.36)	0.57 (0.34)	0.56 (0.41)	0.60 (0.58)
Task (Dots, Gestures)	−0.03 (0.17)	−0.52 (0.28)	0.23 (0.22)	−0.44 (0.18)*	−0.73 (0.26)**	−0.14 (0.26)
Time (Pretest, Posttest)	0.75 (0.27)**	0.80 (0.30)**	1.62 (0.26)***	−0.14 (0.30)	−0.13 (0.31)	0.97 (0.35)**
Knower-Group	2.01 (0.45)***	–	–	1.32 (0.40)***	–	–
Time × Knower-Group	0.88 (0.37)*	–	–	1.13 (0.47)*	–	–

Note. Analyses include participants matched to 3–4-knowers' performance on pretest Ordering 1–7 Dots and Gestures (i.e., were inaccurate on both Ordering 1–7 Dots and Ordering 1–7 Gestures; $n = 31$ 3–4-knowers and $n = 46$ CP-knowers).
*$p < .05$, **$p < .01$, ***$p < .001$

knower, 4-knower, CP-knower) as a between subjects factor, and Age as a covariate. Consistent with the accuracy scoring of this task, we found that CP-knowers significantly outperformed both 3-knowers ($B = 2.39$, $SE = 0.46$, $p < .001$) and 4-knowers ($B = 1.75$, $SE = 0.51$, $p < .001$), who did not significantly differ ($B = 0.63$, $SE = 0.55$, $p = .247$). Task ($B = -0.39$, $SE = 0.24$, $p = .112$) and Age ($B = 0.33$, $SE = 0.38$, $p = .386$) were not significant predictors of pretest Ordering 1–3 Highest Correct. As with accuracy, highest correct scores on ordering 1–3 indicated that CP-knowers were near ceiling on ordering 3 items, 4-knowers showed some emerging ability to order 3 items, and 3-knowers showed no ability to order 3 items.

Effect of knower-group on learning. We again compared 3–4-knowers to low-ordering-knowledge CP-knowers to determine which group(s) improved over time on ordering 1–3 highest correct (see Table 3 for descriptive analyses). We conducted ordinal GEEs on Ordering 1–3 Highest Correct, with Time (pretest, posttest) and Task (Dots, Gestures) as within-subjects factors, Knower-Group as a between-subjects factor, and Age as a covariate (Table 4, Model 4). We found a significant Time × Knower-Group interaction ($B = 1.13$, $SE = 0.47$, $p = .015$), indicating that 3–4-knowers differed from low-ordering-knowledge CP-knowers in how much they improved from pretest to posttest on Ordering 1–3 Highest Correct. Follow-up analyses within Knower-Group indicated there was no significant effect of Time among 3–4-knowers ($B = -0.13$, $SE = 0.31$, $p = .682$; Table 4, Model 5), but there was a significant effect of Time among low-ordering-knowledge CP-knowers ($B = 0.97$, $SE = 0.35$, $p = .006$; Table 4, Model 6). Thus, we found no evidence that 3–4-knowers improved over time in their ability to order sets of 1–3 items; in contrast, CP-knowers significantly improved over time.

3.6. Additional analyses: relation between improvement on succession and exact ordering

Finally, we explored whether there was any relation in improvement on our two main DVs (succession and exact ordering). That is, did children who improved on one task improve on both, and were there any hints of an order of acquisition? We examined this question among CP-knowers, the only group who showed improvement over time, and focused on CP-knowers who were low-knowledge at pretest on both tasks, and therefore had room to grow on both. (For supplementary analyses of the relations among pretest numerical skills among all CP-knowers, see Appendix E.)

We examined the $n = 40$ CP-knowers who, at pretest, were categorized as both low-ordering-knowledge *and* low-successor-knowledge, and therefore had room for growth on both tasks. We categorized children as improvers on exact ordering if they were accurate on either Ordering 1–7 Dots or Gestures at posttest. We categorized children as improvers on succession if their posttest Unit task accuracy minus pretest accuracy was greater than zero. Of the $n = 40$ initially low-ordering-and-successor-knowledge CP-knowers, 50.0% were improvers on exact ordering and 45.0% were improvers on succession. Improvers on exact ordering were marginally more likely than non-improvers to also be improvers on succession, $r(38) = 0.30$, $p = .059$. Among low-ordering-and-low-successor-knowledge CP-knowers, 35.0% did not improve on either succession or exact ordering, 30.0% improved on both, and 35.0% improved on only one task. However, there was no evidence that improvement on one task preceded improvement on another: of those who improved on only one task, 42.9% improved on succession but not ordering, and 57.1% improved on ordering but not succession.

Thus, among CP-knowers who did not start out understanding the successor principle or exact ordering, improvement was (marginally) related across tasks, but there was no evidence that one piece of knowledge preceded the other.

4. Discussion

We provide the first evidence that CP-knowers whose initial knowledge of succession and exact numerical ordering is as low as 3–4-knowers' improve over time in their understanding of these exact numerical relations; 3- and 4-knowers do not. Our data are the strongest evidence to date that knowledge of exact numerical relations (i.e., succession and exact ordering) does not develop at the same time as the cardinal principle, but rather develops *after* children learn the cardinal principle. This is true both on a task requiring children to understand how the cardinal meanings of number words are ordered (the *Unit* task), and on tasks that required children to understand that the cardinal values themselves can be put in exact numerical order (the *Ordering* tasks). Indeed, our findings suggest that the cardinal principle may even be a "gatekeeper" skill, which children need to learn before they can learn about these exact numerical relations. CP-knowers improved over time on a task designed to assess understanding of succession within numbers up to 8 (Unit task)[6], and also improved on tasks tapping exact ordering ability (Ordering 1–7 Dots and Ordering 1–7 Gestures).

By comparing 3–4-knowers to a group of CP-knowers matched on prior knowledge of exact numerical relations, we isolate knowledge of the cardinal principle as the variable most likely to explain group differences in learning over time. Our analyses controlled for child age, indicating that CP knowledge predicted learning over and above children's age. In addition, our analyses suggest that other factors, such as performance on other numerical tasks included in the study, did not predict learning over time over and above knowing the cardinal principle.

Our results suggest that CP-knowers do not need any prior knowledge of succession or exact numerical ordering to learn from short-term instruction or practice. Simply *being* a CP-knower seems to make children ready to benefit from instruction or practice, even without any discernable prior knowledge of exact numerical relations. Our findings suggest that only children who understand the cardinal principle are ready to learn these new concepts.

4.1. Successor principle knowledge

Replicating previous work, our findings show that CP-knowers, as a group, outperform 3- and 4-knowers on tasks tapping successor principle knowledge, and that some CP-knowers show evidence of successor principle knowledge whereas others do not. Here, for the first time, we show that among those CP-knowers who were matched to 3–4 knowers in terms of their knowledge of the successor principle, CP-knowers learned more over time than 3–4-knowers. Importantly, this pattern of growth was specific to the *Unit* task, which involves knowledge of the successor principle. We did not see improvement over time on the *What Comes After* task, a task that is structurally similar to the *Unit* task but does not require an understanding of the successor principle. In fact, CP-knowers without knowledge of the successor principle started out at pretest with more knowledge of *What Comes After* than 3–4-knowers, suggesting that this fluency with the count list precedes understanding succession and hints at an order of acquisition between learning the cardinal principle, learning the count list flexibly, and learning the successor principle. In other words, children may first learn the cardinal principle, then begin to think more flexibly about their count list (as needed for *What Comes After*), and only then learn the conceptual basis for the successor principle—that adding an item to a set increases its size by exactly one, to the next number in the count list. However, while

[6] Previous research suggests that children do not learn the mathematical axiom that *all* numbers have a successor until significantly after they become cardinal-principle knowers (Cheung, Rubenson, & Barner, 2017). In this paper, we were interested in when children learn a more modest generalization about succession—that each successive cardinal value, within numbers they can count, is exactly one more than the previous one.

low-successor-knowledge CP-knowers' flexible understanding of their count list *may* be important to learning the successor principle, performance on *What Comes After* at pretest did not predict learning of the successor principle over and above children's knowledge of the cardinal principle. We note that this study was not designed to test the contribution of flexible counting in learning the successor principle. Future research could therefore help illuminate not only the order of acquisition between the cardinal principle, flexible count list understanding, and the successor principle, but also whether these skills are causally linked in acquisition (i.e., that children's growing fluency with the count list is necessary for them to learn the successor principle).

Taken together, these data suggest three important findings. First, 3–4-knowers did not perform higher than chance on either the *Unit* task or *What Comes After*, suggesting that their deficits in understanding go deeper than just not understanding succession—they may not yet understand how their count lists work, at least when starting from a number other than "one." Second, succeeding on *What Comes After* does not guarantee success on the *Unit* task, replicating work by Davidson et al (2012). Finally, CP-knowers do not improve on *What Comes After* over time, so improvement on the *Unit* task cannot be explained by children simply figuring out the order of the count list. CP-knowers' growth on the *Unit* task must therefore reflect growth in their conceptual understanding of the successor principle.

4.2. Exact ordering knowledge

Adding to previous findings, our results on children's exact ordering ability mirror the successor principle results—CP-knowers are better, as a group, than 3–4-knowers at pretest on tasks tapping exact numerical ordering, and only CP-knowers improve their exact ordering over time. This pattern holds whether we look at exact accuracy in ordering (the most stringent test of exact ordering ability) or at the highest number children put in order before making a mistake (their highest correct score). While 3–4-knowers did grow a little on their highest correct value from pretest to posttest, they moved from ordering a median of zero cards correctly to a median of 1 card in the correct location in Ordering 1–7 Dots—the median remained zero from pretest to posttest on Ordering 1–7 Gestures. (We stress the limited conclusions one can draw from this growth—placing one card in the correct location certainly does not qualify 3–4-knowers as having learned exact ordering on the basis of number, and may not even qualify strictly as "ordering.") In contrast, low-ordering-knowledge CP-knowers went from ordering a median of 2 cards correctly to ordering a median of 4 or 5 cards correctly, including a significant number of children who learned to order the entire sequence correctly.

Strikingly, 3–4-knowers did not order even very small sets correctly. We did not consider success on our Ordering 1–3 task to be a rigorous measure of exact ordering knowledge simply because the Dot and Gesture Ordering 1–3 tasks could be solved using the non-verbal ANS or parallel-individuation systems. It is therefore particularly surprising that 3-knowers were at chance on this task, and 4-knowers were above chance but still quite poor. Moreover, neither 3- nor 4-knowers improved over time on this task. In other words, 3- and 4-knowers not only failed to order cardinal values within their count list correctly, but they also understood very little about the ordinal relations among the numbers that they did know the cardinal meanings for, and showed no signs of gaining this insight in the short term. One possible explanation is that subset knowers are not reasoning relationally yet about the values of cardinalities they have learned. That is, they learn "one," "two" and "three" as individual words that represent exact cardinal values, but not as a set of words and values that relate to each other. One of the developments that may occur when becoming a CP-knower is a recognition that cardinal values do relate to each other in some way. As the data from the current study show, soon after learning the cardinal principle, they begin to understand that the relation between these values is a + 1 function.

It is important to note that although children were shown the correct placement of the set size 1 after their initial response on the Ordering 1–3 tasks to help them understand what the task was about, we cannot rule out the possibility that 3–4-knowers simply did not understand the task. We do note, however, that our results are significant independent of age, so the patterns we find are not simply an issue of subset-knowers being too young to do serial ordering.

Low-ordering-knowledge CP-knowers did have more knowledge of ordering than 3–4-knowers at pretest—their highest correct number on both Dots and Gestures was higher than 3–4-knowers', and they were far more likely to order 1–3 cards correctly. Thus, it could have been that the CP-knowers who were not yet accurate but were able to order at least 2 or 3 cards correctly drew on this understanding to learn to order all of the cards correctly over time. Instead, we found that CP-knowers who ordered 0 or 1 card correctly at pretest (as 3–4-knowers did) were just as likely to learn to order all of the cards correctly as were CP-knowers who started out ordering more cards correctly—and, importantly, more likely than were 3–4-knowers who started out with the same knowledge. In addition, the ability to order 1–3 correctly did not predict learning over and above being a CP-knower. Again, cardinal principle knowledge seems to be the best predictor of learning over time, rather than any other factor such as partial or incomplete knowledge of numerical ordering.

4.3. Relationship between succession and exact ordering

Our main questions involved the relationship between learning the cardinal principle and learning exact numerical relations. However, we also had an opportunity to learn more about how these measures of children's knowledge of exact numerical relations—succession and exact ordering—are related among CP-knowers over time. Specifically, we examined acquisition patterns among CP-knowers who only improved on one of the two tasks to see if there was any hint of an order of acquisition. For example, if learning exact ordering precedes learning succession, we might expect to find more children who improved on exact ordering but not succession, and very few children who improved on succession but not exact ordering. However, the data did not support a strong order of acquisition—the children who improved on only one task were approximately evenly split, with 57% improving on exact ordering and 43% improving on succession. Thus, it is still an open question as to whether most children go through a predictable order of acquisition with respect to exact ordering and succession. Further research is needed to look into how these skills are related, both during acquisition and after, and how children who succeed on only one of these tasks differ from those who succeed on both.

4.4. Overall conclusions

Our data suggest that initial learning of the cardinal principle is on the one hand a more modest step in children's numerical development than previously believed (Sarnecka & Carey, 2008), but on the other hand is an vital precursor to learning other number concepts. The cardinal principle, taken literally, requires only understanding the relation between counting and a specific set size—that counting from 1 to N results in a set of size N. Critically, cardinal principle knowledge does not require understanding the relations *between* set sizes; in other words, it does not require an understanding of exact numerical relations.

Importantly, however, knowing the cardinal principle does make it easier to learn exact numerical relations—in fact, our data suggest that it may even be a *necessary* prerequisite to learning them. Why might children need to understand the cardinal principle in order to learn about both exact ordering and the successor principle? Several possible mechanisms could explain this order of acquisition.

From a theoretical perspective, until children understand the *meaning* of the number words in isolation from their count lists (what the word “five” means conceptually), it may be very difficult, perhaps impossible, to have a conceptual grounding for the *order* of those values within the count list. In other words, until children understand how to categorize their number words with both a lower and an upper bound (i.e., a set labeled “five” has exactly five objects, not four, and not six), their sense of order for those values will likely be imprecise. Children may not be *able* to learn the principles underlying ordering of number words and cardinal values until they understand the cardinal principle because without that principle, they do not yet have the exact numerical categories upon which to build that understanding.

In addition, after learning the cardinal principle, children are in a position to recognize patterns in their everyday interactions involving counting and set size. One of these patterns is the order of *cardinal values* in the list, not just the order of the words. For example, a CP-knower who is counting beads while placing them onto a stick knows that the cardinal value of the beads on the stick is always the last number she has counted. As she places each additional bead onto the stick, she might notice that the cardinal value of the set changes to exactly one later in the count list. There are other patterns the child could now discover through similar processes, such as “adding objects means going later in the count list” (Le Corre, 2014), and “taking away objects means going earlier in the count list.” Discovering these patterns is a first step not only toward learning the successor principle, but also toward learning about simple arithmetic; learning the cardinal principle opens up the potential to learn about a new world of numerical relationships.

We suggest that learning count list fluency (as measured by *What Comes After*) may be a necessary first step for children to think relationally about numbers, both for understanding exact numerical ordering, and for understanding the successor principle. In addition, we suggest that exact ordering and succession involve learning similar concepts beyond count list fluency. That is, learning to order cardinal values exactly involves learning that there *is* a strict order to these values, and that larger values come after smaller values. Learning the successor principle involves learning that each cardinal value represented by a number word in the count list is exactly one less than the value represented by the number word after it. As mentioned earlier, one can see that these underpinning concepts are highly intertwined—to understand that each cardinal number word represents a value that is one less than its successor, one also needs to recognize that there is a strict order to cardinal values and that numbers representing larger cardinal values fall later in the count list. Our data suggest that exact ordering and succession can be learned independently over time, and in either order (we found CP-knowers who started out as low-knowledge on succession and ordering at pretest who improved only on the succession task, but also CP-knowers who started out as low-knowledge on succession and ordering at pretest who improved only on the ordering tasks). We note that our study was not specifically designed to examine the order of acquisition of exact ordering and succession. Nevertheless, these data set the stage for future longitudinal research on the links between ordering and succession tasks, both its acquisition and children’s strategies in solving them. For example, do children recruit the same knowledge to succeed on both ordering and succession tasks, either initially, or eventually? Or, do children perhaps initially solve these tasks in different ways (e.g., invoking a later-is-greater concept to help them solve the ordering task, but not the succession task), but later use more similar relational thinking for both skills? Future research can tease these possibilities apart.

4.5. Limitations and future research

Our findings suggest the need for a longer term longitudinal study in which children can be followed closely from when they are 3-knowers to after they learn the cardinal principle in order to assess when they begin to be able to solve tasks that require understanding the successor principle and exact ordering. Although our findings suggest that those children who have just transitioned to being CP-knowers are most ready to learn about these exact numerical relations, our study was too short to track children in their transition from 3- or 4-knowers to becoming CP-knowers. Such a study could test whether children’s understanding of the successor principle and exact ordering follow quickly after they became CP-knowers, how quickly, and perhaps whether certain kinds of interactions or interventions lead to this next stage of understanding.

One limitation of our study is the lack of difference between the Experimental and Control training conditions. CP-knowers with low prior knowledge of exact numerical relations improved in both conditions; 3- and 4-knowers improved in neither condition. It is not possible to determine whether those children who improved learned from both the Experimental and Control training conditions, or were not affected by either condition. The training provided by *both* the Experimental and Control conditions may have accelerated learning the successor principle in children who were ready to learn the principle, that is, in CP-knowers. Our Control condition involved repeatedly counting and pointing to real objects, and parental talk about large sets of present objects (4–10), including counting, has been associated with relatively early acquisition of cardinal-number knowledge in preschoolers (Gunderson & Levine, 2011). As noted previously, CP-knowers are in a good position to learn from counting interactions because they can mentally supply the cardinal value of the set. In other words, CP-knowers in the Control condition may have supplied the cardinal values that were provided by the experimenter in the Experimental condition. Thus it is possible that the repeated counting in our Control training helped children improve their successor principle knowledge as much as our more explicit Experimental training.

The fact that the amount of time between pretest and posttest was unrelated to change in knowledge of the successor principle suggests that developmental maturation was not the operative factor; however, we cannot definitively rule out this possibility. Further research comparing training conditions similar to ours with a non-numerical control training condition could help us understand whether both or neither trainings were effective.

In addition, because knowledge of the cardinal principle is associated with other numerical skills (such as counting fluency and the ability to order small sets of 1–3 items), it is possible that these other skills, or a suite of skills, contributes to CP-knowers' ability to learn the successor principle. Although we did not find any other measured skill that significantly predicted learning over and above knowledge of the cardinal principle, it is possible that an unmeasured skill might do so, or that lack of power contributed to these null effects. Future research could attempt to tease apart the potential contributions of other skills by training specific skills among CP-knowers and 3–4-knowers and assessing the contribution of these trained skills to successor principle learning. Finally, we cannot rule out the possibility that some other kind of successor principle training might help 3–4-knowers learn the successor principle or exact ordering prior to knowing the cardinal principle. Future research could attempt other kinds of interventions to test this possibility.

Whatever mechanism allowed CP-knowers to improve their understanding of exact numerical relations in the time between pretest and posttest was not sufficient for 3–4-knowers to do the same. Learning the cardinal principle appears to set the stage for rapid development of knowledge about exact numerical relations. More broadly, children appear to learn how counting relates to quantity in a series of smaller conceptual steps, rather than a single conceptual leap. We suggest that the cardinal principle may play a pivotal role in this process, perhaps by increasing children's ability to learn from routine numerical interactions in their environment. Further research is needed to determine the mechanism by which children harness their knowledge of the cardinal principle in learning the successor principle and exact ordering, and to dive more deeply into the relationship between children's learning of those two numerical relational skills and the factors that influence children's learning of the successor principle.

Supplementary material

De-identified data files and analysis scripts for this paper are available at Open Science Framwork: https://osf.io/xuq9h/.

Acknowledgments

This research was supported by the McCormick Foundation and CME Group Foundation grant, "From the Classroom to the Lab and Back: Improving Children's Early Math Skills" 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 National Center for Education Research (NCER) Grant R305C050076 and National Science Foundation CAREER DRL-1452000 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 Gianesin and Emalie Norton for help in data collection. We also thank Susan Carey and two anonymous reviewers for their extremely helpful feedback during the review process.

Appendix A. Ordering task stimuli and ordering task pretest histograms

See Figs. A.1–A.3.

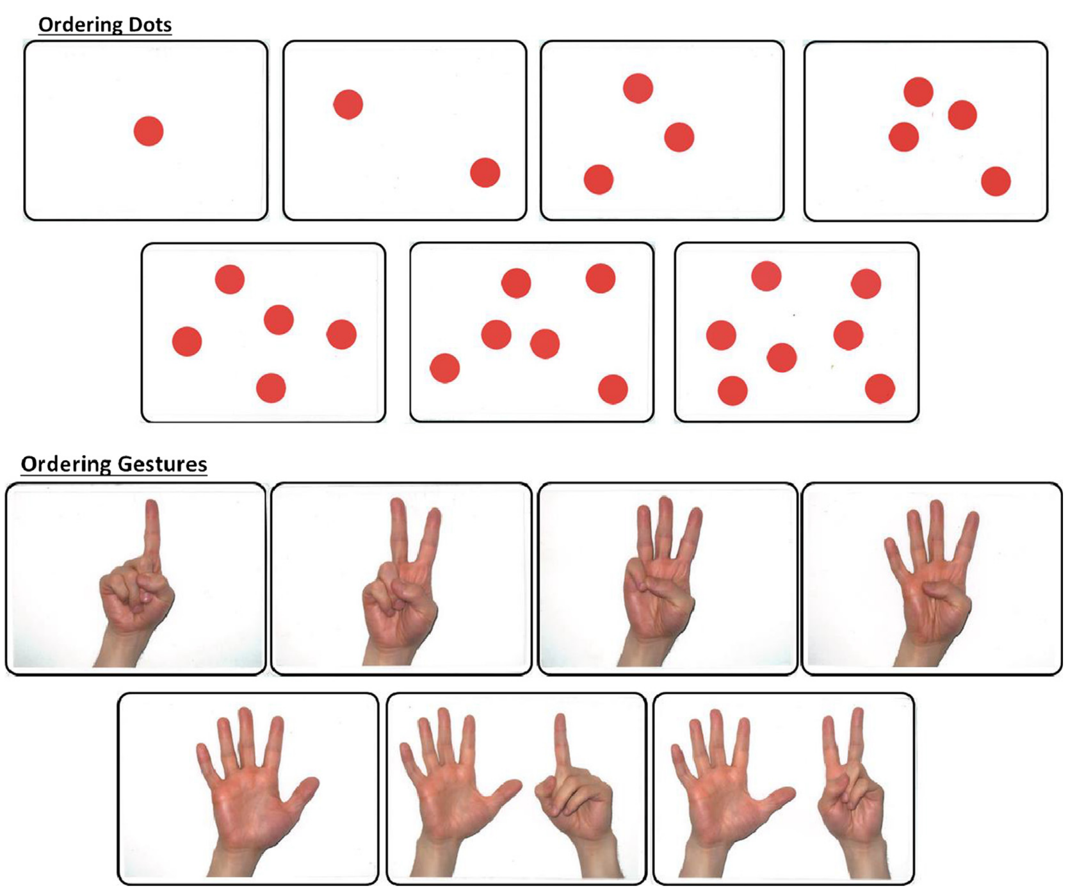

Fig. A.1. Ordering task stimuli.

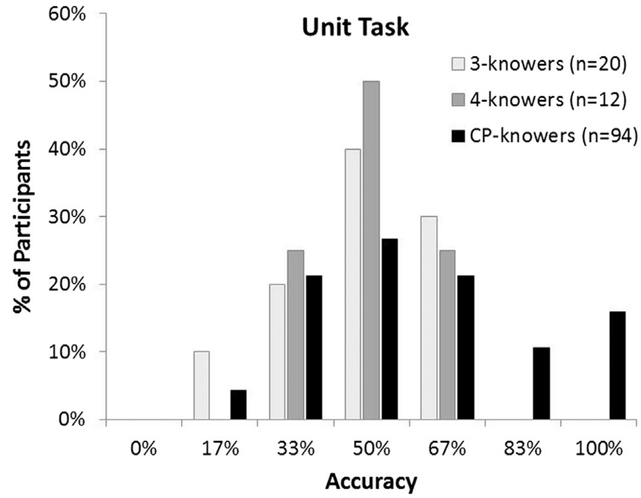

Fig. A.2. Histogram showing pretest performance (accuracy) on Unit task, by knower-level.

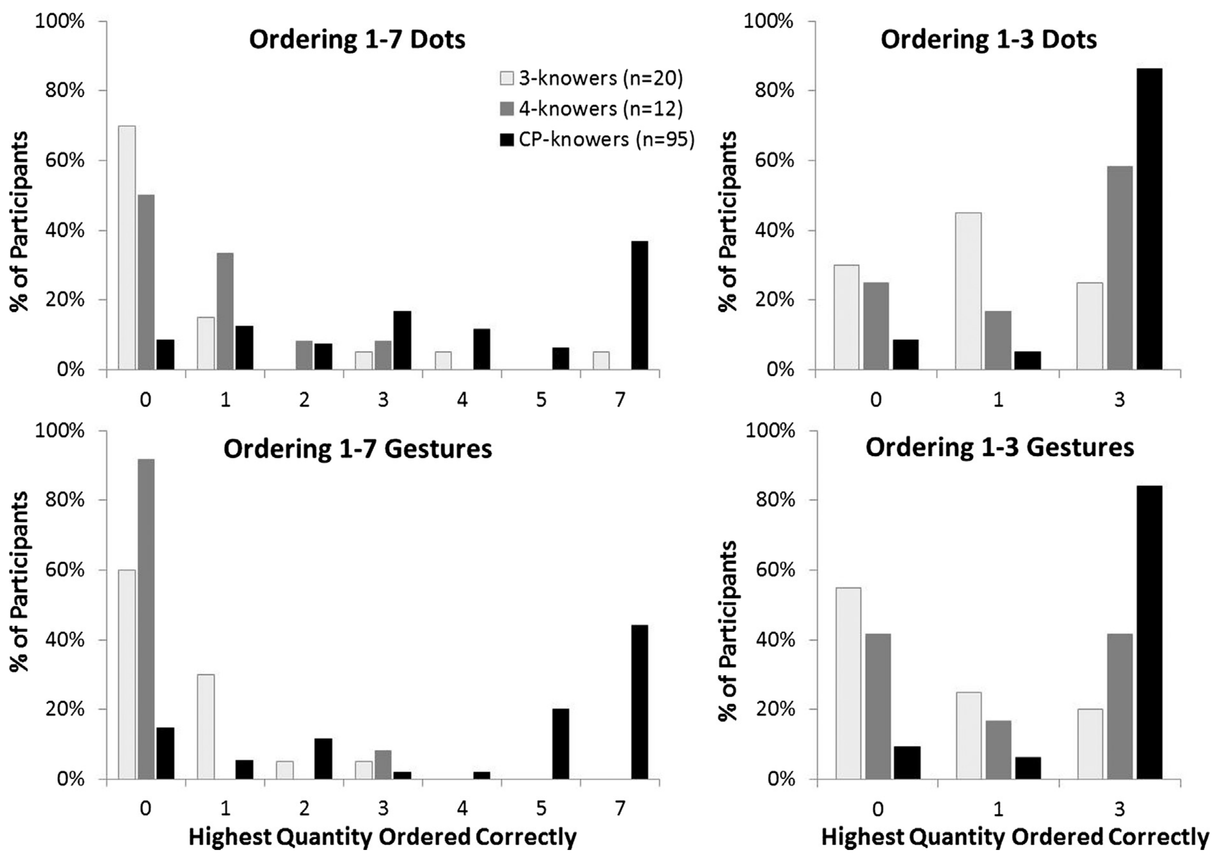

Fig. A.3. Histograms showing pretest performance (highest quantity ordered correctly) on each Ordering task, by knower-level.

Appendix B. Scripts and materials for training conditions

See Fig. B.1.

B.1. Experimental training script

Warm-up: ONE (Milton)

Milton really likes one. See, he has the number one on his hat and he has one dot on his shirt. He likes to have one of everything. But Milton is sad because he doesn't have one wug. See what happens when I give him one wug? Now Milton is happy! *[take off TOY]* Can you give Milton one wug so he'll be happy? *[Give feedback]*

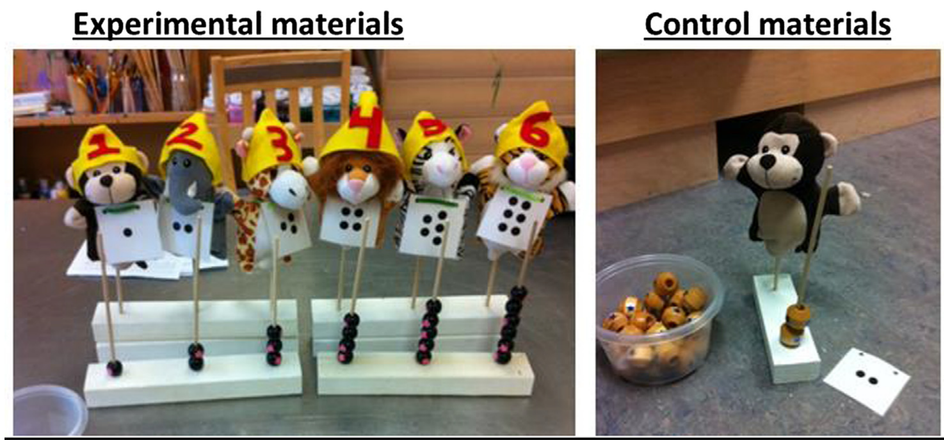

Fig. B.1. Materials used in the experimental and control conditions.

ONE (Milton)

Now let's make Milton happy by giving him one wug. Can you give Milton one wug?
[give feedback on all trials]

Incorrect: Uh oh, Milton isn't happy. Let's fix it. *[Exp fixes]* Yay, now Milton is happy because he has one wug!
Correct: Yay! Milton is happy because he has one wug.

What did we do to make Milton happy? *[After child's response]* He didn't have any wugs and we gave him one *[point to wug]* so he would have one wug. Now he has one wug.

TWO (Ellie)

Now Ellie is here! Ellie really likes two. See, she has the number two on her hat and she has two dots on her shirt < *point to each dot* >. Let's give her ONE wug, just like we gave Milton < *point to Milton's stick* > < *prompt child to count while putting each toy on stick* – exp. *and child count together* >. Now Milton and Ellie have the same, they both have ONE wug. < *point to Milton's and Ellie's sticks* >. Milton's happy, but Ellie's NOT happy! She only has one wug. She likes to have two of everything. She already has ONE wug. How many MORE wugs do we have to give Ellie to make her happy?

Correct: "Right, we have to give her one more so that she'll have two"
Incorrect: "Oops, she already has 1 *[point]*, so we need give her 1 more so that she'll have two. Let's give her one more."
As soon as child adds one bead, allow enough time to see if she goes for more, but stop her before she can put any more beads on.

Let's check and see if Ellie's happy. Yay! Ellie is happy. She has two wugs: < *prompt child to count and point* – exp. *and child count together* > one, two. Remember, Ellie had ONE wug, just like Milton < *point to Milton's stick* >. How many MORE wugs did we have to give Ellie to make her happy? *[After child's response]* She had one wug *[point to bottom wug]* and we gave her one more *[point to top wug]* so she would have two wugs. Now she has two wugs, < *prompt child to count and point* – exp. *and child count together* > one, two. Now they're BOTH happy. Milton has one wug < *point to Milton's stick* > and Ellie has two wugs < *point to Ellie's stick* > *!*

THREE (Ginny)

Now Ginny is here! Ginny really likes three. See, she has the number three on her hat and she has three dots on her shirt < *point to each dot* >. Let's give Ginny TWO wugs, just like we gave Ellie < *point to Ellie's stick* > < *prompt child to count while putting each toy on stick* – exp. *and child count together* >. Now Ellie and Ginny have the same, they both have TWO wugs. < *point to Ellie's and Ginny's sticks* >. Ellie's happy, but Ginny is NOT happy! She only has two wugs. She likes to have three of everything. She already has TWO wugs. How many MORE wugs do we have to give Ginny to make her happy?

Correct: "Right, we have to give her one more so that she'll have three
Incorrect: "Oops, she already has 2 *[point to each]*, so we need give her 1 more so that she'll have three. Let's give her one more."
As soon as child adds one bead, allow enough time to see if she goes for more, but stop her before she can put any more beads on.

Let's check and see if Ginny's happy. Yay! Ginny is happy. She has three wugs < *prompt child to count and point* – exp. *and child count together* > one, two, three. Remember, Ginny had TWO wugs, just like Ellie < *point to Ellie's stick* >. How many MORE wugs did we have to give Ginny to make her happy? *[After child's response]* She had two wugs *[point to each]* and we gave her one more *[point to top wug]* so she would have three wugs. Now she has three wugs, < *prompt child to count and point* – exp. *and child count together* > one, two, three. Now they're ALL happy. Milton has one wug < *point to Milton's stick* > Ellie has two wugs < *point to Ellie's stick* >, and Ginny has three wugs < *point to Ginny's stick* > *!*

FOUR (Leo)

Now Leo is here! Leo really likes four. See, he has the number four on his hat and he has four dots on his shirt. < *point to each dot* >. Let's give Leo THREE wugs, just like we gave Ginny < *point to Ginny's stick* >. < *prompt child to count while putting each toy on stick* – exp. *and child count together* >. Now Ginny and Leo have the same, they both have THREE wugs. < *point to Ginny's and Leo's sticks* >. Ginny's happy, but Leo is NOT happy! He only has three wugs. He likes to have four of everything. He already has THREE wugs. How many MORE wugs do we have to give LEO to make him happy?

Correct: “Right, we have to give him one more so that he’ll have four”
Incorrect: “Oops, he already has 3 *[point to each]*, so we need give him 1 more so that he’ll have four. Let’s give him one more.”
As soon as child adds one bead, allow enough time to see if she goes for more, but stop her before she can put any more beads on.

Let’s check and see if Leo’s happy. Yay! Leo is happy. He has four wugs: < *prompt child to count and point* – exp. *and child count together* > one, two, three, four. Remember, Leo had THREE wugs, just like Ginny < *point to Ginny’s stick* > . How many MORE wugs did we have to give Leo to make him happy? *[After child’s response]* He had three wugs *[point to each]* and we gave him one more *[point to top wug]* so he would have four wugs. Now he has four wugs, < *prompt child to count and point* – exp. *and child count together* > one, two, three, four. Now they’re ALL happy. Milton has one wug < *point to Milton’s stick* > Ellie has two wugs < *point to Ellie’s stick* > *,* Ginny has three wugs < *point to Ginny’s stick* > *,* and Leo has four wugs < *point to Leo’s stick* > *!*

FIVE (Zoe)

Now Zoe is here! Zoe really likes five. See, she has the number five on her hat and she has five dots on her shirt < *point to each dot* > . Let’s give her 4 wugs, just like we gave Leo < point to Leo’s stick > < *prompt child to count while putting each toy on stick* – exp. *and child count together* > . Now Leo and Zoe have the same. They both have FOUR wugs < *point to Leo and Zoe’s sticks* > . Leo is happy, but Zoe’s NOT happy! She only has 4 wugs, she likes to have 5 of everything. She already has FOUR wugs, how many more wugs do we have to give Zoe to make her happy?

Correct: “Right, we have to give her one more so that she’ll have five”
Incorrect: “Oops, she already has 4 *[point to each]*, so we need give her 1 more so that she’ll have five. Let’s give her one more.”
As soon as child adds one bead allow enough time to see if she goes for more, but stop her before she can put any more beads on.

Let’s check and see if Zoe’s happy. Yay! Zoe is happy. She has five wugs < *prompt child to count and point* – exp. *and child count together* > one, two, three, four, five. Remember, Zoe had FOUR wugs, just like Leo < *point to Leo’s stick* > . How many MORE wugs did we have to give to Zoe to make her happy? *[After child’s response]* She had four wugs *[point to each]* and we gave her one more *[point to top wug]* so she would have five wugs. Now she has five wugs, < *prompt child to count and point* – exp. *and child count together* > one, two, three, four, five. Now they’re ALL happy. Milton has one wug < *point to Milton’s stick* > , Ellie has two wugs < *point to Ellie’s stick* > *,* Ginny has three wugs < *point to Ginny’s stick* > Leo has four wugs < *point to Leo’s stick* > , and Zoe has five wugs < *point to Zoe’s stick* > *!*

SIX (Tony)

Now Tony is here! Tony really likes six. See, he has the number six on his hat and he has six dots on his shirt < *point to each dot* > . Let’s give him FIVE wugs, just like we gave Zoe < *point to Zoe’s stick* > < *prompt child to count while putting each toy on stick* - exp. *and child count together* > . Now Zoe and Tony have the same. They both have FIVE wugs < *Point to Zoe and Tony’s sticks* > . Zoe is happy, but Tony’s NOT happy! He only has five wugs. He likes to have six of everything. He already has FIVE wugs. How many MORE wugs do we have to give Tony to make him happy?

Correct: “Right, we have to give him one more so that he’ll have six”
Incorrect: “Oops, he already has 5 *[point to each]*, so we need give him 1 more so that he’ll have six. Let’s give him one more.”
As soon as child adds one bead, allow enough time to see if she goes for more, but stop her before she can put any more beads on.

Let’s check and see if Tony’s happy. Yay! Tony is happy. He has six wugs < *prompt child to count and point* – exp. *and child count together* > one, two, three, four, five, six. Remember, Tony had FIVE wugs just like Zoe < *point to Zoe’s stick* > . How many MORE wugs did we have to give to Tony to make him happy? *[After child’s response]* He had five wugs *[point to each]* and we gave him one more *[point to top wug]* so he would have six wugs. Now he has six wugs, < *prompt child to count and point* – exp. *and child count together* > one, two, three, four, five, six. Now they’re ALL happy. Milton has one wug < *point to Milton’s stick* > , Ellie has two wugs < *point to Ellie’s stick* > , Ginny has three wugs < *point to Ginny’s stick* > , Leo has four wugs < *point to Leo’s stick* > , Zoe has five wugs < *point to Zoe’s stick* > , and Tony has six wugs < *point to Tony’s stick* > !

B.2. Control training script

ONE

This is Milton. Milton is getting ready to start school and he wants to practice his counting! Can you help Milton practice counting so he’ll be ready for school?

< *take 1 bead out of bucket and hold in hand* >
Milton wants to count this. Let’s count it.
< *Place bead in bowl/lid, count along with child* >
1…
< *Provide feedback if child is hesitant – help by starting with “1…”* >
Great! Now let’s help Milton put them on the stick. Can you count it while you put it on the stick?
< *prompt child to count* – exp. *and child count together* >
Uh-oh! Milton forgot. Let’s count them again so he can remember!
< *prompt child to count and point* – exp. *and child count together* >
Great Job!
< *Remove beads – start next trial* >

TWO

< *take 2 beads out of bucket and hold in hand* >
Now Milton wants to count these. Let’s count them.
< *Place beads in bowl/lid, count along with child* >
Great! Now let’s help Milton put them on the stick. Can you count them while you put it on the stick?
< *prompt child to count* – exp. *and child count together* >

Uh-oh! Milton forgot. Let's count them again so he can remember!
< *prompt child to count and point* – exp. *and child count together* >
Great Job!

THREE
< *take 3 beads out of bucket and hold in hand* >
Now Milton wants to count these. Let's count them.
< *Place beads in bowl/lid, count along with child* >
Great! Now let's help Milton put them on the stick. Can you count them while you put it on the stick?
< *prompt child to count* – exp. *and child count together* >
Uh-oh! Milton forgot. Let's count them again so he can remember!
< *prompt child to count and point* – exp. *and child count together* >
Great Job!

FOUR
< *take 4 beads out of bucket and hold in hand* >
Now Milton wants to count these. Let's count them.
< *Place beads in bowl/lid, count along with child* >
Great! Now let's help Milton put them on the stick. Can you count them while you put it on the stick?
< *prompt child to count* – exp. *and child count together* >
Uh-oh! Milton forgot. Let's count them again so he can remember!
< *prompt child to count and point* – exp. *and child count together* >
Great Job!

FIVE
< *take 5 beads out of bucket and hold in hand* >
Now Milton wants to count these. Let's count them.
< *Place beads in bowl/lid, count along with child* >
Great! Now let's help Milton put them on the stick. Can you count them while you put it on the stick?
< *prompt child to count* – exp. *and child count together* >
Uh-oh! Milton forgot. Let's count them again so he can remember!
< *prompt child to count and point* – exp. *and child count together* >
Great Job!

SIX
< *take 6 beads out of bucket and hold in hand* >
Now Milton wants to count these. Let's count them.
< *Place beads in bowl/lid, count along with child* >
Great! Now let's help Milton put them on the stick. Can you count them while you put it on the stick?
< *prompt child to count* – exp. *and child count together* >
Uh-oh! Milton forgot. Let's count them again so he can remember!
< *prompt child to count and point* – exp. *and child count together* >
Great Job!

Appendix C. Relation of training condition to learning

We first examined whether training condition (Experimental vs. Control) was related to change over time in our key DVs (successor principle knowledge and exact ordering), or on any other DVs (What Comes After, additional measures of ordering knowledge, or knower-level). Performance of 3-knowers versus 4-knowers at pretest did not significantly differ on any task (*p*s $< .05$). Therefore, we grouped 3- and 4-knowers together to increase the power of our analyses. Since CP-knowers were diverse in terms of their pretest successor principle and ordering knowledge (see below), in each set of analyses, we matched CP-knowers to 3–4-knowers based on their pretest scores on each task (referred to subsequently as "low-successor-knowledge CP-knowers" and "low-ordering-knowledge CP-knowers"). (For the readers' reference, data for all CP-knowers are provided in Appendix D.)

Unit task. Three- and 4-knowers' ($n = 32$) pretest Unit task scores ranged from 17% to 67% correct ($M = 49\%$, $SD = 15\%$). We classified the 69 (out of 94) CP-knowers whose pretest Unit task scores fell within this range as "low-successor-knowledge CP-knowers" ($M = 48\%$, $SD = 15\%$, range = 17–67%). A 2 (Condition: Control, Experimental) × 2 (Knower-Group: 3–4-knowers, low-successor-knowledge CP-knowers) × 2 (Time: Pretest, Posttest) mixed-effects ANCOVA with age as a covariate revealed no statistically significant interactions of Time × Condition or Time × Knower-Group × Condition (*F*s < 1.0, *p*s > 0.40). However, there was a significant Time × Knower-Group interaction, $F(1, 97) = 4.47$, $p = .037$, $\eta_p^2 = 0.044$, which we discuss further in the main text. Separate ANCOVAs within each Knower-Group confirmed that there was no significant Time × Condition interaction for either 3–4-knowers ($p = 0.685$) or low-successor-knowledge CP-knowers ($p = 0.539$). In other words, Condition was not related to improvement over time on the Unit task.

What Comes After task. We also compared growth of 3–4-knowers to growth of low-successor-knowledge CP-knowers on What Comes After. A 3-way (Condition × Knower-Group × Time) mixed effects ANCOVA with age as a covariate revealed no statistically significant interactions of Time × Condition, Time × Knower-Group, or Time × Knower-Group × Condition (*F*s < 1.0, *p*s > 0.30). Separate ANCOVAs within each Knower-Group showed no significant Time × Condition interaction for either 3–4-knowers ($p = 0.618$) or low-successor-knowledge CP-knowers ($p = 0.305$). Thus, Condition was not related to improvement over time on the What Comes After task.

Ordering 1–7 Dots and Gestures. All but one 3–4-knower (out of 32) got no items correct on Ordering 1–7 Dots and Ordering 1–7 Gestures; we excluded the one 3–4-knower who got one pretest item correct from these analyses. We classified the 46 (out of 95) CP-knowers who got zero Ordering 1–7 items correct at pretest as "low-ordering-knowledge CP-knowers".

Accuracy. Because there was zero variance in pretest scores, we conducted a binomial GEE on posttest Ordering 1–7 accuracy (correctly ordering all sets up to 7), with Task (Dots, Gestures) as a within-subjects factor, Condition and Knower-Group as between-subjects factors, and Age as a covariate. For Ordering 1–7 accuracy, Condition (B = 0.05, SE = 0.47, p = .914) was not a significant predictor of posttest performance; neither was Task (B = 0.70, SE = 0.43, p = .104) nor Age (B = 0.11, SE = 0.38, p = .775). Knower-Group was the only significant predictor of posttest Ordering 1–7 accuracy (B = 3.28, SE = 1.05, p = .002).

We also conducted ordinal GEEs on posttest Ordering 1–7 accuracy within low-ordering-knowledge CP-knowers, with Task as a between-subjects factor, Condition as a within-subjects factor, and Age as a covariate. None of the predictors were significant. (Because only one 3–4-knower scored above zero at posttest, it was not possible to test the impact of Condition on 3–4-knowers.)

Highest correct. We treated highest correct scores (i.e., the highest set size prior to the child's first error) as ordinally distributed, due to their non-normal distribution. To assess any effects of Condition on Ordering 1–7 highest correct scores, we conducted ordinal GEEs on Ordering 1–7 highest correct score, with Time (pretest, posttest) and Task (Dots, Gestures) as within-subjects factors, Condition and Knower-Group as between-subjects factors, and Age as a covariate. A significant Time × Condition interaction would show that Condition affected improvement over time; a Time × Condition × Knower-Group interaction would show that Condition affected improvement over time differentially for each Knower-Group. However, there were no significant effects of Time × Condition or Time × Condition × Knower-Group (ps > .10). Only the main effects of Time (B = 0.90, SE = .36, p = .049) and Knower-Group (B = 1.93, SE = .54, p < .001) were significant. We also examined the effects of Time, Task, Condition, and Age in separate models within each Knower-Group; again, we found no Time × Condition interaction for either Knower-Group (3–4-knowers or low-ordering-knowledge CP-knowers) (ps > .10).

Ordering 1–3 Dots and Gestures. We examined whether Condition related to improvement in performance on the Ordering 1–3 tasks among 3–4-knowers or low-ordering-knowledge CP-knowers.

Accuracy. We conducted binomial GEEs on Ordering 1–3 accuracy (ordering all items correctly up to 3) with Task (Dots, Gestures) and Time (pretest, posttest) as within-subjects factors, Condition and Knower-Group as between-subjects factors, and Age as a covariate. There was no significant Time × Condition or Time × Condition by Knower-Group interaction (ps > .25). Further, binomial GEEs within each Knower-Group with Task, Time, Condition and Age as predictors showed no significant Time × Condition interaction within either knower-group (ps > .25).

Highest correct. We again conducted ordinal Generalized Estimating Equations (GEEs) predicting Ordering 1–3 highest correct scores (highest set ordered correctly before the child's first error), including Time, Task, Condition, and Knower-Group, with Age as a covariate. There were no significant effects of Condition × Time or Condition × Time × Knower-Group (ps > .20). We also examined the effects of Time, Task, Condition, and age in separate models within each Knower-Group; again, we found no Time × Condition interaction for either Knower-Group (ps > .30).

Knower-level. The Experimental training condition might have led to improvement in knower-level among the 3–4-knowers. However, this was not the case. Among 3–4-knowers, a logistic regression with Condition and Age as predictors, and whether the child improved in knower-level as the outcome, found no significant effects of Condition (B = -0.06, SE = 0.78, p = .943) or Age (B = 0.55, SE = 0.69, p = .425). In fact, a paired-samples sign test across Conditions showed no significant change in knower-level from pretest to posttest (exact p = 0.839). In other words, we found no evidence that the 3–4-knowers increased in knower-level over time, overall, or as a function of Condition.

Appendix D. Results for all CP-knowers

We restricted our analyses in the main text to CP-knowers who were comparable to 3–4-knowers in their pretest accuracy on the relevant task (Unit task or Ordering 1–7 tasks). This is important in order to test whether cardinal principle knowledge per se, and not differences in prior knowledge, were driving differences between knower-level groups in change over time. Here, for the purpose of completeness, we report results for all CP-knowers (n = 95).

D.1. Relation of training condition to learning

When including all CP-knowers, the relation of training condition to learning remained the same as in Appendix C: Condition did not relate to change over time on any task.

Unit task. A 2 (Condition: Control, Experimental) × 2 (Knower-Group: 3–4-knowers, CP-knowers) × 2 (Time: Pretest, Posttest) ANCOVA on Unit task accuracy, with age as a covariate, revealed no statistically significant effects of Time × Condition or Time × Knower-Group × Condition (Fs < 1.0, ps > 0.60). An ANCOVA within CP-knowers confirmed that there was no significant Time × Condition interaction (p = 0.809).

What Comes After task. A 2 (Condition: Control, Experimental) × 2 (Knower-Group: 3–4-knowers, CP-knowers) × 2 (Time: Pretest, Posttest) ANCOVA on What Comes After task accuracy, with age as a covariate, revealed no statistically significant effects of Time × Condition or Time × Knower-Group × Condition (Fs < 1.1, ps > 0.30). An ANCOVA within CP-knowers confirmed that there was no significant Time × Condition interaction (p = 0.251).

D.2. Ordering 1–7 dots and gestures

Accuracy. With all participants included, we conducted binomial GEEs on Ordering 1–7 accuracy with predictors Task, Condition, Knower-Group, Age, and pretest accuracy on Ordering 1–7. (Time could not be included as a factor due to lack of variance among 3–4-knowers.) Significant predictors were Knower-Group (B = 2.78, SE = .73, p < .001) and pretest accuracy (B = 2.27, SE = .46, p < .001). Condition was not a significant predictor (B = 0.02, SE = .34, p = .960), nor were Age and Task (ps > .05).

Within only CP-knowers, we conducted a binomial GEE with Time, Task, Condition, and Age as predictors of Ordering 1–7 accuracy. A Time × Condition interaction would indicate that CP-knowers grew more in one condition than the other. There were significant main effects of Time (B = 0.46, SE = .23, p = .049) and Task (B = 0.39, SE = .16, p = .014), but no significant effects of Condition (p = .340) nor the Time × Condition interaction (p = .702).

Highest correct. To assess any effects of Condition on Ordering 1–7 highest correct scores, we conducted an ordinal GEE with the predictors Time, Task, Condition, and Knower-Group, and Age as a covariate. There were no significant effects of Condition × Time or Condition × Time × Knower-Group ($ps > .45$). We also examined the effects of Time, Task, Condition, and Age in a separate model within CP-knowers; again, we found no Time × Condition interaction ($p = .682$).

D.3. Ordering 1–3 dots and gestures

Accuracy. With all participants included, we conducted a binomial GEE predicting Ordering 1–3 accuracy from Task, Tie, Condition, Knower-Group, and Age. Significant predictors were Knower-Group (B = 1.97, SE = .63, $p = .002$), Age (B = 0.86, SE = .38, $p = .023$), and Task (B = -0.45, SE = .19, $p = .017$). Neither the Condition × Time nor the Condition × Time × Knower-Group interaction was significant ($ps > .40$).

Within only CP-knowers, we conducted a binomial GEE predicting Ordering 1–3 from Task, Time, Condition, and Age. The Time × Condition interaction was not significant (B = −0.33, SE = .68, $p = .629$).

Highest correct. An ordinal GEE predicting Ordering 1–3 highest correct scores, including predictors Time, Task, Condition, and Knower-Group, with Age as a covariate, showed no significant effects of Condition × Time or Condition × Time × Knower-Group ($ps > .35$). We also examined the effects of Time, Task, Condition, and Age for CP-knowers; again, we found no Time × Condition interaction ($p = .655$).

D.4. Successor principle

Consistent with our main analyses, when analyzing all CP-knowers, CP-knowers significantly improved from pretest to posttest on the successor principle task, but not on the What Comes After task.

Unit task. Among all CP-knowers, there was significant improvement on the Unit task from pretest ($M = 0.60$, $SD = 0.24$) to posttest ($M = 0.65$, $SD = 0.24$), $t(93) = 2.14$, $p = .035$, $d = 0.22$. However, when all CP-knowers were included and compared to 3–4-knowers in a 2 (Knower-Group) × 2 (Time) ANCOVA, controlling for age, there was a main effect of Knower-Group ($F(1, 123) = 8.65$, $p = .004$, $\eta_p^2 = .066$), but no significant effects of Time or the Time × Knower-Group interaction $Fs < 1.5$, $ps > 0.20$). The lack of a Time × Knower-Group interaction when all CP-knowers are included appears to be due to the fact that including CP-knowers who were already close to ceiling on the Unit task at pretest reduced the average improvement among all CP-knowers (compared to the improvement among low-successor-knowledge CP-knowers).

What Comes After. Among all CP-knowers, there was no significant improvement on the What Comes After task from pretest ($M = 0.77$, $SD = 0.27$) to posttest ($M = 0.79$, $SD = 0.27$), $t(94) = 0.64$, $p = .525$, $d = 0.07$. A 2 (Knower-Group) × 2 (Time) ANCOVA on What Comes After accuracy, controlling for age, revealed a main effect of Knower-Group ($F(1, 124) = 29.53$, $p < .001$, $\eta_p^2 = .192$), and no significant effects of Time or the Time × Knower-Group interaction ($Fs < 1.0$, $ps > 0.60$).

D.5. Exact ordering

Consistent with our main analyses, when analyzing all CP-knowers, CP-knowers improved from pretest to posttest on exact ordering.

Ordering 1–7 Dots and Gestures. See Appendix Table D.1 for pretest and posttest ordering performance for all CP-knowers. In order to assess the effect of Knower-Group on improvement, we conducted a binomial GEE on posttest Ordering 1–7 Accuracy, with Task, Knower-Group, Age, and pretest accuracy on Ordering 1–7 as predictors. Knower-Group (B = 2.78, SE = .74, $p < .001$) and pretest accuracy (B = 2.26, SE = .46, $p < .001$) were significant predictors, whereas Task and Age were not ($ps > .05$). Thus, CP-knowers performed significantly better at posttest than 3–4-knowers, even after controlling for pretest performance and age.

D.6. Additional Ordering Analyses

Ordering 1–7 Dots and Gestures Highest Correct. Performance at pretest and posttest for all CP-knowers ($n = 95$) is displayed in Appendix Table D.1. To compare 3–4-knowers' and all CP-knowers' improvement over time on Ordering 1–7 Highest Correct, we conducted an ordinal GEE, with factors of Time, Task, and Knower-Group, and Age as a covariate. There were significant effects of Time (B = 0.70, SE = .25, $p = .005$), Knower-Group (B = 2.96, SE = .44, $p < .001$), and Age (B = 0.63, SE = .27, $p = .019$), but no significant effect of Task ($p = .159$) and no Time × Knower-Group interaction (B = -0.02, SE = .29, $p = .960$), indicating that the knower-groups did not significantly differ from one another in improvement over time. This contrasts with our main analyses, which showed that low-ordering-knowledge CP-knowers improved significantly more over time than 3–4-knowers. Including all CP-knowers reduced the potential for growth among CP-knowers, since some were at ceiling at pretest.

Ordering 1–3 Dots and Gestures Accuracy. performance at pretest and posttest for all CP-knowers is displayed in Appendix Table D.1. To compare improvement across Knower-Groups, we conducted a binomial GEE on Ordering 1–3 accuracy with Task, Time, Knower-Group, and Age as predictors. Similar to the main text analyses, there were significant main effects of Task, Knower-Group, and Age ($ps < .05$), but no significant Time × Knower-Group interaction ($p = .100$), and no main effect of Time ($p = .841$). Thus, CP-knowers did not significantly differ from 3 to 4-knowers in growth on Ordering 1–3 accuracy.

Ordering 1–3 Dots and Gestures Highest Correct. Performance at pretest and posttest for all CP-knowers ($n = 95$) is displayed in Appendix Table D.1. To compare 3–4-knowers' and CP-knowers' improvement over time on Ordering 1–3 Highest Correct, we conducted an ordinal GEE with factors of Time, Task, and Knower-Group, and Age as a covariate. There was a significant Time × Knower-Group interaction (B = 0.98, SE = .43, $p = .025$), as well as significant main effects of Task (B = -0.44, SE = .17, $p = .010$), Knower-Group (B = 2.01, SE = .38, $p < .001$), and Age (B = 0.73, SE = .35, $p = .035$). Thus, as in the main text, CP-knowers improved significantly more than 3–4-knowers on Ordering 1–3 highest correct.

Table D.1
Pretest and posttest performance for ordering measures, for all CP-knowers.

	CP-knowers (n = 95)		Pretest to posttest change[b]
	M[a] (SD)	Median (range)	
Pretest			
Ordering 1–7			
Dots Accuracy	0.37***	NA	
Gestures Accuracy	0.44***	NA	
Dots Highest Correct (max. 6)	3.77 (2.12)	4.00 (0–6)	
Gestures Highest Correct (max. 6)	4.08 (2.31)	5.00 (0–6)	
Ordering 1–3			
Dots Accuracy	0.86***	NA	
Gestures Accuracy	0.84***	NA	
Dots Highest Correct (max. 2)	1.78 (0.59)	2.00 (0–2)	
Gestures Highest Correct (max. 2)	1.75 (0.62)	2.00 (0–2)	
Posttest			
Ordering 1–7			
Dots Accuracy	0.47***	NA	*n.s.*
Gestures Accuracy	0.59***	NA	*
Dots Highest Correct (max. 6)	4.64 (1.58)	5.00 (0–6)	***
Gestures Highest Correct (max. 6)	4.87 (1.82)	6.00 (0–6)	***
Ordering 1–3			
Dots Accuracy	0.94***	NA	*n.s.*
Gestures Accuracy	0.93***	NA	*
Dots Highest Correct (max. 2)	1.92 (0.35)	2.00 (0–2)	*
Gestures Highest Correct (max. 2)	1.87 (0.47)	2.00 (0–2)	*n.s.*

*$p < .05$, **$p < .01$, ***$p < .001$

[a] Significance of comparison to chance (note: not calculated for Highest Correct). Comparisons to chance for Ordering 1–3 Accuracy were binomial tests versus 16.7%; for Ordering 1–7 Accuracy, binomial tests versus .02%.

[b] Pretest to posttest changes for Accuracy were tested using the McNemar test. Pretest to posttest changes for Highest Correct were tested using the non-parametric Wilcoxon Signed-Rank test.

Appendix E. Relations among tasks at pretest, within CP-knowers

E.1. Results

We first examined zero-order correlations among all pretest measures, within CP-knowers (N = 95) (Table E.1). We also created a composite score for Ordering 1–7 Accuracy by averaging children's accuracy across the two tasks (Ordering 1–7 Dots and Ordering 1–7 Gestures). We created a composite score for Ordering 1–3 Accuracy in the same way. Among CP-knowers, pretest scores on the Unit task, What Comes After task, Ordering 1–7 tasks, and Ordering 1–3 tasks were all positively inter-correlated (all correlations significant at $p < .05$, except the correlation between Unit

Table E.1
Correlations among pretest tasks within CP-knowers (N = 95).

	1.	2.	3.	4.	5.	6.	7.	8.
1. Unit Task accuracy	–	.39***	.33**	.24*	.34***	.25*	.13	.31**
2. What Comes After accuracy	.39***	–	.50***	.44***	.43***	.35***	.29**	.33***
3. Ordering 1–7 Accuracy (Dots & Gestures)	.35***	.50***	–	*NA*	*NA*	.39***	.35***	.34***
4. Ordering 1–7 Dots Accuracy	.26*	.44***	*NA*	–	.54***	.32**	.29**	.27**
5. Ordering 1–7 Gestures Accuracy	.36***	.43***	*NA*	.55***	–	.37***	.33**	.33**
6. Ordering 1–3 Accuracy (Dots & Gestures)	.26*	.35***	.40***	.32**	.38***	–	*NA*	*NA*
7. Ordering 1–3 Dots Accuracy	.16	.29**	.37***	.30**	.35***	*NA*	–	.60***
8. Ordering 1–3 Gestures Accuracy	.30**	.33***	.34***	.27**	.33***	*NA*	.58***	–
9. Age	.18	.04	.16	.10	.18	.11	.20*	.00

Notes: Values below the diagonal are zero-order correlations. Values above the diagonal are partial correlations controlling for age.
Unit and What Comes After treated as continuous. Ordering Accuracy scores are dichotomous.
Pearson's r is reported for all continuous and dichotomous variables (for two dichotomous variables, this simplifies to the phi coefficient; for one continuous and one dichotomous variable, this simplifies to the point-biserial correlation).
*$p < .05$, **$p < .01$, ***$p < .001$

Table E.2
Simultaneous regressions predicting CP-Knowers' Pretest Unit Task Accuracy, Ordering 1–7 Accuracy, and What Comes After Task Accuracy (N = 94).

	Pretest Unit Task Accuracy Linear regression β Model 1	Pretest Ordering 1–7 Accuracy (Dots & Gestures) Binary logistic regressionB (SE) Model 2	Pretest What Comes After Accuracy Linear regression β Model 3
Age	0.14	0.46 (0.35)	−0.06
Pretest Unit	*NA*	1.26~ (0.76)	0.26**
Pretest What Comes After	0.29**	4.07*** (0.92)	*NA*
Pretest Ordering 1–7 Accuracy (Dots & Gestures combined)	0.19~	*NA*	0.41***
R-squared	20.5		29.7
F-stat (df)	7.75*** (3, 90)		12.67*** (3, 90)
Likelihood ratio chi-square (df)		47.1*** (3)	

Notes. NA: Could not be entered in the model due to lack of variance or collinearity.
~$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$.

task and Ordering 1–3 Dots accuracy). Further, partial correlations controlling for age (Table E.1) were quite similar in magnitude and in statistical significance to the zero-order correlations, indicating that Age did not confound these relations. Indeed, Age was not significantly correlated with any measure except Ordering 1–3 Dots Accuracy (Table E.1).

As predicted, Unit and Ordering 1–7 Accuracy were significantly related ($r(91) = .33$, $p = .001$, controlling for age). Further, there were significant relations between Unit and What Comes After tasks ($r(91) = 0.39$, $p < .001$, controlling for age), and Ordering 1–7 Accuracy and What Comes After tasks ($r(92) = 0.50$, $p < .001$, controlling for age). Thus, at pretest, CP-knowers who understood the successor principle were more likely to understand exact ordering. Further, procedural fluency with the count list was related to both succession and exact ordering.

Although we originally included the What Comes After task as a control measure, the strong relations between this task and each of our key DVs among CP-knowers (Unit task and Ordering 1–7 Accuracy) prompted us to explore these three tasks further. To do so, we conducted three regression analyses, with pretest performance on each task as a DV, and the other two pretest tasks, and age, as predictors (Table E.2). CP-knowers' Unit task accuracy was significantly predicted by What Comes After task accuracy and marginally predicted by Ordering 1–7 Accuracy; age was not a significant predictor (Table E.2, Model 1). Similarly, CP-knowers' Ordering 1–7 Accuracy was significantly predicted by What Comes After task accuracy and marginally predicted by Unit task accuracy; again age was not a significant predictor (Table E.2, Model 2). Finally, What Comes After task accuracy was significantly predicted by both Unit task accuracy and Ordering 1–7 Accuracy; again age was not significant (Table E.2, Model 3).

E.2. Discussion

Perhaps unsurprisingly, CP-knowers' performance on many of the numerical relations tasks at pretest was highly correlated. Even controlling for age, children's performance on the *Unit* task, the *Exact Ordering 1–7* tasks, and *What Comes After* were all significantly correlated. Perhaps most interestingly, regression analyses indicated that *What Comes After* performance was a unique predictor of both succession *and* exact ordering performance. We had hypothesized a relationship between *What Comes After* and the *Unit* task because of the structural similarities between the tasks. The fact that *What Comes After* also predicts children's ordering suggests that fluency and flexible thinking about the count list is an important competency for understanding exact numerical relations more generally, at least among CP-knowers.

In addition, the regression analyses revealed a marginal relationship between the *Unit* task and the *Exact Ordering* tasks, over and above *What Comes After*'s contribution to both tasks. This finding suggests, as noted in the introduction, that succession and exact ordering likely invoke similar, although not identical, reasoning about exact numerical relations. In fact, the two tasks may complement each other in interesting ways. They require some of the same relational thinking, but different skills may come into play for each task. For example, in the Ordering task, children can always start from "one" to create their orders, which may provide scaffolding to create a correct order for some children. In the Unit task, children must show understanding of pairs of larger numbers and number words in isolation. The Ordering tasks also requires keeping track of many more numerical relationships at once, so might require higher levels of executive function than the Unit task, where children only need to track two numbers at a time. However, in the Ordering tasks, children were able to check their work and fix mistakes, whereas the Unit task required children to respond without visual aids. All of these factors could be investigated in future research.

Appendix F. Arabic numeral identification and ordering

In addition to Ordering Dots and Gestures, participants were also asked to order Arabic Numerals. Participants completed an Arabic numeral identification task in which they were asked to name each Arabic numeral (from 1 to 10), presented on flash cards in a random order. Three- and four-knowers were quite poor at this – only 20% of 3-knowers and 17% of 4-knowers successfully named all 7 Arabic numerals used in the Ordering tasks (1–7). In addition, only 65% of CP-knowers successfully named the Arabic numerals from 1 to 7. Given children's low success rates at Arabic numeral identification, and the differences in these rates across knower-levels, performance on the Arabic numeral ordering tasks is difficult to interpret. However, for the interested reader, we present children's pretest accuracy by knower-level in Table F.1, and pretest and posttest accuracy by knower-group in Table F.2.

Table F.1
Pretest arabic numeral task performance by knower-level.

	Three-knowers (n = 20) M[a] (SD)	Four-knowers (n = 12) M[a] (SD)	CP-knowers (n = 95) M[a] (SD)	3-knowers vs. 4-knowers[b]	3-knowers vs. CP-knowers[b]	4-knowers vs. CP-knowers[b]
Arabic numeral identification (% correct on numerals from 1 to 7)	0.47 (.35)	0.62 (.29)	0.90 (.20)	*n.s.*	***	**
Ordering 1–7						
Arabic Numerals Accuracy	0.05**	0.17***	0.72***	*n.s.*	***	***
Arabic Numerals Highest Correct (max. 6)	1.05 (1.57)	1.25 (2.26)	5.07 (1.82)	*n.s.*	***	***
Ordering 1–3						
Arabic Numerals Accuracy	0.25	0.17	0.91***	*n.s.*	***	***
Arabic Numerals Highest Correct (max. 2)	0.80 (0.83)	0.50 (0.80)	1.85 (0.48)	*n.s.*	***	***

Notes. For tasks with a binary outcome (i.e., Ordering Accuracy), the percent of participants who were correct is reported.
*p < .05, **p < .01, ***p < .001
[a] Asterisks indicate significance of comparison to chance (note: not calculated for Ordering Highest Correct scores or Arabic numeral identification). Comparisons to chance for Ordering 1–7 Accuracy were binomial tests versus 0.02%. Comparisons to chance for Ordering 1–3 Accuracy were binomial tests versus 16.7%.
[b] Arabic numeral identification compared using independent-samples t-tests. Ordering Accuracy compared using Pearson's χ^2. Ordering Highest Correct compared non-parametrically using Mann-Whitney U test.

Table F.2
Pretest and Posttest Performance for Arabic Numeral Ordering Tasks, by Knower-Group.

	M[a] (SD)		Group difference[b] *p-value*	Pretest to posttest change[c] *p-value*	
	Three- and four-knowers (n = 31)	CP-knowers (n = 46)		3–4-knowers	CP-knowers
Pretest					
Ordering 1–7					
Arabic Numerals Accuracy	0.10***	0.48***	***		
Arabic Numerals Highest Correct (max. 6)	1.03 (1.78)	4.15 (2.28)	***		
Ordering 1–3					
Arabic Numerals Accuracy	0.23	0.83***	***		
Arabic Numerals Highest Correct (max.2)	0.68 (0.83)	1.72 (0.66)	***		
Posttest					
Ordering 1–7					
Arabic Numerals Accuracy	0.03**	0.65***	***	*n.s.*	*
Arabic Numerals Highest Correct (max. 6)	0.87 (1.41)	4.67 (2.17)	***	*n.s.*	**
Ordering 1–3					
Arabic Numerals Accuracy	0.42***	0.87***	***	*n.s.*	*n.s.*
Arabic Numerals Highest Correct (max.2)	0.97 (0.95)	1.80 (0.54)	***	*n.s.*	*n.s.*

Notes. Analyses includes 3–4-knowers (n = 31) and low-ordering-knowledge CP-knowers (n = 46, matched to 3–4-knowers based on pretest Ordering 1–7 Dots and Gestures accuracy).
*p < .05, **p < .01, ***p < .001
[a] Significance of comparison to chance (note: not calculated for Highest Correct). Comparisons to chance for Ordering 1–7 Accuracy were binomial tests versus 0.02%. Comparisons to chance for Ordering 1–3 Accuracy were binomial tests versus 16.7%.
[b] Group differences for Accuracy were tested using Pearson's χ^2. Group differences for Highest Correct were tested using the non-parametric Mann-Whitney U test.

Appendix G. Ordering task response slopes

Ordering 1–7 Slopes. We also asked whether participants might show some knowledge of "later-is-greater" on the Ordering 1–7 tasks, beyond that which could be detected using our measure of Highest Correct. To test this, we examined the slope relating participants' actual response to the correct response, for those numbers beyond their Highest Correct number. For example, if a child's order was "1, 3, 5, 4, 6, 2, 7", their Highest Correct would be 1, and we would analyze the slope relating the remaining numbers, "3, 5, 4, 6, 2, 7", to "2, 3, 4, 5, 6, 7". We reasoned that participants who made a mistake early in the sequence would have the most room to show this type of knowledge, and therefore calculated these slopes for participants with Highest Correct scores of 0 or 1 on each task. The slopes were not significantly greater than zero for any knower-level on any task, at pretest or posttest (see Appendix Table G.1). To check the robustness of these results, we also calculated these slopes including all participants with Highest Correct scores of 4 or lower. These slopes did not significantly differ from zero for any knower-level, on any Ordering 1–7 task, at pretest or posttest (ps > .10). Therefore, we did not find any evidence that children showed use of "later-is-greater" ordering strategies beyond the point where they made their highest correct number (i.e, beyond their first mistake). Therefore, we use Highest Correct to measure ordering skill (in addition to Accuracy).

Table G.1
Ordering 1–7 Tasks: Slope of Responses Above Highest Correct, for Participants who Ordered Only 0 or 1 Correctly.

	Three-knowers			Four-knowers			CP-knowers		
	n	M (SD)	*p*	*n*	M (SD)	*p*	*n*	M (SD)	*p*
Pretest Ordering 1–7									
Dots	17	-.04 (.47)	.736	10	-.19 (.31)	.091	20	-.12 (.42)	.231
Gestures	18	-.01 (.41)	.948	11	-.12 (.39)	.325	19	-.18 (.37)	.042
Posttest Ordering 1–7									
Dots	14	.03 (.29)	.697	9	.04 (.45)	.814	6	-.21 (.28)	.125
Gestures	17	-.10 (.37)	.259	10	-.14 (.35)	.235	9	.27 (.39)	.073

Note. p-values are from one-sample t-tests versus zero.

References

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition, 83*(3), 223–240.

Brannon, E. M., & Van de Walle, G. A. (2001). The development of ordinal numerical competence in young children. *Cognitive Psychology, 43*(1), 53–81.

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

Cheung, P., Rubenson, M., & Barner, D. (2017). To infinity and beyond: Children generalize the successor function to all possible numbers years after learning to count. *Cognitive Psychology, 92*, 22–36. http://dx.doi.org/10.1016/j.cogpsych.2016.11.002.

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

Duncan, G. J., Claessens, A., Huston, A. C., Pagani, L. S., Engel, M., Sexton, H., ... Duckworth, K. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446. http://dx.doi.org/10.1037/0012-1649.43.6.1428.

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.

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., Gibson, D., Goldin-Meadow, S., & Levine, S. C. (2015). Gesture as a window onto children's number knowledge. *Cognition, 144*, 14–28. http://dx.doi.org/10.1016/j.cognition.2015.07.008.

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.

Le Corre, M. (2014). Children acquire the later-greater principle after the cardinal principle. *British Journal of Developmental Psychology, 32*(2), 163–177. http://dx.doi.org/10.1111/bjdp.12029.

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.

Lyons, I. M., & Beilock, S. L. (2009). Beyond quantity: Individual differences in working memory and the ordinal understanding of numerical symbols. *Cognition, 113*(2), 189–204.

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*(2), 256–261. http://dx.doi.org/10.1016/j.cognition.2011.07.009.

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. [Article]. *Developmental Science, 17*(5), 714–726. http://dx.doi.org/10.1111/desc.12152.

Mix, K. S. (2008). Children's equivalence judgments: Crossmapping effects. *Cognitive Development, 23*(1), 191–203.

Mix, K. S., Sandhofer, C. M., Moore, J. A., & Russell, C. (2012). Acquisition of the cardinal word principle: The role of input. *Early Childhood Research Quarterly, 27*(2), 274–283. http://dx.doi.org/10.1016/j.ecresq.2011.10.003.

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.

Sarnecka, B. W., & Wright, C. E. (2013). The idea of an exact number: Children's understanding of cardinality and equinumerosity. *Cognitive Science, 37*(8), 1493–1506. http://dx.doi.org/10.1111/cogs.12043.

Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2017). Preschool children use space, rather than counting, to infer the numerical magnitude of digits: Evidence for a spatial mapping principle. *Cognition, 158*, 56–67. http://dx.doi.org/10.1016/j.cognition.2016.10.010.

Spaepen, E. (2008). Language for number: The limits of numerical reasoning without linguistic input. [Doctoral dissertation], University of Chicago. Retrieved from ProQuest Dissertations Publishing (AAT 3322647).

Spaepen, E., Gunderson, E., Gibson, D., Goldin-Meadow, S., & Levine, S. C. (2018). Meaning before order: cardinal principle knowledge predicts improvement in understanding the successor principle and exact ordering. Retrieved from osf.io/xuq9h.

Wagner, K., Kimura, K., Cheung, P., & Barner, D. (2015). Why is number word learning hard? Evidence from bilingual learners. *Cognitive Psychology, 83*, 1–21. http://dx.doi.org/10.1016/j.cogpsych.2015.08.006.

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352–360. http://dx.doi.org/10.3102/0013189x14553660.

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*, 155–193. 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.

Xu, C., & LeFevre, J.-A. (2016). Training young children on sequential relations among numbers and spatial decomposition: Differential transfer to number line and mental transformation tasks. *Developmental Psychology, 52*(6), 854–866. http://dx.doi.org/10.1037/dev0000124.