# Counting and the ontogenetic origins of exact equality

Rose M. Schneider [*], Erik Brockbank, Roman Feiman, David Barner

[a] Department of Psychology, University of California, San Diego, CA, United States of America
[b] Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, United States of America

A B S T R A C T

Humans are unique in their capacity to both represent number exactly and to express these representations symbolically. This correlation has prompted debate regarding whether symbolic number systems are necessary to represent large exact number. Previous work addressing this question in innumerate adults and semi-numerate children has been limited by conflicting results and differing methodologies, and has not yielded a clear answer. We address this debate by adapting methods used with innumerate populations (a "set-matching" task) for 3- to 5-year-old US children at varying stages of symbolic number acquisition. In five studies we find that children's ability to match sets exactly is related not simply to knowing the meanings of a few number words, but also to understanding how counting is used to generate sets (i.e., the cardinal principle). However, while children were more likely to match sets after acquiring the cardinal principle, they nevertheless demonstrated failures, compatible with the hypothesis that the ability to reason about exact equality emerges sometime later. These findings provide important data on the origin of exact number concepts, and point to knowledge of a counting system, rather than number language in general, as a key ingredient in the ability to reason about large exact number.

## 1. Introduction

Humans, like pigeons, rats, fish, and other non-human animals, are sensitive to numerical properties of our environment, and can represent and discriminate large sets on the basis of their approximate magnitudes (Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004). Unlike other animals, however, humans are also capable of creating and transmitting external symbolic representations of numerical information via verbal numerals, body counting (Bender & Beller, 2012; Saxe, 1981; Wassmann & Dasen, 1994) written notation of numerals (Chrisomalis, 2021), physical artifacts such as clay tokens (Mattessich, 1987; Schmandt-Besserat, 1992), and calculating devices like the abacus (Frank & Barner, 2012; Hatano, Miyake, & Binks, 1977; Stigler, 1984). Also unlike other animals, humans routinely use these symbolic systems to perform exact numerical computations that appear to be unavailable to non-human animals. This relation between symbolic representations and the ability to reason about exact number has emerged as the center of an active debate about the role of language in abstract numerical thought.

One hypothesis suggested by this tight correlation between symbols and numerical thought is that the ability to reason about large exact number depends on learning the meanings of corresponding words, like

"seven" and "fifty-two". On this hypothesis, humans who lack a system of number words might also lack the ability to represent both large exact magnitudes and their properties: In the absence of a symbolic number system, humans might not only have no concept of exactly "seven," but also no understanding that magnitudes can be exactly equal to one another. However, an alternative hypothesis is that the exact concepts encoded by number words depend on some form of pre-existing understanding of exact number. One possibility, for example, is that non-verbal number systems are capable of representing large number precisely (Clarke & Beck, 2021; Gallistel & Gelman, 2000; Gelman & Butterworth, 2005), and that these concepts become accessible to humans when they learn to count (e.g., Gallistel & Gelman, 1992; Leslie, Gelman, & Gallistel, 2008). Another possibility is that humans lack representations of large exact magnitudes, but nevertheless understand that sets can match or differ with respect to the number of items they contain. For example, it is possible to verify that two sets are equal without knowing exactly how many items each set contains, so long as every item in each set has a corresponding item in the other - what Boolos (1986) called "Hume's principle." This is important because it suggests that humans could, in principle, represent and reason about exact equality without verbal labels. Given this, although counting provides

an especially powerful tool for representing number, other, non-symbolic, procedures that draw on one-to-one correspondence might also be used to establish equality, including tallies (e.g., Owens & Lean, 2018), tit-for-tat procedures (Sheldon, 1999; Silk, 2003), or simply pairing off individuals from two sets until none remain. The existence of such practices raises the possibility that humans who do not have a symbolic system for expressing exact numbers might nevertheless recognize that number *can* be represented exactly.

### 1.1. Can innumerate people use one-to-one correspondence to establish equinumerosity?

Perhaps one of the most influential attempts to determine whether a concept of exact number is dependent on symbolic number systems is Gordon's (2004) study of the Pirahã, an indigenous group who live near the Maici River in the Brazilian Amazon, and who lack linguistic labels for any exact numbers, even "one" and "two."[1] Gordon presented his study as a test of the Sapir-Whorf hypothesis (Kay & Kempton, 1984; Whorf, 1956), and argued that if large exact number representations depend on language, then an innumerate group like the Pirahã should lack such concepts, and fail tests that depend upon them. To test this idea, Gordon used a series of "set-matching" tasks in which participants could use one-to-one correspondence to match sets generated by an experimenter (i.e., to establish exact equality). These encompassed not only relatively taxing tasks that required tracking a set of nuts hidden in a can (i.e., the "Nuts-in-a-can" task, Fig. 1C), or aligning a row orthogonally to the experimenter's (i.e., the Orthogonal Task, Fig. 1B), but also much less demanding tasks, including a "parallel" set-matching task, in which the experimenter's set was fully visible and the participant was asked to create a line in parallel to this model (see Parallel Match Task, Fig. 1A). Gordon found that, instead of using a one-to-one procedure to create exact matches for large sets, participants approximated for all but the smallest sets (2–3 items), compatible with the use of the Approximate Number System, which furnishes imprecise magnitude representations of large numerosities (Dehaene, 1997; Feigenson et al., 2004).[2] On the basis of his findings, Gordon argued that the Pirahã lack a concept of exact numerical equality, and therefore that such knowledge depends on language and culturally constructed systems of symbolic number, compatible with the Sapir-Whorf hypothesis.

Subsequent investigations of this question have adopted Gordon's logic but report varying results. In one study, Frank, Everett, Fedorenko, and Gibson (2008) provided another group of Pirahã participants with feedback and training, and found that they produced exact matches when sets could be visually aligned (as in the parallel task), but approximated for sets beyond the parallel individuation range when they could not (as in the Nuts-in-a-can task). Frank and colleagues took this as evidence that (a) Gordon's original participants may not have sufficiently understood the task, and (b) the Pirahã understand that number can be represented exactly, despite lacking exact number words. Challenging this conclusion, however, Everett and Madora (2012) argued that the Pirahã individuals who had participated in the Frank et al. study had been exposed to number language and one-to-one training by missionaries operating in the area. This later work reported that Pirahã participants who had not received this exposure failed to use one-to-one correspondence on the Parallel set-match task, replicating Gordon's initial findings.

Work done with other semi-numerate groups—who have either a limited count list, or limited knowledge of their culture's counting

system—also leaves this question unresolved. For example, using tasks adapted from Gordon (2004), Flaherty and Senghas (2011) concluded that speakers of Nicaraguan Sign Language who were unable to count were also unable to exactly match quantities greater than 3 when sets could not be aligned in parallel. However, like Frank et al. (2008), they found that participants performed well at a Parallel set-matching task that was used as a warm-up to other measures, raising questions about the extent of their failures. In another study, Pica, Lemer, Izard, and Dehaene (2004) found that the Mundurukú, an indigenous Amazonian group which has no words for numbers beyond "five," failed to compute arithmetic operations exactly, and differentiated sets greater than five on the basis of their approximate magnitudes. At the same time, how-ever, Pica et al. also reported that some Mundurukú participants were capable of placing objects into one-to-one correspondence with their fingers and toes (suggesting the presence of a body count system in some individuals). Finally, Butterworth and Reeve (2008) compared numerate English-speaking Australian children to innumerate children who spoke Walpiri or Anindilyakwa (indigenous languages spoken in the Northern Territory of Australia), and found no differences between groups in their ability to match sets, but also that all participants approximated in the tasks used, leaving open the question of how numeracy impacts reasoning about exact number.

In addition to the inconsistencies found in these cross-cultural case studies, another limitation of much previous work in this literature is that it has been restricted to asking whether a symbolic number system is necessary for reasoning about exact number, without differentiating the various mechanisms by which a number system might have such an effect. Both cross-culturally and in development, numeracy is not a bi-nary outcome defined by possession of a single symbolic capacity, but is better characterized as a combination of many different components that emerge at different points across the lifespan (e.g., knowledge of a culturally shared count list, the meanings of individual number words in this list, counting procedures and principles, arithmetic abilities, etc.; Schaeffer, Eggleston, & Scott, 1974; Stephens, Landeros, Perkins, & Tang, 2016; Wynn, 1990, 1992). Consequently, while it is possible that exposure to number words is the single component needed to reason about equality, it is also possible that number words are only one ingredient among many that support reasoning about exact number. In addition, numeracy can vary among individuals in idiosyncratic ways that reflect differences in access to education as well as context-specific economic needs and cultural practices (Coben et al., 2003; Garegae & Lekoko, 2006; Kifleyesus, 2009). Prior work with innumerate groups does not address the many ways in which experiences with symbolic number might impact the ability to reason about exact number, and thus does not provide a clear test of why innumerate groups differ from numerate groups, and what aspects of numeracy might explain differ-ences between groups, if any. Finally, past work on this question, by adopting Gordon's logic, has assumed that the set-matching task is a transparent test of one-to-one correspondence, and, by extension, un-derstanding of exact equality. However, here, too, it is possible that innumerate participants have an understanding of exact equality, but fail to understand how to express this in a set-matching task, as previous studies have argued (Frank et al., 2008; Laurence & Margolis, 2007). To address these questions, we investigated a different semi-numerate group that varies with respect to numeracy within a single cultural context: children who are at different points in the process of learning the meanings of number words.

### 1.2. Numeracy as a continuum: children as a semi-numerate case study

In the US, and in many other numerate cultures, various distinct components of numeracy begin to emerge in the first 5–6 years of life in the form of number word learning, providing a natural experiment for investigating how numeracy impacts reasoning about exact equality. Most children in industrialized cultures have access to a culturally shared count list early in life, and can even recite portions of the count

---

[1] While the Pirahã were initially thought to have exact meanings for "one" and "two," these meanings were later shown to be approximate (Frank et al., 2008).

[2] Here, we elect to use the term "numerosity" to refer to the numerical properties of sets in the world, in friendly defiance of those who would have us call all things numerical "number" (see Clarke & Beck, 2021).
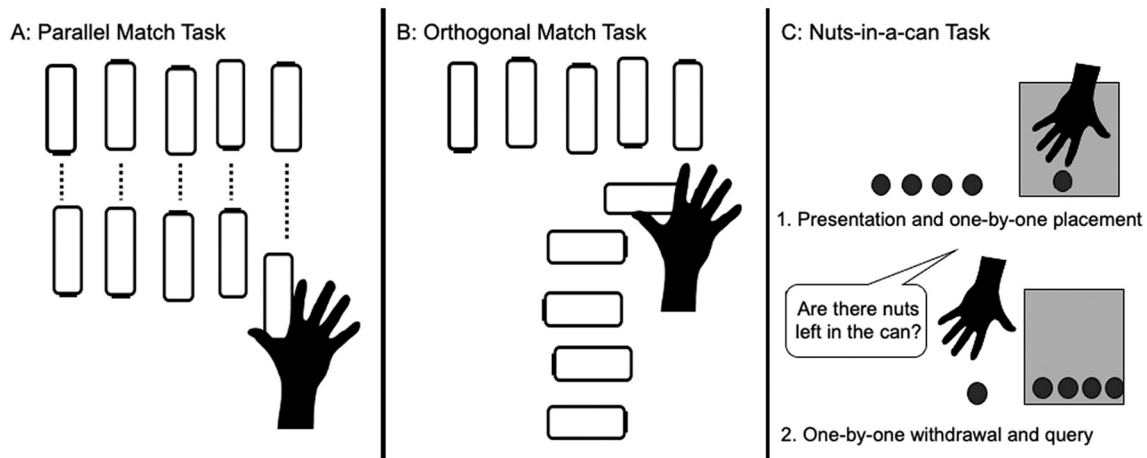
**Fig. 1.** Example of task set-up for Parallel, Orthogonal, and Nuts-in-a-can set-matching tasks from Gordon (2004). Dashed lines in (A) indicate that rows can be spatially aligned using one-to-one correspondence.

list by around 2 years of age, although the words in this count list likely lack numerical content at this stage, and appear to form part of an "unbroken" chain similar to the *ABC*s (for review, see Fuson, 1988). At around 2½ to 3 years of age children in many cultures begin to learn the meanings of these words, beginning with the word "one", followed by "two" 6–9 months later, then eventually "three" some months after that (Wynn, 1990, 1992). During these months, children are called "subset knowers," because they have acquired meanings for only a subset of the number words in their count list and cannot yet accurately label or count sets larger than 3–4. Also during this period, children begin to learn to recite number words while pointing at objects in one-to-one correspondence, although initially these counting routines lack numerical content; children who accurately count often cannot say how many items are in the final set, or cannot use counting to construct sets themselves (Briars & Siegler, 1984; Frye, Braisby, Lowe, Maroudas, & Nicholls, 1989; Fuson, 1988; Fuson, Lyons, Pergament, Hall, & Kwon, 1988; Fuson, Secada, & Hall, 1983; Sarnecka & Carey, 2008; Schaeffer et al., 1974; Siegler, 1991; Wynn, 1990). Sometime between 3 and 5 years of age, children become "counting principle"- or "cardinal principle"-knowers (CP-knowers), meaning that they can accurately label and construct large sets via counting. However, even at this stage children appear to lack a complete understanding of how counting represents number, and take several years to learn that as you count up in the list, you should add one object for each word counted (Davidson, Eng, & Barner, 2012; Cheung, Rubenson, & Barner, 2017; see also Schneider, Sullivan, Guo, & Barner, 2021; Schneider et al., 2020; see Barner, 2017; Carey & Barner, 2019, for review).

These distinct developmental changes offer the opportunity to investigate how different early components of numeracy relate to the ability to reason about exact equality. First, compatible with Frank et al. (2008), some argue that knowledge of number words is not necessary for reasoning about exact equality. For example, on some accounts, knowledge of equality might stem from innate counting principles that entail an understanding of exactness (Butterworth, 1999; Gelman & Butterworth, 2005; Gelman & Gallistel, 1978; Leslie et al., 2008). Other studies argue that preverbal infants can compare sets "via one-to-one correspondence to establish numerical equivalence" (Feigenson & Carey, 2003; p. 568;) (Feigenson et al., 2004; Xu, Spelke, & Goddard, 2005). However, there is currently no evidence that pre-verbal humans can reason about exact equality for large sets. First, although infants can represent and compare sets that contain up to 3 items, they either fail completely for larger sets (e.g., Feigenson & Carey, 2005), or approximate (Cantrell, Boyer, Cordes, & Smith, 2015; Cantrell & Smith, 2013). Second, while infants may use one-to-one correspondence to compare small sets, it is unclear whether they use this information to infer that

the sets are numerically equal. For example, Izard, Streri, and Spelke (2014) argue that when older children (who know the meanings of number words, but not the cardinal principle) use one-to-one correspondence to track hidden objects, they track only the identity of hidden things, rather than their number.

While some studies propose that the seeds of exact equality are present in preverbal infants, others have argued that children only begin to reason about large exact number once they have begun to learn the meanings of their first number words, compatible with Gordon (2004) and Everett and Madora (2012). Some studies posit that this occurs when children become subset knowers, and have not yet learned to count. For instance, Sarnecka and Gelman (2004) argue that although subset knowers only know the meanings of a few small number words, they may nevertheless have a concept of large exact number. As evidence for this, they report that subset knowers judge that a set labeled with an unknown number word, like "seven" must receive a different label if an object is added to it, compatible with an understanding of exact equality. However, others have argued that these effects reflect the pragmatics of mutual exclusivity, are not specific to number words (Brooks, Audet, & Barner, 2013), and cannot be found in other, highly similar tasks that do not involve mutual exclusivity (Condry & Spelke, 2008; Sarnecka & Gelman, 2004).

Based on evidence from tasks that avoided these pragmatic concerns, Sarnecka and Wright (2013) argued that children's intuitions about large exact number emerge at the moment that they become CP-knowers. In their study, they showed subset knowers and CP-knowers two sets of 5 or 6 items aligned in one-to-one correspondence and found that only CP-knowers judged that the two sets should receive the same label when they were equal, but not when they were unequal. From these findings, Sarnecka and Wright proposed that becoming a CP-knower may involve acquiring an understanding of exact equality, since counting establishes a one-to-one correspondence between sets which receive the same cardinal label, highlighting that two sets of, e.g., "five" must be equal. However, much like the earlier work that argued for knowledge of exact equality in subset knowers (Sarnecka & Gelman, 2004), the tasks used by Sarnecka and Wright relied on children's assessments of number *words*, leaving open not only whether subset knowers might successfully reason about large exact sets in non-verbal tasks, but also whether CP-knowers' abilities are limited to reasoning about number words, or if they could also deploy one-to-one correspondence to match sets.

Other studies have used non-verbal tests of equality and have found some evidence that CP-knowers and subset knowers differ, though these experiments have focused mainly on small sets. For example, Mix (1999) found that when preschoolers were shown a card with one set of 2–4

items and then asked to select which one of two additional cards had the same number of items, children with little or no understanding of counting were only slightly better than chance, and performed even worse for items of different kinds (e.g., matching different objects to dots). Although Mix did not assess cardinal principle knowledge, children who could correctly generate sets for large and small numerosities were both more likely to match sets and more likely to ignore perceptual dissimilarities. Similarly, Negen and Sarnecka (2009) showed that while subset knowers struggled to match sets of 1 to 4 items generated by an experimenter, CP-knowers were more likely to do so for all set sizes. Thus, although these studies do not assess reasoning about large sets, they do suggest that learning to count may play some role in children's understanding of exact equality. Other studies temper this conclusion, however. First, studies of semi-numerate Tsimane children (Jara-Ettinger, Piantadosi, Spelke, Levy, & Gibson, 2017) have argued that although the CP- transition is correlated with understanding of exact equality, some subset knowers succeed at a test of set matching and some CP-knowers fail (a topic we revisit in the General Discussion). Second, there is evidence that CP-knowers are better than subset knowers at approximate matching, which could also explain improved set-matching for both small and large sets. For example, in a study by Shusterman and colleagues (Shusterman et al., 2017), children were asked to examine a large set on one table and then find a corresponding number of items on another table based on their memory of the first set. They found that CP-knowers outperformed subset knowers despite not being allowed to count, suggesting that CP-knowers were better than subset knowers at matching sets approximately.

Finally, other studies argue that the capacity to reason about exact equality for large numerosities emerges very late in development, several years after children learn the cardinal principle. Most famously, work beginning with Inhelder and Piaget (1953) found that children as old as 6 years of age sometimes judge that lines of objects placed in one-to-one correspondence are not the same if one of the two sets is lengthened or shortened—a failure of what they called "conservation of number." However, multiple subsequent studies argued that the transformations involved in conservation underestimate children's numerical knowledge due to pragmatic (McGarrigle & Donaldson, 1974; Mehler & Bever, 1967; Rose & Blank, 1974), linguistic (Markman, 1979), and attentional demands (Gelman, 1969). Despite this, some later work by Piaget (1965) found that when 5- and 6-year-old children were asked to match sets in a task analogous to Gordon's Parallel match task, they often matched according to length rather than number (see Russac, 1978, for a similar result in children as old as 7).

Currently, conflicting findings and differing methodologies in the developmental literature make it difficult to determine when the ability to reason about exact equality emerges in children, what form such knowledge might take, and whether it is impacted by changes in numeracy. Developmental tests of exact equality vary with respect to whether they test small or large numbers, whether they offer perceptual cues to scaffold one-to-one correspondence, whether they require children to explicitly reason about equality, and whether children need to reason about equality between two numeral labels or between sets of objects. In addition, there is no work which evaluates the contribution children's general exposure to number language, or their mastery of early-learned rote procedures, like pointing at objects and reciting number words in one-to-one correspondence—an ability which might scaffold reasoning about one-to-one correspondence between sets. Further, despite the variety of paradigms used in the developmental literature, almost no studies have used methods directly comparable to those used with the Pirahã, making it unclear how past studies in children relate to work in innumerate groups, and therefore whether semi-numerate children are a valid model for testing whether numeracy facilitates reasoning about exact number.

In the current work we tested how early changes in numeracy relate to children's ability to exactly match large sets using the set-matching task previously used with the Pirahã (Gordon, 2004). We reasoned

that if the ability to generate exact matches emerges independent of number word knowledge, as Frank et al. (2008) suggest, then children should do well on the set-matching task independently of whether they are CP- or subset knowers. However, if the ability to generate exact matches depends on learning a system of large exact number words, as Gordon (2004) and Everett and Madora (2012) suggest, then subset knowers should fail to exactly match large sets since they lack meanings for large numbers, and CP-knowers should succeed. Finally, if simply learning a counting system is not enough, then even CP-knowers may not generate exact matches as consistently as numerate adults. In addition to testing how set matching abilities relate to numeracy, we also explored the validity of the set-matching task. As noted above, some past studies have argued that innumerate individuals who fail the set-matching task may be capable of reasoning about exact equality, but may fail to comprehend instructions, lack motivation to perform well, or simply lack an effective procedure for matching sets (Frank et al., 2008; Laurence & Margolis, 2007). While these possibilities are difficult to assess in remote innumerate groups for a host of reasons (language, culture, remoteness, and small groups), testing children offers the possibility of high-powered follow-up studies that investigate the role of alternative instructions, motivational factors, and different strategies that the task makes available (e.g., see Frank et al., 2008). This, in addition to exploring effects due to variable numeracy, was a primary motivation for the developmental approach we take here.

In Study 1, we tested subset- and CP-knowers' ability to use one-to-one correspondence to exactly match sets. However, unlike work done with innumerate groups, we broke down children's numeracy into several components. First, we measured how high children could count, as children's count list knowledge can be seen as a general proxy for their previous exposure to counting procedures (LeFevre et al., 2010), and is predictive of other numerical knowledge (Cheung et al., 2017; Davidson et al., 2012; Schneider et al., 2021; Schneider et al., 2020). Second, we tested children's ability to respect one-to-one correspondence when reciting the count routine and pointing at objects in a set. Previous work finds that this ability begins to emerge before children become CP-knowers (Frye et al., 1989; Sarnecka & Carey, 2008), raising the possibility that set-matching via one-to-one correspondence could depend on counting abilities that emerge prior to the CP transition. Given this, we reasoned that this ability, once learned in the context of counting, might impact children's ability to execute analogous one-to-one matching of objects, potentially allowing them to set-match prior to becoming CP-knowers. Third, we measured children's status as either subset knowers (who know a small set of number word meanings) or CP-knowers (who can accurately count and give large sets). Together, these measures allowed us to test which components of counting knowledge, if any, might contribute to the ability to make exact matches. In Studies 2, 3, and 4 we tested three possible alternative explanations for children's performance on these tasks, including attention to number (Study 2), motivation to create equal sets (Study 3), and the ability to surreptitiously count while set-matching (Study 4). Finally, in Study 5, we constructed a set of model simulations to identify whether children's set-matching behavior reflects the deployment of procedures which respect one-to-one correspondence.

## 2. Study 1

### 2.1. Method

All methods and analyses were pre-registered at https://osf.io/3wta2 unless stated otherwise in-text.

#### 2.1.1. Participants

We recruited 160 English-speaking children between 3;0 and 5;0 from local preschools, daycare centers, and museums in San Diego, CA. Sixteen children were tested but excluded from analysis for the following reasons: Failure to finish the experiment ($n = 9$); experimenter

error or broken stimuli ($n = 6$); and failure to pass the training trials for both the parallel and orthogonal orientations ($n = 1$). After these exclusions, our analyzable sample included 144 participants ($M_{age} = 3.94$ years, $SD_{age} = 0.53$ years; $n$ female $= 62$, $n$ male $= 82$). Demographic information by knower-level is shown in Table 1. While we did not collect other demographic information, for this and all other studies our sample was drawn from a population with the following statistics: White (75.5%), Black (5.5%), Asian (12.6%), American Indian or Alaska Native (1.3%), Pacific Islander (0.6%), Multiracial (4.5%) (US Census Bureau).

### 2.1.2. Stimuli, design, and procedure

Children were tested individually in a quiet spot within the classroom or museum, or in a room set apart from the classroom. Participants received the tasks in a fixed order (Set-matching, How Many, Highest Count, and Give-N).

#### 2.1.2.1. Set-matching.
We adapted the set-matching task used by Gordon (2004), framing it as a "matching game": Children were told that the goal of the game was to make their "pond" (4.5"x30" blue rectangle) look like the experimenter's pond (Fig. 2). At the beginning of the task, the experimenter presented children with an empty pond and 15 identical plastic fish, saying, "This is a matching game. Do you know what matching means? Matching means to make things look the same! So, in this game, you're going to be making your pond *look like* my pond." The experimenter then placed a board with one fish glued to its center directly above the child's pond, and said, "Using your fish, can you make your pond look like my pond?" If the child responded correctly the experimenter said, "That's right! Your pond looks like my pond because there is a fish here *(indicating fish on experimenter's board)* and a fish here *(indicating fish on child's board)*. Great job! You made the ponds match." If the child did not respond correctly, the experimenter said, "Hmm, I don't think these ponds match," and pointed out the discrepancies. Children then received an identical training trial for a set of two fish. If a child failed the training trials, the experimenter repeated them once more. Only children who demonstrated an understanding of the training trials by correctly generating sets for these trials with or without feedback were included in analyses.

After the two training trials, children were tested on sets of 3, 4, 6, 8, and 10 with neutral feedback. All children completed trials with the boards first in a parallel orientation (experimenter's board presented directly above the child's, with about $1/2''$ of separation between them), and then in the orthogonal orientation (experimenter's board rotated 90 degrees and to the right of the child's, Fig. 2B). Trial order was fixed for the parallel (3, 4, 10, 8, and 6) and orthogonal (4, 3, 8, 10, and 6) orientations. Finally, because we wished to test how children's ability to use one-to-one correspondence in this task was affected by the perceptual similarity of items across sets, we manipulated the identity of items between subjects. For all trials, half of all CP-knowers and subset knowers received fish which matched the experimenter's (Identical condition), and half were tested with fish that did not match (Non-identical condition).

Children were prevented from counting or using number language, and the experimenter did not use numerical language in their instructions or feedback for two reasons: First, we wished to replicate Gordon (2004)—where participants did not have access to symbolic number—as closely as possible. Second, we wished to ensure that differences between CP- and subset knowers did not simply reflect the availability of a symbolic tool (counting) that only CP-knowers could
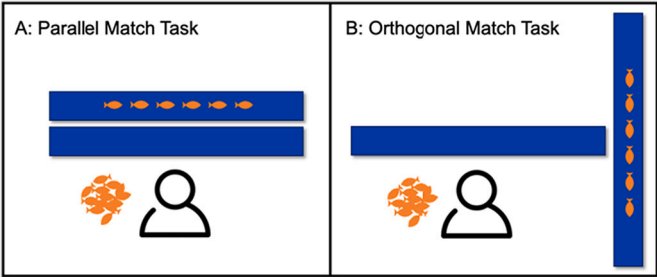


**Fig. 2.** Setup for parallel (A) and orthogonal (B) set-matching task.

use. If the experimenter observed a child counting they covered both boards and said, "This isn't a counting game! This is just a matching game!" and restarted the trial, removing fish from the child's board if necessary. Overt attempts to count were relatively rare: 15 CP-knowers made noticeable attempts to count on 64/700 trials, while 6 subset knowers tried to count on 28/740 trials. Only 10 children attempted 3 or more counts in either the parallel or orthogonal tasks. Counting attempts did not appear to have an impact on performance, as children produced incorrect responses on 47/92 trials where they attempted to count.

#### 2.1.2.2. How many?.
We used a modified version of the "How Many?" task (Cantlon, Fink, Safford, & Brannon, 2007) to test children's procedural counting mastery. This task was chosen because it measures children's ability to execute a one-to-one counting procedure, in which each number word is assigned to a distinct object, analogous to the one-to-one procedure that children must use to match sets of objects in one-to-one correspondence. As noted in the Introduction, the ability to use a one-to-one counting procedure is known to emerge prior to the acquisition of the CP, and therefore offers a measure of numeracy that is distinct from knower level. Also, the task captures intermediate steps in the mastery of this ability, by differentiating between children who correctly count sets on a first effort vs. those who make errors and correct them vs. children who are completely unable to execute a one-to-one counting procedure.

After the final trial of the set-matching task, the experimenter placed a board containing either 8 or 10 fish in front of the child and said, "Can you count these fish out loud and tell me how many there are?" After the child finished counting the first board, the experimenter repeated this process with the other board, with order presentation randomized across participants. Following Cantlon et al., children received a score between 0 and 3 for each instance of counting. A score of 3 indicated errorless counting (coordination of one-to-one tagging and labeling for all objects and correct determination of cardinality). A score of 2 indicated a correct count with errors fixed (either by restarting the count or backtracking). A score of 1 indicated correct counting for a minimum of two items; and a score of 0 indicated random counting.

#### 2.1.2.3. Highest count.
We included Highest Count as a measure of children's verbal counting mastery, as well as a proxy for variability in additional number knowledge after the CP-knower stage. The experimenter introduced this task by saying, "In this game, I want you to count as high as you can. Can you start counting?" If the child did not begin counting, the experimenter prompted them by saying, "Let's start together! *One*…" with rising intonation. Children were allowed to continue counting if they made errors. If a child stopped counting, the experimenter prompted them once by saying, "Do you know what comes after *N*?" Children were only prompted to continue counting once. Children's highest count was the highest number to which they could count, allowing for one error.

#### 2.1.2.4. Give-N.
We assessed children's knowledge of the cardinal principle with a titrated version of the Give-N task (Wynn, 1992). The

**Table 1**
Demographic information by knower-level for Study 1.

|  | $n$ | $M_{age}$ ($SD$) |
|---|---|---|
| Subset knowers | Total $n = 74$ ($n$ female $= 35$; $n$ male $= 39$) | 3.69 (0.41) |
| CP-knowers | Total $n = 70$ ($n$ female $= 27$; $n$ male $= 43$) | 4.20 (0.52) |

experimenter placed ten small plastic items (e.g., bananas) and a plate in front of the child, and said, "In this game, your job is to put these bananas on the plate! Can you put $N$ bananas on the plate for me?" After the child generated the set, the experimenter asked, "Is that $N$? Can you count and make sure?" If the child realized they had made a mistake, they were permitted to fix it. If a child correctly generated a set in response to a given $N$, the next number queried was $N + 1$ (up to six). If they did not, the next number queried was $N$-1. This staircasing procedure was used to determine trial order until a child's knower level could be established. Three was the first number queried for all children.

Children who could correctly generate sets of six at least two out of three times were classified as CP-knowers. If children correctly constructed a set of six once, they were asked to do so a second time. If they succeeded, they were classified as a CP-knower, and the task ended. If they failed, the experimenter asked for a set of five, and the session continued until their knower level could be identified.

If children could not meet the criteria for being classified as CP-knowers, they were identified as $N$-knowers if they correctly provided $N$ on at least two out of the three trials that $N$ was requested and, of those times that the child provided $N$, two-thirds of the time it was in response to a request for $N$. Because we could not ensure an even distribution of $N$-knower levels, in our confirmatory analyses we treat all non-CP-knowers as subset knowers.

## 2.2. Results

Descriptive statistics of set-matching accuracy are shown in Table 2. Closely mirroring the performance of innumerate adults reported in some previous studies (Everett & Madora, 2012; Gordon, 2004), subset knowers were able to match small sets with some degree of accuracy but failed to exactly match large sets in both the parallel and orthogonal conditions. Also, suggesting that the set-matching task is sensitive to changes in numeracy, CP-knowers outperformed subset knowers, with higher accuracy on both parallel and orthogonal trials. While both CP- and subset knowers were more likely to generate exact matches for small sets across both orientations, CP-knowers had greater accuracy even in the parallel condition.

In our first set of analyses we test whether set-matching accuracy is significantly related to acquisition of the cardinal principle by constructing a generalized linear mixed effects model (GLMM) with the formula Accuracy (0/1) ~ Cardinal principle knowledge (CP/Subset) + Set size (continuous) + Orientation (parallel/orthogonal) + Age (continuous) + (1 | Participant).[3] A likelihood ratio test indicated a

**Table 2**
Accuracy on the set-matching task by knower-level and orientation.

| Knower-level | Overall *M (SD)* | Parallel *M (SD)* | Orthogonal *M (SD)* |
|---|---|---|---|
| Subset knower | 0.28 (0.45) | Overall: 0.32 (0.47) | Overall: 0.25 (0.43) |
| | | Small sets: 0.60 (0.49) | Small sets: 0.41 (0.49) |
| | | Large sets: 0.13 (0.33) | Large sets: 0.14 (0.34) |
| CP-knower | 0.51 (0.50) | Overall: 0.61 (0.49) | Overall: 0.40 (0.49) |
| | | Small sets: 0.89 (0.32) | Small sets: 0.74 (0.44) |
| | | Large sets: 0.42 (0.32) | Large sets: 0.18 (0.38) |

significant effect of cardinal principle knowledge ($\chi^2_{(1)} = 25.22$, $p < .001$; Model 1, Appendix),[4] with CP-knowers significantly more likely than subset knowers to generate exact matches in this task overall, even when controlling for age ($\chi^2_{(1)} = 8.00$, $p = .005$). In addition to this main effect of cardinal principle knowledge, the model also revealed significant decreases in accuracy with increasing set sizes ($\chi^2_{(1)} = 324.94$, $p < .001$), as well as greater accuracy for trials in the parallel orientation ($\chi^2_{(1)} = 41.25$, $p < .001$). These main effects are shown in Fig. 3. While CP-knowers outperformed subset knowers overall, they had much lower accuracy for large numerosities in the orthogonal orientation ($M = 0.18$) in comparison to large numerosities in the parallel orientation ($M = 0.42$). A follow-up analysis confirmed this three-way interaction between orientation, cardinal principle knowledge, and set size ($\chi^2_{(1)} = 9.93$, $p = .002$; Model 3, Appendix).

We also explored the magnitude of children's errors on incorrect trials as a less conservative measure of their attention to exact equality. This was done in an effort to identify children who may have attempted to use one-to-one correspondence, but made occasional errors. For these analyses we used children's absolute error (|Target quantity - Response|) on incorrect trials, with the logic that children who attempt a one-to-one match but make an error should generate a response relatively close to the target quantity, regardless of set size, resulting in lower levels of absolute error. To test this hypothesis, we constructed a linear mixed effects model (LMM) with the formula Absolute error ~ Cardinal principle knowledge (CP/Subset) + Set size (continuous) + Orientation (parallel/orthogonal) + Age (continuous) + (1 | Participant). A likelihood ratio test indicated that the a significant effect of CP-knower status($\chi^2_{(1)} = 9.76$, $p = .002$; Model 10, Appendix),[5] with lower absolute error for CP-knowers ($M = 2.42$) in comparison to subset knowers ($M = 3.27$), as shown in Fig. 4. Once again, we found that this effect of cardinal principle knowledge was significant when controlling for the effects of age ($\chi^2_{(1)} = 2.71$, $p = .10$), and also that absolute error significantly increased with set size ($\chi^2_{(1)} = 25.94$, $p < .001$) and for trials in the orthogonal orientation ($\chi^2_{(1)} = 17.17$, $p < .001$).

In both these analyses, CP-knowers were more likely to generate exact matches not only for large numerosities, but also for small ones, resulting in higher accuracy and lower error. In contrast, and analogous to the Pirahã participants in Gordon (2004) and Everett and Madora (2012), subset knowers did not create exact set matches for large numerosities, and even struggled to do so for sets of 4. However, despite their higher accuracy, CP-knowers performed well below numerate adult levels (Frank, Fedorenko, Lai, Saxe, & Gibson, 2012), suggesting that acquiring the cardinal principle is insufficient to succeed on this task, regardless of the reasons underlying children's failures. In addition, these findings raise the question of why some CP-knowers fail while others succeed. One possibility is that some children shift to a one-to-one strategy around the time they become CP-knowers, while others do not. The other is that CP-knowers, like subset knowers, use approximation to solve the task, but are more capable approximators, perhaps due to changes in their approximate number representations, or mappings between number words and the Approximate Number System (Shusterman, Slusser, Halberda, & Odic, 2016).

We conducted a preliminary exploration of these potential mechanisms by investigating differences in response patterns between the parallel and orthogonal trials for CP- and subset knowers (Fig. 5A and B). While parallel trials arguably highlight one-to-one correspondence by making it easy to visually align individuals, items within sets are not

---

[3] All GLMMs and LMMs were fit in R using the 'lme4' package (Bates et al., 2015). All GLMMs used a logit link function.

[4] Although we pre-registered a model containing a CP-knower status*Set size interaction, a likelihood ratio test indicated that this interaction did not improve the fit of the main effects model ($\chi^2_{(1)} = 1.2$, $p = .27$; Model 2, Appendix).

[5] Although we pre-registered a model containing a CP-knower status*Set size interaction, we found that this term did not significantly improve the fit of the model ($\chi^2_{(1)} = 2.74$, $p = .10$; Model 11, Appendix)
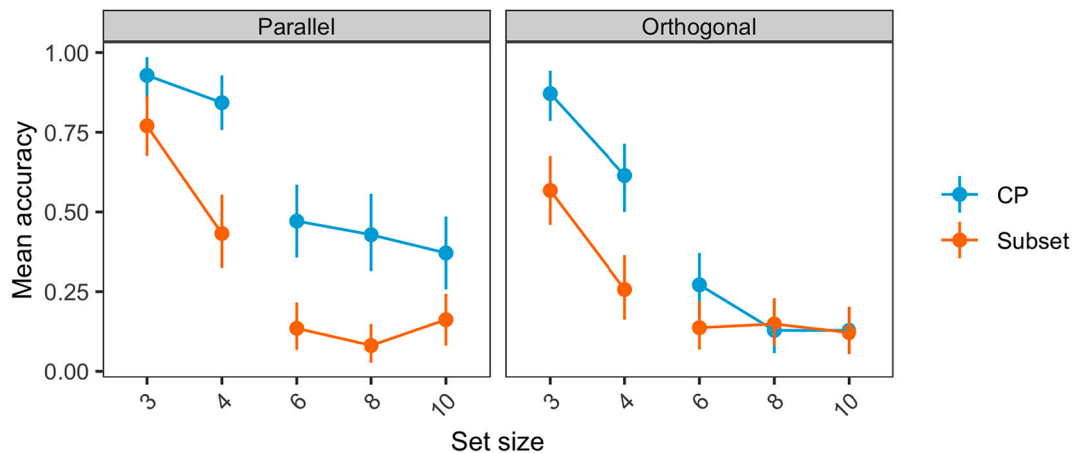
**Fig. 3.** Mean set-matching accuracy for parallel and orthogonal trials, grouped by CP-knower status. Error bars represent 95% confidence intervals computed by non-parametric bootstrap.
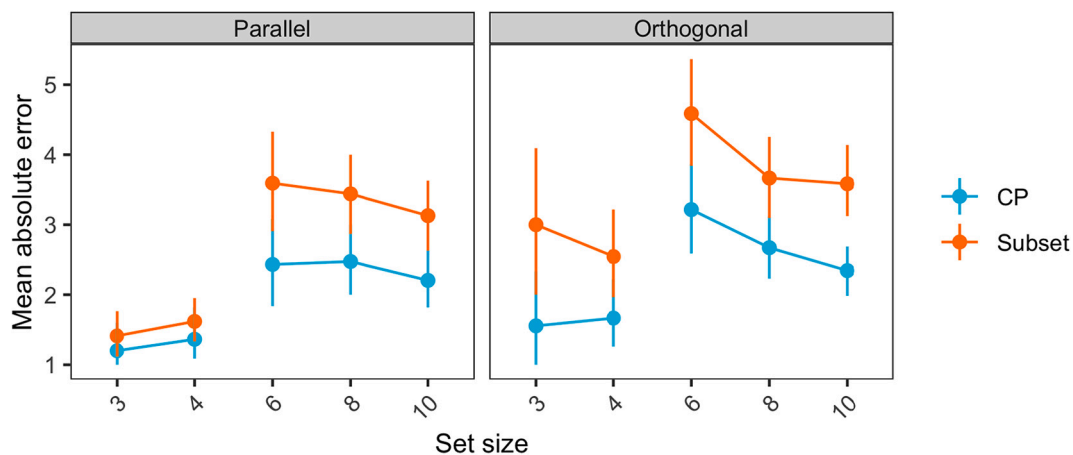


**Fig. 4.** Mean absolute set-matching error by CP-knower status. Error bars represent 95% confidence intervals computed by nonparametric bootstrap.

spatially aligned on orthogonal trials, making it more likely that individuals will generate sets through approximation. Given this, we reasoned that if the difference between CP- and subset knowers was driven by a general improvement in approximation, then this should be similar across the parallel and orthogonal trials. However, if it resulted from an increased use of a one-to-one strategy, then the difference between groups may be greatest in the parallel condition. Compatible with this reasoning, in follow-up analyses using Levene's Test we found that although subset knowers' response variance for sets within the parallel individuation range was significantly different between the parallel and orthogonal trials (both $ps < 0.02$), there was no difference for large numerosities (all $ps > 0.10$). Also, subset knowers frequently appeared to use a "give-all" heuristic for these larger sets, and gave the maximum number of fish (15). This suggests that subset knowers likely used a similar strategy (e.g., approximation) for large numerosities regardless of whether individuals could be spatially aligned. In contrast, CP-knowers had significantly different distributions of responses between the parallel and orthogonal tasks for all set-sizes except three (all $ps < 0.05$). In addition, the overall shape of CP-knowers' responses was strikingly different in the parallel and orthogonal trials, with their responses heavily centered on the target set size in parallel trials, and a wider distribution of responses in the orthogonal trials. These findings provide some preliminary evidence that CP-knowers' set-matching reflects a shift towards a one-to-one strategy, a possibility we address more thoroughly in Study 5.

In our final analysis of this section, we asked whether cardinal

principle knowledge impacted children's ability to ignore perceptual differences between members of sets (e.g., when the two sets contained different kinds of fish) in a follow-up GLMM with the formula Accuracy (0/1) ~ Identity (Identical/Non-identical)*cardinal principle knowledge (CP/Subset) + Set size (continuous) + Orientation (parallel/orthogonal) + Age (continuous) + (1 | Participant) (Model 4, Appendix). While we did not find a significant main effect of identity ($\chi^2_{(1)} = 0.32, p = .57$), this model indicated a significant interaction between identity and cardinal principle knowledge ($\chi^2_{(1)} = 9.81, p = .002$). Post hoc pairwise comparisons revealed that, consistent with past work showing that CP-knowers are more capable of inhibiting differences in identity to establish numerical equivalence (Mix, 1999; Negen & Sarnecka, 2009), CP-knowers had significantly greater accuracy ($p < .001$) on Non-identical trials ($M = 0.54$) in comparison to subset knowers ($M = 0.23$). While there was only a marginal difference ($p = .05$) between groups on Identical trials, CP-knowers' performance did not significantly differ between Identical ($M = 0.47$) and Non-identical trials ($p = .07$). In contrast, subset knowers had significantly higher performance on Identical trials ($M = 0.33$) in comparison to Non-identical trials ($p = .008$), suggesting that they were more able to create numerical set matches when item identity did not differ across sets.

### 2.2.1. Counting

In our next analyses, we tested whether children's ability to exactly match sets was impacted by their procedural or verbal counting mastery. First, we used children's performance on the How Many counting task to
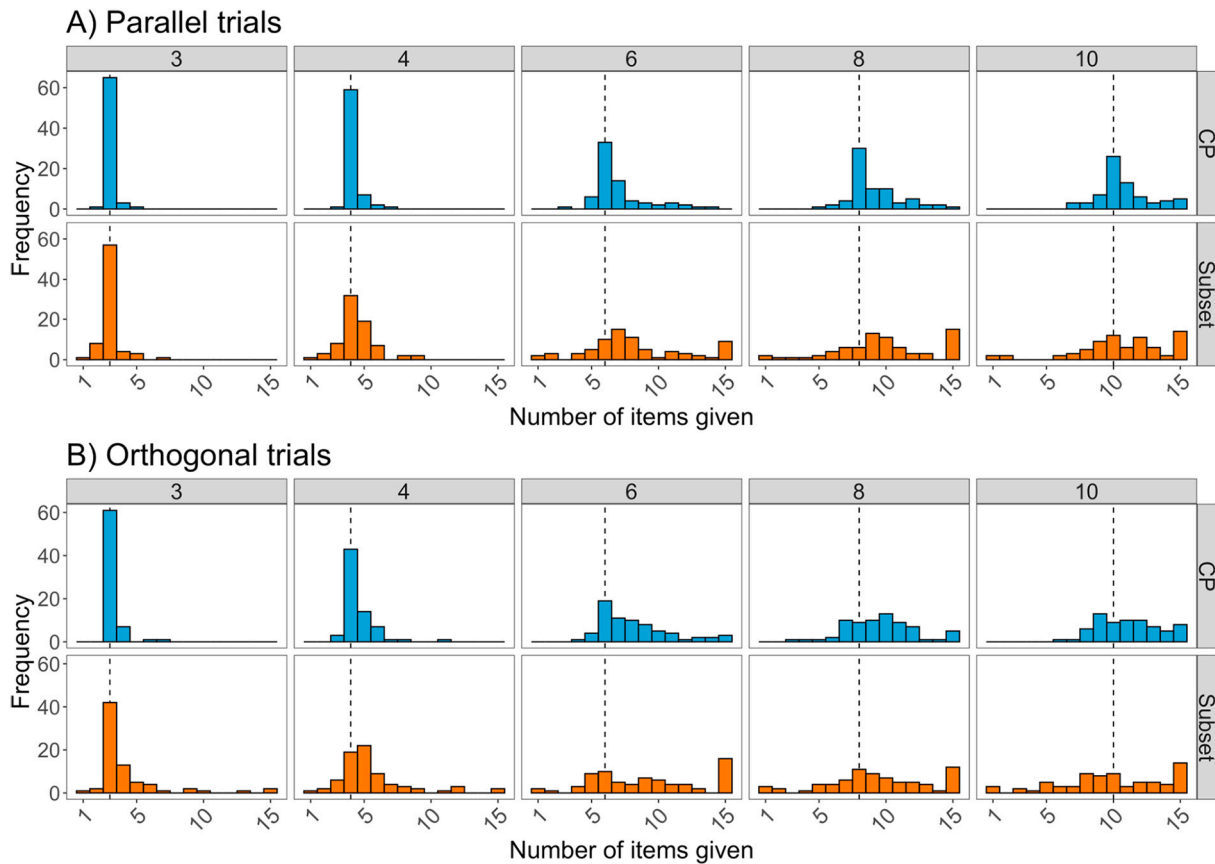
**Fig. 5.** Histogram of response distributions on each set size for a) parallel trials and b) orthogonal trials, grouped by cardinal principle knowledge. Dashed line indicates the correct response for each target quantity.

identify their procedural counting skill. One child did not complete this task and was excluded from this analysis. We constructed a base GLMM with the formula Accuracy (0/1) ~ Set size (continuous, centered and scaled) + Orientation (parallel/orthogonal) + Age (continuous, centered and scaled) + (1 | Participant), and then tested whether the addition of either children's averaged How Many score or their Highest Count improved the fit of the model. A breakdown of children's scores on the How Many task is shown in Table 3.

Model comparison indicated that procedural counting proficiency (measured with the How Many task, included as an ordinal term) was significantly related to set-matching accuracy when not accounting for cardinal principle knowledge ($\chi^2_{(2)} = 8.30, p = .02$; Model 6, Appendix). However, performance on the task did not explain unique variance beyond children's subset-/CP-knower classification ($\chi^2_{(2)} = 0.14, p = .93$; Model 7, Appendix). In contrast, we found a significant effect of cardinal principle knowledge even when controlling for performance on the How Many task ($\chi^2_{(2)} = 17.29, p < .001$; Model 8, Appendix). Because subset knowers, unlike CP-knowers, exhibited variability on the How Many task, we conducted a follow-up analysis that included only

subset knowers, and again found that the addition of a term indicating counting proficiency on the How Many task did not significantly improve the fit of the base model ($\chi^2_{(2)} = 0.16, p = .92$; Model 5, Appendix).[6] Together, the results from the How Many Task indicate that although knowledge of one-to-one counting procedures begins to emerge before children become CP-knowers, this knowledge is not alone sufficient to explain set-matching performance, which only begins to improve when children become CP-knowers. One interpretation of this pattern of results is that one-to-one counting procedures play no role in the development of set-matching. However, the results are also compatible with the idea that one-to-one procedures play a role, but only once they are completely mastered and children become CP-knowers - i.e., that they are necessary but not sufficient. We revisit this question of how counting procedures and set-matching are related in the General Discussion.

In a second analysis, we found that children's verbal counting fluency (measured with the Highest Count task) was unrelated to set-matching performance for children overall ($\chi^2_{(1)} = 0.86, p = .36$; Model 9, Appendix), and failed to account for significant variance beyond cardinal principle knowledge ($\chi^2_{(1)} = 0.22, p = .64$; Model 10,

**Table 3**
Averaged scores from the "How Many" task by knower-level*.

|  | *n* subset knowers | *n* CP-knowers |
|---|---|---|
| Random counters (averaged score 0–0.99) | 20 | 0 |
| Minimal counters (averaged score 1–1.99) | 23 | 2 |
| Proficient counters (averaged score 2–3) | 30 | 68 |

* A score of 3 = errorless counting (coordination of one-to-one tagging and labeling for all objects and correct determination of cardinality); 2 = counting with errors that were fixed (either by restarting the count or backtracking); 1 = maximum of two items counted without error; 0 = random counting.

---

[6] We also replicated these results in a post hoc analysis collapsing over Minimal and Random counters and comparing them to Proficient counters to test whether completely errorless counting predicted set-matching accuracy. Procedural counting proficiency on the How Many task (coded categorically as Proficient/Non-proficient) was significantly predictive of set-matching accuracy when not accounting for cardinal principle knowledge ($\chi^2_{(1)} = 8.12, p = .004$); however, it did not explain unique variance beyond subset-/CP-knower classification ($\chi^2_{(1)} = 0.03, p = .86$), while CP-knower status was uniquely predictive of set-matching performance beyond procedural counting proficiency ($\chi^2_{(1)} = 17.36, p < .001$).

Appendix). In summary, although the CP-/Subset knower distinction was an important predictor of children's set-matching performance, children's verbal counting fluency (Highest Count task) and knowledge of one-to-one counting procedures (How Many task) failed to explain additional variance, leaving open what causes changes in set-matching beyond the CP transition. We return to this topic in the General Discussion.

### 2.3. Discussion

In Study 1 we adapted methods previously used with innumerate groups to explore the effect of symbolic number acquisition on exact number concepts in young US children. Mirroring the findings of some previous studies of innumerate adults, we found that subset knowers were unable to exactly match large sets, whereas CP-knowers performed significantly better, compatible with the proposal that set matching ability is related to numeracy (Everett & Madora, 2012; Gordon, 2004). However, while CP-knowers were more likely to exactly match sets, especially in the parallel condition (where sets could be visually aligned), and regardless of the perceptual similarity of set members, they were well below adult levels (Frank et al., 2012), averaging only around 40% accuracy on parallel trials. This suggests that knowing how to accurately count and generate large sets may not alone provide the ability to match sets, and that this ability emerges later in development. Importantly, these results also indicate that rigorously addressing the question of whether language influences numerical thought need not be conducted in hard-to-access populations, such as the Pirahã. Instead, we found that subset knowers perform almost identically to the innumerate adults reported by Gordon (2004), suggesting that semi-numerate children offer an accessible model for conducting robust and replicable investigations on this topic.

One interpretation of our data is that reasoning about exact equality requires knowledge of the cardinal principle as well as additional learning that occurs after the CP transition. However, another possible explanation of children's struggles on the set-matching task is pragmatic rather than conceptual. Previous tests of numerical equality have found that both CP- and subset knowers often focus on other dimensions of stimuli, such as length (Gelman, 1969; Piaget, 1965; Russac, 1978), and color, shape, or size (Izard et al., 2014; Chan & Mazzocco, 2017; Sarnecka & Wright, 2013), rather than number. However, when the intentions of the experimenter are made more explicit, children's performance sometimes improves markedly. Compatible with this, Negen and Sarnecka (2015) found that although CP-knowers are better than subset knowers at judging which of two dot arrays contains "more dots," this difference disappears when subset knowers are given feedback emphasizing numerosity over cumulative surface area.

Such findings raise the possibility that failures to make exact matches among young children—and perhaps innumerate adults like the Pirahã—may be due to the pragmatically ambiguous nature of the set-matching task, rather than a lack of numerical knowledge. In particular, it has been proposed that the Pirahã might fail to match sets not because they do not have a concept of exact number, but because they do not understand the task (Frank et al., 2008; Laurence & Margolis, 2007). Studies 2 and 3 tested this question in children.

## 3. Study 2

To test whether children's difficulties with set-matching might stem from a failure to understand the task, in Study 2 we explicitly told children that the goal of the task was to match according to number.

### 3.1. Method

This study was not pre-registered.

#### 3.1.1. Participants

We recruited 41 English-speaking children between 3;0 and 5;0 from local preschools, daycare centers, and museums in San Diego, CA. Nine of these children were tested but excluded from analyses for the following reasons: experimenter error ($n = 5$); failure to finish Give-N ($n = 2$); and failure to understand the task ($n = 2$). Demographic information for our final analyzable sample ($n = 32$, $M_{age} = 4.01$ years, $SD_{age} = 0.55$ years; $n$ female $= 19$, $n$ male $= 9$) is shown in Table 4. While we planned to collect data from 80 participants ($n = 40$ each CP- and subset knowers), data collection was halted due to COVID-19.

#### 3.1.2. Stimuli, design, and procedure

##### 3.1.2.1. Set-matching.
The stimuli and procedure for set-matching were identical to Study 1, with the exception of instructions and feedback, that all children received identical fish, that there were no trials for set size 6, and that there were no orthogonal trials. We did not include trials for set size 6 to simplify and shorten the procedure. In addition, because Study 1 found considerable variability in the ostensibly easier parallel task, we omitted orthogonal trials and focused instead on methods that might improve performance for the parallel task.

At the start of the task, the experimenter said, "We're going to play a number matching game with fish in a pond. In this game, your job is to make sure that your pond has the same exact number of fish as my pond *without counting*." On the first training trial (1 fish) the experimenter said, "I'm putting a fish here in my pond. Without counting, can you make your pond have the same number of fish as my pond?" If the child succeeded, the experimenter said, "Great! Your pond has the same exact number of fish as my pond, because there is one fish here, and one fish here!" If the child failed, the experimenter said, "Hmm, your pond doesn't have the same exact number of fish as my pond," pointed out the error, and restarted the training trial. The same process was repeated for the second training trial (2 fish). Only children who could pass both training trials were included in our sample. Once children passed both training trials the experimenter gave only neutral feedback. Order of trials (3, 4, 10, 8; or 4, 3, 8, 10) was counterbalanced across children. Although we used number language in our instructions, we did not allow counting during the task: Counting attempts were again rare, with two CP-knowers attempting to count on 5/78 trials and no subset knower attempts.

##### 3.1.2.2. Give-N.
We again assessed children's knowledge of the cardinal principle with the Give-N task as in Study 1.

### 3.2. Results

We found evidence that explicitly telling children that the purpose of the task was to create a numerical match did not lead either CP- or subset knowers to create more exact matches. A GLMM constructed with both CP- and subset knower data with the formula Accuracy (0/1) ~ Study (Study 1/Study 2) + Set size (continuous, centered and scaled) + Age (continuous, centered and scaled) + (1 | Participant) indicated no effect of Study ($\chi^2_{(1)} = 2.57$, $p = .11$; Model 12, Appendix). Accuracy in Study 2 was equivalent to that of Study 1 for CP-knowers (Fig. 6), with 88% mean accuracy for small numerosities (compared to 89% in Study 1) and 40% for large numerosities (compared to 42% in Study 1). There is some preliminary evidence that drawing attention to number may have slightly improved subset knowers' performance, as they had marginally higher accuracy in comparison to Study 1 with 70% accuracy for small

**Table 4**
Demographic information by knower-level for Study 2.

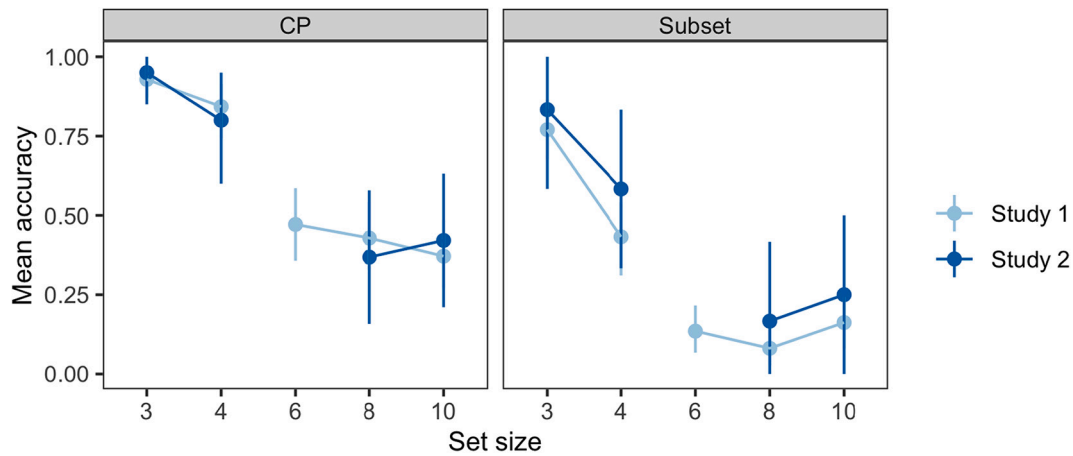|  | $n$ | $M_{age}$ (SD) |
|---|---|---|
| Subset knowers | Total $n = 12$ ($n$ female $= 8$; $n$ male $= 4$) | 3.61 (0.65) |
| CP-knowers | Total $n = 20$ ($n$ female $= 11$; $n$ male $= 9$) | 4.41 (0.45) |

**Fig. 6.** Mean accuracy for CP- and subset knowers on the parallel set-matching task in Studies 1 and 2. Error bars represent 95% confidence intervals computed by nonparametric bootstrap.

numerosities (compared to 61%) and 21% for large numerosities (compared to 13%). However, due to the small number of subset knowers ($n = 12$) and the small differences in performance, these data are not conclusive and overall suggest that if mention of number had an effect, it was very small.

### 3.3. Discussion

In Study 2, we explicitly told children to match sets on the basis of number and found no significant difference in children's pattern of responses. Instead, the results from Study 2 replicate our main findings from Study 1, with subset knowers failing to exactly match large sets, better performance among CP-knowers for both large and small numerosities, and a marked decrease in accuracy in both groups for large numerosities relative to small ones.

These results suggest that children's struggles with set-matching are likely not due to a failure to attend to number in this task. Nevertheless, it remains possible that children still did not understand the goals of the study, since children may be able to make exact matches before they understand the meanings of words contained in the instructions of Study 2, like "exact number." Also, it is possible that children understood the goal of creating exact matches, but were simply unmotivated to pursue this goal for larger sets (perhaps because of the additional effort required).

To address these possibilities, in Study 3 we framed the set-matching task as a "sharing" game, in which each row was given to a character, and children had to ensure that the distribution of resources was "fair" and "the same." Historically, human concerns for fairness and equitable distribution of resources motivated some of the earliest creations of symbolic number systems, often including one-to-one tallies to represent goods and trading (Ifrah, 2000; Schmandt-Besserat, 1992). Children have early-developing (Sloane, Baillargeon, & Premack, 2012) and strong expectations for fairness, and have normative beliefs about equal resource distribution by 3 years of age (Rakoczy, Kaufman, & Lohse, 2016). In addition, recent work has argued that symbolic number knowledge supports equal resource distribution when fairness is at stake. In that work, children were more likely to prefer 'fair' (equal) resource distributions after they had acquired the cardinal principle, suggesting that counting abilities allow children to more accurately assess the equality of sets (Chernyak, Sandham, Harris, & Cordes, 2016). Relatedly, and most relevant to our study, work by Hamamouche, Chernyak, and Cordes (2020), found that children were more likely to divide a set of 8 items equally between two characters when the task was presented as a sharing game than when it was a non-social task. This raises the possibility that children might also match large sets exactly when prompted to create a fair distribution of resources. Tempering this,

however, when children in the Hamamouche et al. study were shown two characters that had unequal amounts (e.g., 6 vs. 2) their ability to correctly add or subtract items to balance the distribution of resources did not differ across sharing and non-social conditions. Also, because children's successes in the division condition were limited to small sets (two pairs of 4), it is unclear which mechanism they used to succeed - e.g., approximation, counting, or one-to-one correspondence. Overall, the sharing literature suggests that children are motivated to favor equal distributions of resources, but leaves open whether children can equally distribute large sets by establishing a one-to-one correspondence between them. Given this, we asked whether fairness concerns might motivate children to attend to exact numerical equality for large sets, leading to improved performance on the set-matching task.

## 4. Study 3

Study 3 tested set-matching in the context of a sharing game.

### 4.1. Methods

This study was pre-registered on OSF (https://osf.io/n9t5p/?view_only=dca1acb1eeed496899e84a9b96015150), and all methodological and analytical choices were pre-registered, unless stated otherwise.

#### 4.1.1. Participants

We recruited 102 English-speaking children between 3;0 and 5;0 from local preschools, daycare centers, and museums in San Diego, CA. Sixteen of these children were tested but excluded from analyses for the following reasons: failure to finish ($n = 11$); parental interference ($n = 3$); experimenter error ($n = 1$); and failure to understand the task ($n = 1$). Demographic information for our final analyzable sample ($n = 86$, $M_{age} = 4.13$ years, $SD_{age} = 0.56$ years) is shown in Table 5.

#### 4.1.2. Stimuli, design, and procedure

##### 4.1.2.1. Sharing game.
We modified the set-matching task from Study 1 by framing it as a "sharing game" between two animals. The experimenter presented children with two blue rectangular boards and two

**Table 5**
Demographic information by knower-level for Study 3.

|  | *n* | $M_{age}$ (*SD*) |
|---|---|---|
| Subset knowers | Total *n* = 33 (*n* female = 16; *n* male = 17) | 3.86 (0.42) |
| CP-knowers | Total *n* = 53 (*n* female = 31; *n* male = 22) | 4.30 (0.56) |

stuffed animals of the same kind (e.g., penguins). Next, the experimenter showed children a bowl with plastic fish and said, "These penguins worked together to catch these fish, but they need our help to share them in their ponds. Do you know what sharing means? Sharing means to make things fair so that people have the same. Because these penguins both worked hard to catch these fish, they should have the same, because that's fair, right?" We emphasized equal collaboration to encourage children to distribute resources equitably (Ng, Heyman, & Barner, 2011).

To ensure children understood the purpose of the task, the experimenter then provided a demonstration of one fair and one unfair share. First, the experimenter placed one fish in each penguin's "pond," and said "Let's ask the penguins if this is fair!" The experimenter made each penguin say, "Yay! That's fair!" and then told the child, "This is fair, because there is a fish in this penguin's pond and a fish in this penguin's pond, so they have the same." Next, the experimenter showed children an unfair share by placing one fish in one pond and no fish in the other pond and saying, "Now, let's ask the penguins if this is fair." The experimenter made both penguins protest, saying "That's not fair!" and then told the child, "This isn't fair, because there's a fish in this penguin's pond, but no fish in this penguin's pond. Let's try again, and this time you can help me make it fair."

After these two demonstrations the experimenter began the training trial. The experimenter removed all fish from the boards, added two fish to one penguin's board and then offered the bowl to the child and saying, "Now it's your turn to share with the other penguin. Remember to make it fair!" After the child was done, the experimenter said, "Let's ask the penguins if this is fair," making the penguins say "Yay! It's fair!" if the child placed two fish in the pond, or "Hey! That's not fair!" if they had not. If the child failed the training trial, the experimenter repeated it a second time. Only children who could pass the training trial and demonstrated an understanding of the task were included in the study.

After training, children received five test trials with neutral feedback for sets of 3, 4, 6, 8, and 10. Order of trials (3, 4, 10, 6, and 8; or 4, 3, 8, 6, and 10) was counterbalanced across participants. The experimenter always placed their set on one board before allowing the child to generate their set. Both the experimenter and child drew fish from the same bowl to emphasize that fish were a common resource; before each trial the experimenter covertly added or subtracted some number of fish to the bowl so that the bowl always contained 15 fish when the child generated their set.

As in Studies 1 and 2, children were not allowed to count. If the experimenter observed a child counting they covered both boards and said, "This isn't a counting game, it's a sharing game!" Counting attempts were again rare: Four CP-knowers attempted to count on 8/265 trials, while four subset knowers attempted counts on 9/170 trials.

*4.1.2.2. Give-N.* We again assessed children's knowledge of the cardinal principle with the Give-N task as in Studies 1 and 2.

*4.2. Results*

To test whether concern for fairness motivates children to generate exact numerical matches for both small and large numerosities we compared data from the sharing task to data from the parallel task in Study 1. To do this we constructed a GLMM with the formula Accuracy (0/1) ~ Study (Study 1/Study 3) + Set size (continuous) + Cardinal principle knowledge (CP/Subset) + Age (continuous) + (1 | Participant). A likelihood ratio test indicated a significant effect of Study ($\chi^2_{(1)}$ = 9.1, $p$ = .003; Model 13, Appendix); surprisingly, however, accuracy was significantly lower for the sharing ($M = 0.41$), relative to matching condition ($M = 0.46$), as shown in Fig. 7. We pursued this difference with a set of follow-up *t*-tests comparing mean accuracy between conditions for CP- and subset knowers by numerosity, and found that it was due to subset knowers' low accuracy on small numerosities in the

sharing condition in comparison to the matching condition ($t(105) = -2.37$, $p = .02$). Consistent with our findings in Studies 1 and 2, we found a significant effect of numeracy, with CP-knowers ($M = 0.53$) more accurate than subset knowers ($M = 0.24$), even controlling for age ($\chi^2_{(1)} = 21.46$, $p < .001$) and set-size ($\chi^2_{(1)} = 7.78$, $p = .006$).

We again investigated differences in the magnitude of error between the two datasets as a less conservative measure of whether framing the task as a sharing game encouraged children to attempt exact matches. As in Study 1, we calculated the absolute error of incorrect trials (|Target item - Response|), and then constructed a LMM with the formula Absolute error ~ Study (Study 1/ Study 3) + Set size (continuous) + Cardinal principle knowledge (CP/Subset) + Age (continuous) + (1 | Participant). Reflecting children's lower accuracy in the sharing task, a likelihood ratio test indicated a significant effect of condition ($\chi^2_{(1)} = 5.4$, $p = 02$; Model 14, Appendix), with higher absolute error for the sharing condition ($M = 2.90$) in comparison to matching ($M = 2.71$). Once again, however, follow-up t-tests of mean performance between conditions indicated that this difference was due to significantly higher absolute error in subset knowers for small numerosities in the sharing task ($t(67) = 2.66$, $p = .009$), while there was no difference in absolute error for CP-knowers between the two conditions ($ps > 0.3$). Overall, children's pattern of responses was generally similar in both the sharing and matching conditions (Fig. 8).

*4.3. Discussion*

The results of Study 3 indicate that children's failures to establish exact matches in the set-matching task are not improved by asking them to ensure resources are shared fairly. In fact, compared to results from Study 1, subset knowers were actually less likely to equally distribute resources for small numerosities in the sharing condition, in line with other findings that the ability to distribute resources equitably for even small sets is linked to CP acquisition (Chernyak et al., 2016). Instead, we replicate the finding that performance on the set-matching task is significantly related to numeracy and that subset knowers struggled to create exact matches for sets >3 items, while CP-knowers were significantly more likely to do so for all set sizes. This suggests that, while numerical competence improves the ability to share fairly, being in a context that requires sharing fairly may not make children any more likely to establish exact equality between the shares.

Studies 2 and 3 tested whether the set-matching task might underestimate children's knowledge. However, another possibility is that the task actually overestimates knowledge of one-to-one correspondence by unwittingly allowing children to count. Although CP-knowers who create exact matches may have learned to do so via one-to-one correspondence, an alternative explanation is that they outperform subset knowers by surreptitiously counting the two sets. While we prevented children from counting out loud in Studies 1–3, more advanced counters may nevertheless count subvocally to generate exact matches, thereby explaining why CP-knowers perform better on average than subset knowers, and why some CP-knowers perform better than others. We tested this possibility in Study 4.

# 5. Study 4

Study 4 tested whether subvocal counting might explain the difference between CP-knowers and subset knowers on set-matching.

*5.1. Method*

All methods and analyses were conducted as pre-registered at https://osf.io/pj4zy unless stated otherwise in-text.

*5.1.1. Participants*

We recruited 38 English-speaking children between 3;0 and 5;0 from local preschools, daycare centers, and museums in San Diego, CA. Ten of
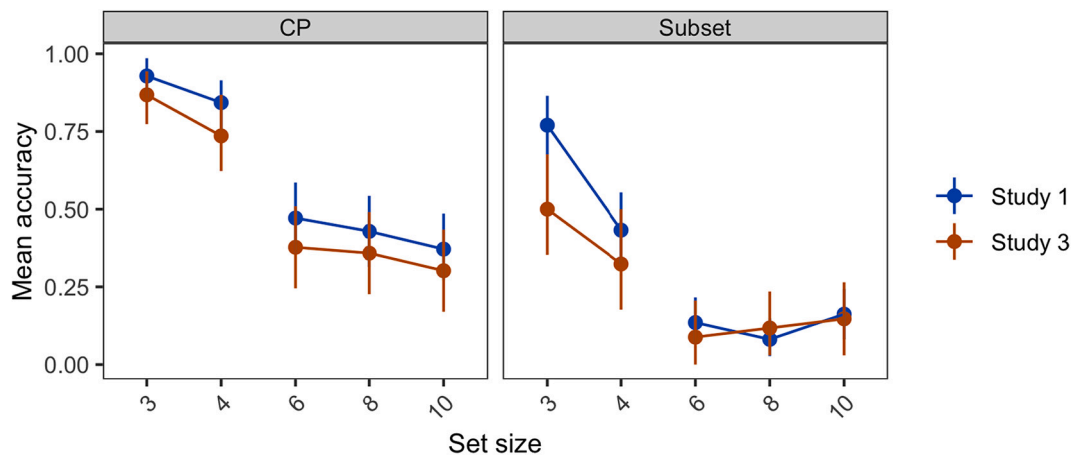
**Fig. 7.** Mean accuracy on Matching (Study 1) and Sharing (Study 3) tasks by knower-level. Error bars indicate 95% confidence intervals, computed by nonparametric bootstrap.
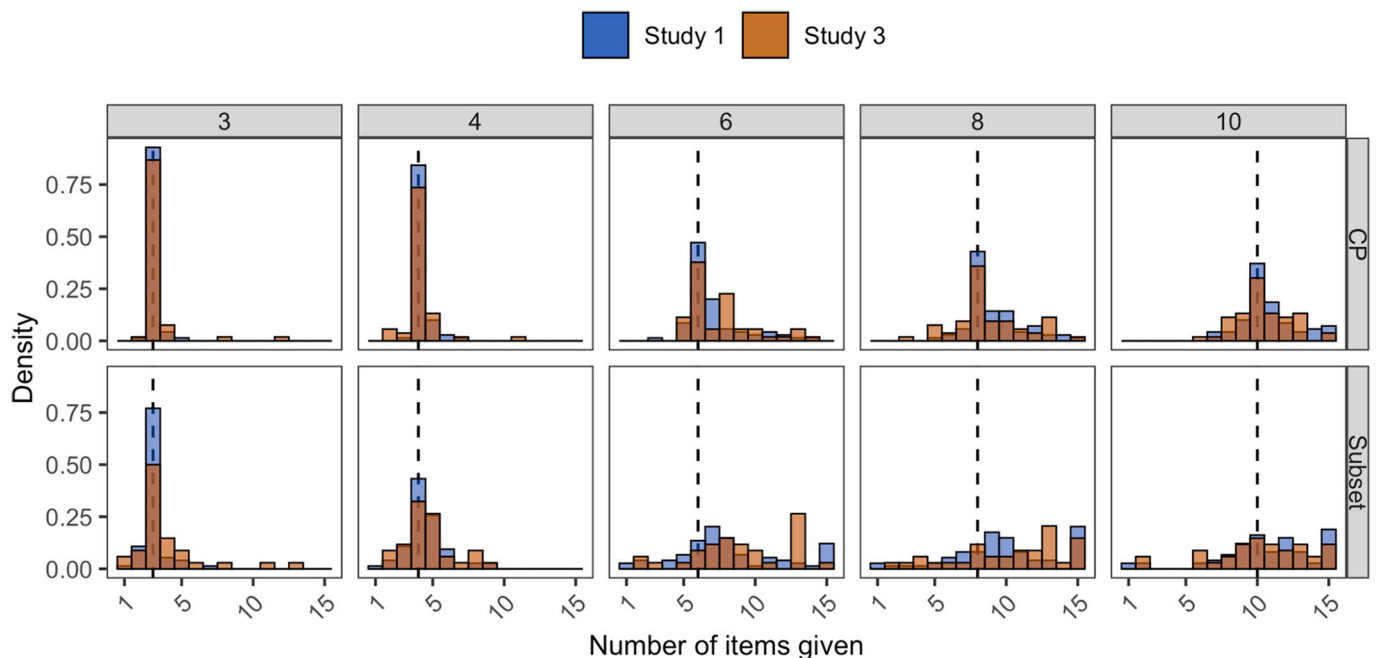


**Fig. 8.** Distribution of number of items given in response to a prompt in Matching (Study 1) and Sharing (Study 3) tasks by knower-level. Dashed lines indicate target set size.

these children were tested but excluded from analyses for the following reasons: not being a CP-knower ($n = 9$); and experimenter error ($n = 1$). After exclusions, our final analyzable sample included 28 CP-knowers ($n$ female $= 12$, $n$ male $= 16$; $M_{age} = 4.38$ years, $SD_{age} = 0.52$ years). While we pre-registered an $n$ of 40, data collection was halted due to COVID-19. We therefore temper our conclusions accordingly.

#### 5.1.2. Stimuli, design, and procedure

*5.1.2.1. Set-matching with follow-up question.* To test whether CP-knowers used subvocal counting to succeed in Studies 1–3, we tested children on the parallel version of the set-matching task. Procedures for training and test trials were identical to Study 1, except that all children received identical fish, and trial order (3, 4, 10, 6, 8; or 4, 3, 8, 6, 10) was counterbalanced across participants. To test whether children counted the experimenter's set, after the last trial the experimenter covered both boards with their arms, and asked the child, "How many fish are in my

pond?" If the child responded "I don't know," they were prompted to guess. The logic of this follow-up question was that if a child had used a count of the experimenter's board to generate their set, they should provide an accurate response when asked to generate the target set's cardinality, and also that this count should reflect the cardinality of the set they generated.

To help control for differences in working memory, and to ensure that all children could remember the cardinality of a recently-counted set, after the child gave their response the experimenter uncovered the boards and said, "Why don't you count the fish in my pond and see how many fish there are?" Immediately after the child finished counting, the experimenter covered both boards again and said, "Now, how many fish are in my pond?"

*5.1.2.2. Give-N.* We again assessed children's knowledge of the cardinal principle with the Give-N task as in Study 1.

*5.2. Results*

We found several pieces of evidence indicating that CP-knowers' performance in the first three studies was not explained by subvocal counting. First, we closely replicated our findings from Study 1: A GLMM with the formula Accuracy (0/1) ~ Study (Study 1/Study 4) + Set size (continuous) + Age (continuous) + (1 | Participant) indicated no effect of experiment ($\chi^2_{(1)} = 0.006$, $p = .94$; Model 15, Appendix). CP-knowers had 93% accuracy for small sets (compared with 89% in Study 1), and 41% accuracy for large sets (compared with 42% in Study 1). Second, around half of children (15/28, 54%) replied "I don't know" when asked how many fish were in the experimenter's pond, despite nearly all children giving the correct count when asked to recall a recently counted set (26/28, 93%). Thirteen children did initially provide a number in response to the experimenter's question before being invited to count, but only two of these children guessed correctly. The remaining 15 children first said "I don't know" in response to the experimenter's question; when these children were prompted to guess, only seven did so, and only one was correct. Finally, and most conclusively, no child's guess matched the number of fish they actually placed on their board, and verbal responses differed from the target by about four on average ($M = 4.25$, $SD = 3.24$). Follow-up analyses showed that children who gave numerical responses did not match sets significantly more accurately than children who were unable to answer the experimenter's initial question ($t(26) = -0.65$, $p = .52$). Together, these results suggest that CP-knowers were not subvocally counting either to (a) count the experimenter's set, or (b) generate their own sets in Studies 1–3.

*5.3. Discussion*

Thus far, we have found evidence that children's ability to exactly match sets is poor prior to acquiring knowledge of the cardinal principle, that it improves when they become CP-knowers, but also that CP-knowers exhibit substantial variability in their performance. The findings from Study 4 suggest that CP-knowers' increased accuracy on this task does not reflect a change in their ability to subvocally count. Although the sample size included in this study was smaller than planned, somewhat limiting generalizability, the results were nevertheless quite categorical, and none of the 28 CP-knowers were able to correctly report how many objects were contained within a row of fish they had just matched. These findings therefore suggest that CP-knowers' advantage over subset knowers is likely due to other changes in strategy. In Study 5, we tested two possibilities—a shift towards the use of one-to-one correspondence versus changes in the acuity of approximation while set-matching.

## 6. Study 5

In Studies 1–4 we found evidence that acquiring the cardinal principle may be implicated in the ability to reason about exact equality. However, as noted by Izard et al. (2014), matching by approximation may frequently yield the same outcome as one-to-one correspondence on the set-matching task. For example, it is possible that all children in Studies 1–4 used an approximation strategy that simply improved as they became CP-knowers, resulting in increasing numbers of correct responses (Shusterman et al., 2016). However, another possibility is that becoming a CP-knower is one step towards learning procedures for matching sets via one-to-one correspondence, and that some children take longer than others to converge on this procedure, resulting in a mix of approximation and one-to-one strategies. In principle, both of these possibilities might produce a distribution of responses similar to the one we observed in Study 1 (Fig. 5). To investigate this issue, we used a modeling approach to identify the strategies underlying children's set-matching behavior in Study 1.

*6.1. Methods*

We compared two models that each represented distinct hypotheses about the source of CP-knowers' accuracy: a) that CP-knowers' increased set-matching accuracy is driven by a refinement of the same approximation strategy used by subset knowers, or, b) that in some cases, CP-knowers instead use a qualitatively different strategy—one-to-one correspondence—rather than approximation.

The first model assumes a purely approximation-based response, with CP-knowers' higher set-matching accuracy due to increases in the acuity of their Approximate Number System. This "Approximation-only" model represents the null hypothesis that CP-knowers' set-matching behavior relies on a numerical mechanism shared with subset knowers, and their performance reflects improvement in this mechanism rather than a qualitative shift in their ability to reason about exact equality. On the Approximation-only model, we expect the distribution of CP-knowers' responses to reflect the scalar variability associated with the ANS, but featuring greater precision than subset knower responses.[7] To test this hypothesis, we used maximum likelihood estimation to fit a Coefficient of Variation (COV) to CP-knowers' responses for large sets (6, 8, and 10) on the parallel task.[8] COV, a measure of how the error in participants' responses (i.e., their generated set) increases with the magnitude of the stimulus, captures the scalar variability associated with approximation (Gordon, 2004; Whalen, Gallistel, & Gelman, 1999). The fitted COV for the Approximation-only model therefore reflects a best estimate of participants' approximation accuracy, allowing for a quantitative assessment of first, whether CP-knower behavior is well-characterized by an approximation strategy, and second, whether the difference between CP-knowers and subset knowers is primarily one of more accurate approximation. Concretely, the model estimates a COV by assuming that subject responses are sampled from a normal distribution centered at the trial magnitude (6, 8, or 10) and with a standard deviation that is a product of the trial magnitude and the COV. In other words, participant responses $y$ are independently sampled according to $y \sim N(m, \sigma)$ for trial magnitude $m \in \{6, 8, 10\}$ and $\sigma = m * COV$. The model estimates the most likely COV given the participant responses $y$.

To test the alternative possibility that CP-knower responses reflect a combination of approximation and exact matching, rather than approximation alone, we also tested a "One-to-one mixture" model in which responses can *either* be generated by approximation or an exact match strategy. Concretely, the model assumes that a proportion $k$ of responses were made through an error-free one-to-one matching procedure, while the remaining *1-k* were generated through approximation as above (Zeigenfuse & Lee, 2010). The model therefore estimates a match percentage $k$, as well as a COV parameter to characterize the approximation behavior in *1-k* percentage of trials. Intuitively, the more correct responses participants make above and beyond what would be expected by approximation alone, the better the One-to-one mixture model should characterize CP-knower data relative to the Approximation-only model.

To evaluate the models, we first compare model fits for the Approximation-only and One-to-one mixture models applied to CP-

---

[7] Another possibility is that CP-knower performance reflects a greater precision in their mapping from ANS magnitudes to verbal labels than subset knowers, rather than a greater precision in ANS representations themselves. For our purposes, we treat these alternatives as equivalent, since both are consistent with an improved "Approximation-only" strategy.

[8] We did not include small numerosities (3, 4) in the fitting process. It is possible that participants may have recruited a mechanism other than approximation (e.g., subitizing, object-tracking) for these small sets (Feigenson et al., 2004); given this uncertainty, we restricted our analysis to the range where children would need approximation in the absence of exact matching, which is a conservative assumption, since subitizing behavior likely looks more like exact matching.

knower responses in Study 1. Further, using the estimated model parameters, we simulate response data for each model to provide an intuitive comparison to the empirical distribution obtained in Study 1. Next, we fit both models to the subset knower responses from Study 1; we expect that subset knowers are primarily using a noisy approximation strategy, which therefore provides a useful comparison for the model fits and parameter estimates from the CP-knower responses. In this vein, we also fit both models to the *orthogonal* responses from Study 1, since we expected responses there to reflect an approximation strategy by both subset and CP-knowers.

If the responses from CP-knowers simply reflect an accurate approximation strategy, we should see little if any improvement by adding the exact match parameter $k$ in our One-to-one mixture model and the fitted value for $k$ should be small. In this case, we would also expect the simulated data from the Approximation-only model to closely match the CP-knower responses. On the other hand, if CP-knowers are using one-to-one correspondence in their responses, as the data in Fig. 5A suggests, then the One-to-one mixture model should provide a better fit and may have a large estimated match percent $k$, and should produce simulated data that more closely resemble the results from Study 1. Finally, because subset knowers are less likely to use an exact match strategy than CP-knowers, the results of the model comparison with CP-knower data should be attenuated among subset knowers.

### 6.2. Results

The Approximation-only model estimated a COV of 0.32 for the CP-knower responses. This COV is slightly higher than 0.15, the value typically associated with approximation for both numerate and innumerate adult approximation (Whalen, Gallistel, & Gelman, 1999; Gordon, 2004; Frank et al., 2008; Frank et al., 2012), but is comparable to COVs reported in previous developmental work using a similar task that could only be solved by approximation (Shusterman et al., 2017).[9] Using summary values reported from Experiment 1 of Shusterman et al. (2017) for sets of 5–9 (a similar range as the large numerosities in Study 1), we calculated a mean COV of 0.27, which aligns with our fitted value of 0.32 in the present task.

Using the COV of 0.32 estimated by the Approximation-only model, we simulated independent responses for 10,000 participants approximating sets of 6, 8, and 10, and compared the resulting distribution to CP-knower data from Study 1 (Fig. 9, top). Although the simulated data matches the spread of the behavioral data, it fails to capture one key aspect: CP-knowers' modal responses were centered on the correct set size for all large numerosities, resulting in a large peak. For simulated responses from the Approximation-only model to capture the large number of correct responses in the CP-knower data, it would be necessary to assume a much lower COV than the fitted estimate (reflecting greater approximation precision). However, the simulated response distribution resulting from this modification would deviate from the empirical CP-knower distribution in two ways: it would include many more responses which are *close,* but not identical, to the correct number than what we find among CP-knowers, and it would fail to capture the spread of the empirical response distribution. In short, the high variance and the large number of exact matches exhibited by CP-knowers is not readily captured by the Approximation-only model, in which responses are normally distributed based on a stable, group-level COV.

In contrast, the One-to-one mixture model provides a close fit to the empirical distribution of CP-knower responses from Study 1. It estimated an exact match proportion $k = 0.3$, suggesting that roughly 30% of trials were associated with exact matches rather than approximating.

As a reminder, since the model fitting treats trials independently, this $k$ value indicates the percentage of all trials that are best fit by assuming exact matching behavior rather than responses generated via approximation. The COV estimate for approximation responses was 0.32, identical to the value obtained by the Approximation-only model. The stability of this fitted COV estimate in both models, along with the relatively high exact match proportion, further supports the conclusion that a combination of approximate and one-to-one strategies better captures CP-knowers' behavior on the set-matching task. Indeed, the simulated data from the One-to-one mixture model (Fig. 9, bottom) provides a close approximation of the empirical distribution of CP-knower responses (Fig. 9, top). In addition to the qualitatively better fit provided by the One-to-one mixture model, model comparison using Bayesian Information Criterion (which penalizes the One-to-one mixture model for its additional parameter) suggests that the One-to-one mixture model provides a better fit to our behavioral data (BIC = 867.4, $n = 210$, $k = 2$, $L = -428.3$) than the Approximation-only model (BIC = 945.1, $n = 210$, $k = 1$, $L = -469.9$).

Finally, we asked whether this mixture of approximation and one-to-one strategies was unique to CP-knowers by repeating the same process with the subset knower data from Study 1. Here, the Approximation-only model estimated a COV of 0.57[10] (again, similar to a COV of 0.53 calculated for subset knowers from Shusterman et al., 2017). Critically, the One-to-one mixture model produced a similar COV estimate of 0.56, but with an exact match estimate $k$ of only 4% of all trials. Indeed, unlike our finding with CP-knowers, the addition of the exact match term did not produce a better fit to the subset knower data (BIC = 1213.3, $n = 222$, $k = 2$, $L = -601.3$) relative to the Approximation-only model (BIC = 1211.9, $n = 222$, $k = 1$, $L = -603.2$). Consistent with the possibility that many subset knowers may have simply provided the maximum allowable value as a distinct "strategy", a separate model in which we estimated a "give-all percentage" $p$ similar to the exact match percentage in our One-to-one mixture model estimated that 15% of trials followed the give-all strategy and produced a much better fit to the data than either the One-to-one mixture or Approximation-only model ("give-all" BIC = 1140.1, $n = 222$, $k = 2$, $L = -564.7$).

The previous model results focused on set-matching behavior in the *parallel* condition of Study 1, illustrating the role of one-to-one correspondence among CP-knowers but not subset knowers. In contrast, CP-knower responses in the *orthogonal* trials of Study 1 show no evidence of exact matching behavior for sets larger than 3–4 (Fig. 5B), suggesting that the Approximation-only model may better capture these data. Indeed, similar to the pattern of results with subset knower data on the parallel trials, the Approximation-only model estimates a COV of 0.46 for the CP-knower orthogonal responses while the One-to-one mixture model estimates a similar COV of 0.49 and an exact match value $k$ of only 7%. Unsurprisingly given these results, a model comparison reveals little improvement from the added exact match parameter (Approximation-only: BIC = 1068.9, $n = 210$, $k = 1$, $L = -531.8$; One-to-one mixture: BIC = 1067.7, $n = 210$, $k = 2$, $L = -528.5$).[11] Thus, the value of the One-to-one mixture model's exact match parameter appears

---

[9] In the study, 3- to 5-year-old CP-knowers were asked to fetch socks for a caterpillar with 1–20 ft (Shusterman et al., 2017). Because the caterpillar and socks were on different tables and all socks had to be retrieved in a single trip children could only use an approximation strategy.

[10] While earlier model fits did not use a prior in the maximum likelihood estimation, here we included a loose prior on the COV equal to the fitted value from the CP-knower responses. Without such a prior, the model estimates a subset knower COV closer to 0.7; this likely reflects an attempt to fit the many subset knower responses of 15 (the maximum value) as approximations greater than or equal to 15. The results reported here remain consistent even if we do not include a prior in the model fitting process for subset knower responses.

[11] The responses from subset knowers offer little point of comparison here. In fact, both models fail to converge on parameter estimates for the orthogonal trials from subset knowers. This is likely due to the fact that while CP-knower responses are roughly normally distributed, subset knower responses appear almost uniform for magnitudes above 4 (Figure 5B), with little discernible difference between response distributions for 6, 8, and 10, making it impossible to estimate a COV for subset knower responses.
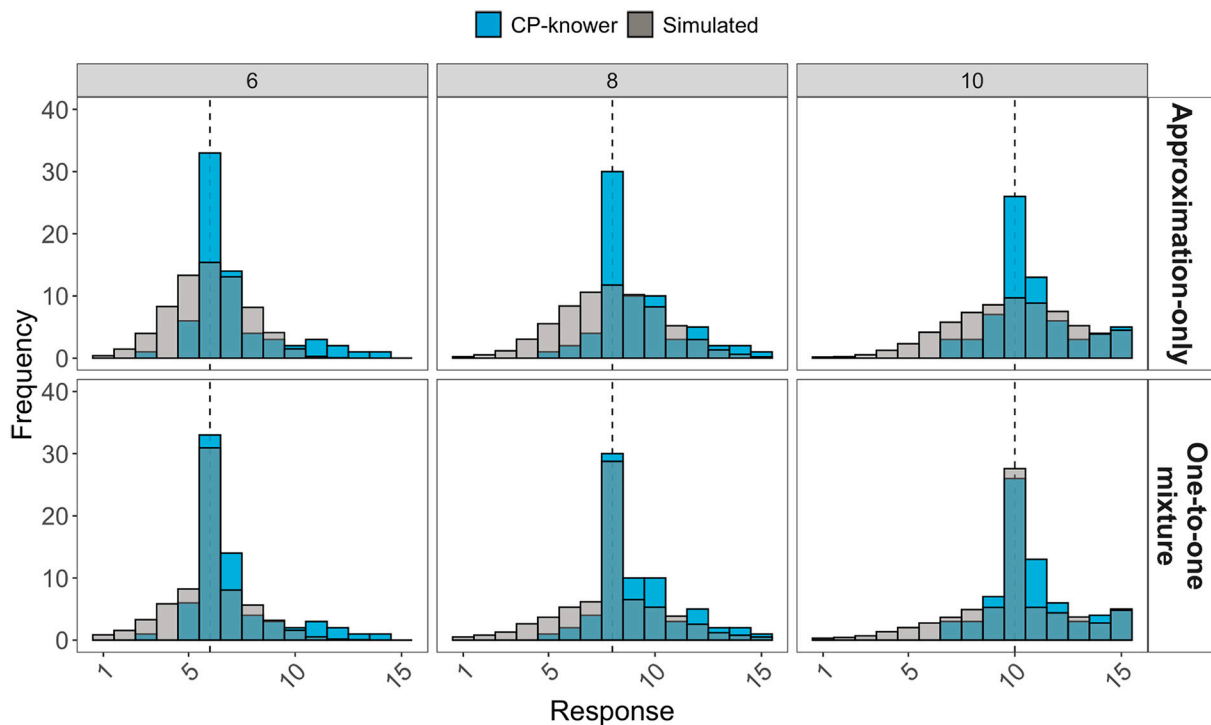
**Fig. 9.** Behavioral data from CP-knowers in Study 1 overlaid with simulated data from the Approximation-only (top) and One-to-one mixture (bottom) models for set sizes of 6, 8, and 10.

to apply selectively to CP-knower responses on parallel trials, but not subset knower responses on the same trials or CP-knower responses on the orthogonal trials in Study 1.

*6.3. Discussion*

The findings of Study 5 suggest that CP-knowers' increased set-matching accuracy relative to subset knowers reflects a qualitative shift in the strategies they used to generate exact matches, rather than a quantitative change in the acuity of approximation. Using a modeling approach, we found that a One-to-one mixture model, which assumed that CP-knowers deployed a combination of approximation and one-to-one strategies, was better able to capture CP-knower behavior compared to an Approximation-only model. Importantly, this effect was limited to CP-knowers on parallel trials; we found little evidence that any subset knowers were using one-to-one correspondence over approximation to establish exact matches, and little evidence for exact matching behavior in CP-knower responses on orthogonal trials. These results suggest that becoming a CP-knower may constitute an important step in learning how one-to-one correspondence can be used to establish exact equality. However, the results of these models also provide support for the hypothesis that additional learning must occur after this point, as our One-to-one mixture model indicated a large proportion of trials on which CP-knowers were likely still approximating.

**7. General discussion**

Across five studies, we found evidence that the ability to establish exact equality between sets is related to children's ability to accurately construct large sets through counting, but also that counting ability is not alone sufficient. Similar to innumerate adults described in some previous studies (Everett & Madora, 2012; Gordon, 2004), many children failed to create exact matches for sets with greater than 3 items even when asked to match sets in parallel lines. In particular, we found that subset knowers—and some CP-knowers—resembled innumerate adults in this respect. Although subset knowers were unlike the Pirahã in

that they did understand the meanings of some number words, they nevertheless generated sets either by noisily approximating or through the use of error-prone heuristics like giving all items in a set. In contrast, CP-knowers generated exact matches for both small and large sets significantly more often, though some proportion failed to do so consistently. Studies 2, 3, and 4 found that these results were not easily explained either by failure to understand that the goal of the task was to match sets by number, insufficient motivation to do so, or by surreptitious counting on the part of CP-knowers. Finally, model simulations in Study 5 provided evidence that CP-knowers' advantage on this task was not due to increases in approximation acuity after the subset knower stage, but instead reflected a shift towards a one-to-one matching strategy.

These results have important implications for understanding the relationship between symbolic number and exact equality. First, the results address conflicting findings from studies in innumerate populations. Previous evidence from innumerate groups was limited in several respects. First, because these studies included participants who lacked any number language, they were unable to isolate which components of symbolic learning might impact numerical reasoning. Second, because these studies have focused on small and remote indigenous populations, researchers have understandably been limited to relatively small sample sizes, and replication and extension of results by independent labs is challenging. By using the methods developed for studying innumerate populations to study semi-numerate children, our studies address the same debate, but allow for a finer-grained understanding of the relation between numeracy and the ability to reason about exact equality. In particular, we ruled out several hypotheses, including the idea that reasoning about large exact number might emerge after learning 2–3 small number words, or that it might depend exclusively on the ability to perform rote counting procedures (e.g., pointing at objects and placing them in one-to-one correspondence with number words). Also, we found that knowledge of the cardinal principle appears to play some role, though even this is likely not sufficient, since even CP-knowers are accurate only about half the time when matching larger sets. These findings are compatible with the proposal that

symbolic number is implicated in the ability to reason about exact numerical equality for large numerosities (Everett & Madora, 2012; Gordon, 2004), while also greatly narrowing the possible roles that symbolic number might play. Also, this work suggests that to further study effects of symbolic number on non-verbal numerical thought, it may be unnecessary to study members of remote indigenous groups since young children perform similarly and offer greater variability in their exposure to symbolic number.

Although our studies were designed to address the debate regarding the role of symbolic systems in reasoning about exact equality that began in work on innumerate indigenous groups, caution should be taken in generalizing from our results back to those populations. The evidence we present is compatible with the thesis that symbolic number systems —specifically, verbal numerals—play an important role in human reasoning about exact equality, compatible with the conclusion reached by Gordon (2004) and others (Everett & Madora, 2012; Pica et al., 2004). However, members of the Pirahã, Mund{}uruku, and other innumerate groups are different from children in very many ways, including their age, cognitive and linguistic development, exposure to formal education, exposure to formal testing environments, familiarity with western researchers, etc. Therefore, it remains possible that the way in which semi-numerate US children interpret experiments is entirely different, and that different mechanisms explain ostensibly similar behaviors across groups. For example, it remains possible that innumerate Pirahã adults are capable of making exact matches, but that they only fail when they do not understand that exact matches are the goal intended by the experimenter (Frank et al., 2008; Laurence & Margolis, 2007). As noted in the Introduction, some existing work begins to address this, by testing semi-numerate Tsimane children's reasoning about exact number (Jara-Ettinger et al., 2017; Piantadosi, Jara-Ettinger, & Gibson, 2014). Although these studies use slightly different methods from both the work reported here and studies in the Pirahã and Mund{}uruku, they converge in finding that the ability to generate exactly matching sets correlates with mastery of counting, but that knowledge of counting may not be the lone factor in explaining performance.[12] Such work provides preliminary reason to believe that mechanisms at play in semi-numerate US children may also apply to other groups.

In addition to addressing differing accounts from work in innumerate groups, our findings also provide an important framework for synthesizing previous studies of children's understanding of exact equality. As noted in the Introduction, past work in children has arrived at different conclusions that have been difficult to reconcile due to the use of different methods. For example, some studies have argued that because subset knowers judge that adding items to a set requires changing the number word that describes it—even for large numbers they do not yet know—children must know that large numbers have exact meanings, and therefore must have some understanding of exact equality, too. Other studies have argued that, to become CP-knowers, children must acquire knowledge of one-to-one correspondence and exact equality, and so claim that this knowledge should be present in all CP-knowers (Lipton & Spelke, 2006; Sarnecka & Carey, 2008; Sarnecka & Wright, 2013). Finally, other studies have reported poor understanding of equality in much older CP-knowers (Piaget, 1965; Russac, 1978), sometimes even for small sets, though these studies failed to test whether understanding improved with changes in numeracy, or whether children understood the goals of the experiment.

Our findings help to adjudicate between these previous claims by

relating changes in numeracy in early development to children's abilities to make exact, non-symbolic matches of sets. Our repeated finding that subset knowers failed to match sets greater than 3 items, and that their set-matching was impacted by the similarity of objects across sets, is compatible with the finding of Izard et al. (2014) that subset knowers do not reliably track the numerosity of sets of objects (or empty branches) greater than 3, and that object identity may be especially important to matching sets for children at this level of numeracy. Also, our finding that CP-knowers' are better at creating exact matches for large numerosities, and that they do so even when sets are made up of dissimilar objects, indicates that learning how to generate sets via counting likely plays an important role in helping children to reason abstractly about numerical equality for both small and large sets (Mix, 1999; Negen & Sarnecka, 2009; Sarnecka & Wright, 2013). However, contrary to some claims in the literature (Carey, 2004; Sarnecka & Carey, 2008; Sarnecka & Wright, 2013), we fail to find evidence that acquiring the cardinal principle entails a complete understanding of exact equality. Instead, our results, coupled with the model simulations of Study 5, show that while CP-knowers as a group were more likely to match sets using one-to-one correspondence, many children appear to still approximate. Thus, though counting has a significant impact on children's ability to match large sets, this ability may continue to develop for some years after children become CP-knowers.

Perhaps the most interesting question raised by this work is precisely *how* learning to accurately count might facilitate reasoning about exact equality, and why counting alone might be insufficient. One reading of the findings presented here is that knowledge of exact equality is constructed in development; children may initially represent large numerosities only approximately, and may build new conceptual resources when they acquire symbolic number representations in language, consistent with the Whorfian claims of Gordon (2004). However, another possibility is that a concept of exact equality is innately available, and that innumerate and semi-numerate humans simply lack the necessary procedures for applying it to large sets. Such a hypothesis would make sense of the observation that children and innumerate groups are relatively good at matching small sets, and the claim that infants can represent and compare sets of up to ~3, even when hidden from sight. In their work on parallel individuation, Feigenson and Carey (2005) argued that representations of individual objects, which they call 'object-files', "implicitly contain information about the number of objects in an array, in that there is a one-to-one correspondence between object-files and objects." (p. 297). Critically, on their hypothesis, object files that correspond to each object in a set are activated simultaneously and in parallel, such that the number of objects that can be represented at any given time is constrained by the resources of working memory, with a limit of around 3–4 (Feigenson & Carey, 2005; Leslie, Xu, Tremoulet, & Scholl, 1998). One live possibility, then, is that innumerate humans *can* represent exact equality via one-to-one correspondence, but do so only with a parallel, capacity-limited, verification procedure.

When children become CP-knowers, they master a sequential verification procedure—i.e., counting—that allows the evaluation of one-to-one correspondence for sets of any size, effectively bypassing the capacity limits of working memory. As noted by Gelman and Gallistel (1978), a key characteristic of all human counting systems is that they allow sets to be partitioned into two groups, i.e., those that have been counted, and those that have not. Objects in an array are ticked off by labeling them, touching them, setting them aside, etc., often using a unique label (or body position) to distinguish each step in the partitioning procedure. Crucially, the working memory resources required by such partitioning procedures do not differ as a function of set size—the 5th step in the procedure is no more or less resource intensive than the 10th step, or the Nth step. In other words, counting constitutes a Markov process that requires a constant amount of memory: As the size of the set increases, it is always represented by the single state of the current count, while the elements of the set—no matter how many—need never

---

[12] The findings of Jara-Ettinger et al. (2017) are consistent with the results reported here in that they found approximately half of CP-knowers (12/26) could not state whether two sets in one-to-one correspondence were "of equal quantity" through a number of transformations. Although they highlight that some subset knowers succeeded, only a few (6/37) were at ceiling on this task, a rate potentially compatible with an approximation strategy.

be stored in memory at all. Consequently, unlike parallel individuation, which is capacity limited due to its parallel one-to-one verification procedure, counting allows humans to assess equality via a similar one-to-one operation, but without a set-size limit.

Notably, like the counting procedure, matching sets via one-to-one correspondence is also a Markov process. Much like counting, it requires tracking which items in a set have most recently been paired off and partitioning them from items that have not been paired off yet. Given this structural similarity, we might expect children who have learned this basic counting procedure to notice, by analogy, that a similar process might be used to match large sets. However, as we showed here, knowledge of counting alone is clearly not sufficient, since even some subset knowers know how to count and partition sets (despite failing to recognize its significance to cardinality). Critically, though, understanding Hume's principle requires more than knowing how to deploy a one-to-one correspondence procedure—it also requires recognizing the significance of this relation to number, and that two sets that exhibit one-to-one correspondence are numerically equal (Decock, 2008; Frege, 1884; Heck, 2000; Hume, 1739/2003; Wright, 1999). By most accounts, this is precisely what subset knowers do not appreciate about counting—that the one-to-one relation it creates between objects and labels provides a system for labeling the cardinalities of sets, and thus for exactly matching any two sets with the same label.

Still, although children might have the conceptual basis for forming an analogy between counting and set-matching by the time they become CP-knowers, there is clearly something more that is required to notice this analogy, since not all CP-knowers successfully make exact matches. One hypothesis, tested in Study 3, is that the same social concerns for fair divisions of resources that led humans to develop symbolic number systems in the first place (Ifrah, 2000; Schmandt-Besserat, 1992) might lead children to want to ensure that two recipients get exactly matching shares. Although previous research has established that CP-knowers generate more exact matches than subset knowers when asked to share (Chernyak et al., 2016; Chernyak, et al., 2019), we found that presenting the set matching task in a sharing context did not make either CP or subset knowers more likely to generate exact matches, even for smaller set sizes (in contrast with Hamamouche et al., 2020). While those children who are already capable of generating exactly equal sets may apply that ability in the context of fair sharing, our findings suggest that a sharing context may not enable children who would not otherwise do so to generate exactly equal distributions.

An alternative candidate hypothesis is that making this analogy depends on experience generating and evaluating the outcomes of counting procedures. Practice generating sets through counting provides opportunities to notice how counts with the same outcome contain the same number of items and vice versa, an ability which appears to emerge after children become CP-knowers (Frydman & Bryant, 1988; Muldoon, Lewis, & Freeman, 2003). Also, it may allow children to notice that adding exactly 1 item to a set changes a count, and therefore that differences of even one item result in unequal amounts. This too, appears to emerge after the CP-knower stage (Condry & Spelke, 2008; Lipton & Spelke, 2006; Sarnecka & Wright, 2013). In general, practice counting provides CP-knowers the opportunity to notice how a one-to-one procedure like counting produces similar or different results. Such experiences might help children converge on the use of a similar non-symbolic procedure for assessing the equality of sets. In other words, simple learned procedures may form the basis for creating models that allow children to discover the principles that these procedures instantiate.

While our findings suggest that reasoning about exact equality may emerge through learning how counting is used to generate sets, there are several limitations to the studies described here that raise alternative explanations. First, it is possible that some proportion of the set-matching failures we report in both subset- and CP-knowers arise from performance demands that mask an otherwise mature understanding of one-to-one correspondence. Although the set-matching task is designed to be minimally demanding, children still had to visually saccade across

both ponds, remember which fish they had already paired off, and coordinate the retrieval and arrangement of new fish; it is possible that even a child who has learned to count and has acquired a non-symbolic set partitioning strategy based on this might still struggle to implement such a procedure due to limits in working memory. Second, although we attempted to highlight the numerical nature of the task in Study 2, it is still possible that some children may not have fully understood the task, perhaps because they did not understand what was meant by a request to provide "the same exact number." Likewise, although Study 3 found that highlighting the importance of equitable resource allocation in a collaborative game did not help children generate exact matches, it is possible that a context in which children share their own more valuable resources, instead of allocating symbolic resources to third parties, could provide the requisite motivation. Finally, although our modeling approach provides some preliminary evidence that CP-knowers' increased accuracy relative to subset knowers reflects a shift towards one-to-one procedures, larger datasets can both refine this model (e.g., by allowing us to compute individual-level COV values) and provide additional empirical tests of its validity.

Together, our results point to a new direction in the study of the relation between symbolic representations and the ability to reason about exact equality. First, the finding that this ability may emerge gradually after learning to count indicates that the most fruitful future studies may not lie in studies of innumerate groups who completely lack a symbolic number system, or have only a few number words. Instead, this question may be better tested in children from diverse numerate cultures which vary along more dimensions of numeracy (such as exposure to counting or the structure of the count list; Almoammer et al., 2013; Jara-Ettinger et al., 2017; Schneider et al., 2021). Second, our results suggest that such work should not only focus on changes in the ability to reason about exact equality before and after acquisition of the cardinal principle, but also on learning unfold after the acquisition of the cardinal principle, as children gain additional experience deploying counting procedures. Such investigations may provide insights not only on the ontogenesis of exact number concepts, but also on processes by which children extract conceptual knowledge from procedural routines.

## Acknowledgements

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104952.

## References

Almoammer, A., Sullivan, J., Donlan, C., Marušič, F., O'Donnell, T., & Barner, D. (2013). Grammatical morphology as a source of early number word meanings. *Proceedings of the National Academy of Sciences, 110*(46), 18448–18453.

Barner, D. (2017). Language, procedures, and the non-perceptual origin of number word meanings. *Journal of Child Language, 44*(3), 553–590.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Dai, B. (2015). *lme4 (Version 1.1-8)*.

Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition, 124*(2), 156–182.

Boolos, G. (1986, January). Saving Frege from contradiction. In *, 87. Proceedings of the Aristotelian Society* (pp. 137–151). Aristotelian Society, Wiley.

Briars, D., & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology, 20*(4), 607.

Brooks, N., Audet, J., & Barner, D. (2013). Pragmatic inference, not semantic competence, guides 3-year-olds' interpretation of unknown number words. *Developmental Psychology, 49*(6), 1066.

Butterworth, B. (1999). *The mathematical brain*. London: Macmillan.

Butterworth, B., & Reeve, R. (2008). Verbal counting and spatial strategies in numerical tasks: Evidence from indigenous Australia. *Philosophical Psychology, 21*(4), 443–457.

Cantlon, J., Fink, R., Safford, K., & Brannon, E. M. (2007). Heterogeneity impairs numerical matching but not numerical ordering in preschool children. *Developmental Science, 10*(4), 431–440.

Cantrell, L., Boyer, T. W., Cordes, S., & Smith, L. B. (2015). Signal clarity: An account of the variability in infant quantity discrimination tasks. *Developmental Science, 18*(6), 877–893.

Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition, 128*(3), 331–352.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus, 133*(1), 59–68.

Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences, 23*(10), 823–835.

Chan, J. Y. C., & Mazzocco, M. M. (2017). Competing features influence children's attention to number. *Journal of Experimental Child Psychology, 156*, 62–81.

Chernyak, N., Sandham, B., Harris, P. L., & Cordes, S. (2016). Numerical cognition explains age-related changes in third-party sharing. *Developmental Psychology, 52*, 1555–1562. https://doi.org/10.1037/dev0000196

Chernyak, N., Harris, P. L., & Cordes, S. (2019). Explaining early moral hypocrisy: Numerical cognition promotes equal sharing behavior in preschool-aged children. *Developmental science, 22*(1), Article e12695.

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.

Chrisomalis, S. (2021). The scope of linguistic relativity in graphic and lexical numeration. *Language & Communication, 76*, 1–12.

Clarke, S., & Beck, J. (2021). The number sense represents (rational) numbers. *Behavioral and Brain Sciences*, 1–57.

Coben, D., Colwell, D., Macrae, S., Boaler, J., Brown, M., & Rhodes, V. (2003). *Adult numeracy: Review of research and related literature*. London: National Research and Development Centre for Adult Literacy and Numeracy.

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General, 137*(1), 22.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123*(1), 162–173.

Decock, L. (2008). The conceptual basis of numerical abilities: One-to-one correspondence versus the successor relation. *Philosophical Psychology, 21*(4), 459–473.

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

Everett, C., & Madora, K. (2012). Quantity recognition among speakers of an anumeric language. *Cognitive Science, 36*(1), 130–141.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6*(5), 568–584.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*(3), 295–313.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Flaherty, M., & Senghas, A. (2011). Numerosity and number signs in deaf Nicaraguan adults. *Cognition, 121*(3), 427–436.

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General, 141*(1), 134.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition, 108*(3), 819–824.

Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology, 64*(1–2), 74–92.

Frege, G. (1884). *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl. w. Koebner*.

Frydman, O., & Bryant, P. (1988). Sharing and the understanding of number equivalence by young children. *Cognitive Development, 3*(4), 323–339.

Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development*, 1158–1171.

Fuson, K. C. (1988). *Children's counting and concepts of number* (Springer Science & Business Media).

Fuson, K. C., Lyons, B. G., Pergament, G. G., Hall, J. W., & Kwon, Y. (1988). Effects of collection terms on class-inclusion and on number tasks. *Cognitive Psychology, 20*(1), 96–120.

Fuson, K. C., Secada, W. G., & Hall, J. W. (1983). Matching, counting, and conservation of numerical equivalence. *Child Development*, 91–97.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1–2), 43–74.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*(2), 59–65.

Garegae, K. G., & Lekoko, R. N. (2006). Intuitive mathematical knowledge as an essential aspect of contemporary adult learning: A case of women street vendors in the city of Gaborone. *Literacy and Numeracy Studies, 15*(1), 61–77.

Gelman, R. (1969). Conservation acquisition: A problem of learning to attend to relevant attributes. *Journal of Experimental Child Psychology, 7*(2), 167–187.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences, 9*(1), 6–10.

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

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496–499.

Hamamouche, K., Chernyak, N., & Cordes, S. (2020). Sharing scenarios facilitate division performance in preschoolers. *Cognitive Development, 56*, 100954.

Hatano, G., Miyake, Y., & Binks, M. G. (1977). Performance of expert abacus operators. *Cognition, 5*(1), 47–55.

Heck, R. G. (2000). Cardinality, counting, and equinumerosity. *Notre Dame Journal of Formal Logic, 41*(3), 187–209.

Hume, D. (1739/2003). *A treatise of human nature*. Courier Corporation.

Ifrah, G. (2000). *The universal history of numbers*. London: Harvill.

Izard, V., Streri, A., & Spelke, E. S. (2014). Toward exact number: Young children use one-to-one correspondence to measure set identity but not numerical equality. *Cognitive Psychology, 72*, 27–53.

Jara-Ettinger, J., Piantadosi, S., Spelke, E. S., Levy, R., & Gibson, E. (2017). Mastery of the logic of natural numbers is not the result of mastery of counting: Evidence from late counters. *Developmental Science, 20*(6), Article e12459.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist, 86*(1), 65–79.

Kifleyesus, A. (2009). Jeberti women Traders' innumeracy. Its impact on commercial activity in Eritrea. *L'Homme. Revue française d'anthropologie, 189*, 49–80.

Laurence, S., & Margolis, E. (2007). Linguistic determinism and the innate basis of number. *The Innate Mind, 3*, 139–169.

LeFevre, J. A., Polyzoi, E., Skwarchuk, S. L., Fast, L., & Sowinski, C. (2010). Do home numeracy and literacy practices of Greek and Canadian parents predict the numeracy skills of kindergarten children? *International Journal of Early Years Education, 18*(1), 55–70.

Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences, 12*(6), 213–218.

Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing what and where systems. *Trends in Cognitive Sciences, 2*(1), 10–18.

Lipton, J. S., & Spelke, E. S. (2006). Preschool children master the logic of number word meanings. *Cognition, 98*(3), B57–B66.

Markman, E. M. (1979). Classes and collections: Conceptual organization and numerical abilities. *Cognitive Psychology, 11*(4), 395–411.

Mattessich, R. (1987). Prehistoric accounting and the problem of representation: On recent archeological evidence of the middle-east from 8000 BC to 3000 BC. *The Accounting Historians Journal, 14*(2), 71–91.

McGarrigle, J., & Donaldson, M. (1974). Conservation accidents. *Cognition, 3*(4), 341–350.

Mehler, J., & Bever, T. G. (1967). Cognitive capacity of very young children. *Science, 158* (3797), 141–142.

Mix, K. S. (1999). Similarity and numerical equivalence: Appearances count. *Cognitive Development, 14*(2), 269–297.

Muldoon, K., Lewis, C., & Freeman, N. H. (2003). Putting counting to work: Preschoolers' understanding of cardinal extension. *International Journal of Educational Research, 39* (7), 695–718.

Negen, J., & Sarnecka, B. W. (2009). *Young children's number-word knowledge predicts their performance on a nonlinguistic number task*.

Negen, J., & Sarnecka, B. W. (2015). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *British Journal of Developmental Psychology, 33*(1), 92–105.

Ng, R., Heyman, G. D., & Barner, D. (2011). Collaboration promotes proportional reasoning about resource distribution in young children. *Developmental Psychology, 47*(5), 1230.

Owens, K., & Lean, G. (2018). Body-part tally systems. In *History of number* (pp. 61–72). Cham: Springer.

Piaget, J. (1953). *The origin of intelligence in the child*. New Fetter Lane, New York: Routledge & Kegan Paul.

Piaget, J. (1965). *The Child's conception of number*. New York: W. W. Norton & Company, Inc.

Piantadosi, S. T., Jara-Ettinger, J., & Gibson, E. (2014). Children's learning of number words in an indigenous farming-foraging group. *Developmental Science, 17*(4), 553–563.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*(5695), 499–503.

Rakoczy, H., Kaufmann, M., & Lohse, K. (2016). Young children understand the normative force of standards of equal resource distribution. *Journal of Experimental Child Psychology, 150*, 396–403.

Rose, S. A., & Blank, M. (1974). The potency of context in children's cognition: An illustration through conservation. *Child Development*, 499–502.

Russac, R. J. (1978). The relation between two strategies of cardinal number: Correspondence and counting. *Child Development*, 728–735.

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

Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition, 92*(3), 329–352.

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.

Saxe, G. B. (1981). Body parts as numerals: A developmental analysis of numeration among the Oksapmin in Papua New Guinea. *Child Development*, 306–316.

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6*(3), 357–379.

Schmandt-Besserat, D. (1992). *Before writing, vol. I: From counting to cuneiform. 1*. University of Texas Press.

Schneider, R. M., Sullivan, J., Guo, K., & Barner, D. (2021). What counts? Sources of knowledge in children's acquisition of the successor function. *Child Development, 92* (4), e476–e492.

Schneider, R. M., Sullivan, J., Marušič, F., Biswas, P., Mišmaš, P., Plesničar, V., & Barner, D. (2020). Do children use language structure to discover the recursive rules of counting? *Cognitive Psychology, 117*, 101263.

Sheldon, K. M. (1999). Learning the lessons of tit-for-tat: Even competitors can get the message. *Journal of Personality and Social Psychology, 77*(6), 1245.

Shusterman, A., Cheunga, P., Taggartb, J., Bassc, I., Berkowitzd, T., Leonarde, J. A., & Schwartzf, A. (2017). Conceptual correlates of counting: Children's spontaneous matching and tracking of large sets reflects their knowledge of the cardinal principle. *Journal of Numerical Cognition, 3*(1), 1–30.

Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLoS One, 11*(4), Article e0153072.

Siegler, R. S. (1991). *Children's thinking*. Prentice-Hall, Inc.

Silk, J. B. (2003). Cooperation without counting. *Genetic and cultural evolution of cooperation*, 37–54.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science, 23*(2), 196–204.

Stephens, M., Landeros, K., Perkins, R., & Tang, J. H. (2016). Highlights from TIMSS and TIMSS advanced 2015: mathematics and science achievement of US students in grades 4 and 8 and in advanced courses at the end of high School in an International Context. In *NCES 2017–002*. National Center for Education Statistics.

Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology, 16*(2), 145–176.

Wassmann, J., & Dasen, P. R. (1994). Yupno number system and counting. *Journal of Cross-Cultural Psychology, 25*(1), 78–94.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*(2), 130–137.

Whorf, B. L. (1956). *Language, thought, and reality: selected writings of Benjamin Lee Whorf. (Edited by John B. Carroll.)*. Technology Press of MIT.

Wright, C. (1999). Is Hume's principle analytic? *Notre Dame Journal of Formal Logic, 40*(1), 6–30.

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*(2), 155–193.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology, 24*(2), 220–251.

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8*(1), 88–101.

Zeigenfuse, M. D., & Lee, M. D. (2010). A general latent assignment approach for modeling psychological contaminants. *Journal of Mathematical Psychology, 54*(4), 352–362.