The Development of Cardinal Extension: From Counting to Exact Equality

Anonymized Author 1, Anonymized Author 2, and & Anonymized Author 3

[anonymized]

Author Note

Abstract

Numerate adults know that when two sets are equal they should be labeled by the same number word. We explored the development of this principle - sometimes called "cardinal extension" - and how it relates to children's other numerical abilities. Experiment 1 revealed that 2- to 5-year-old children who could accurately count large sets often inferred that two equal sets should be labeled with the same number word, unlike children who could not accurately count large sets. However, not all counters made this inference, suggesting that learning to construct and label large sets may be a necessary but not sufficient step in learning how numbers represent exact quantities. Experiment 2 found that children who extended labels to equal sets were not actually sensitive to exact equality, and that they often assigned two sets the same label when they were approximately equal, but differed by just 1 item (violating one-to-one correspondence). These results suggest a gradual, stage-like, process in which children learn to accurately count, learn to extend labels to perceptually similar sets, and then eventually restrict cardinal extension to sets that are exactly equal.

Public significance statement: This study suggests that learning to count and construct sets of large quantities may be a necessary but not sufficient step in learning how numbers represent exact quantities. While children who can accurately construct large sets show greater understanding that only sets of the same quantity can be labeled by the same number word, they also often assign the same number word to sets that are approximately equal.

Keywords: number words, number concepts, exact equality, cardinal extension, language development, preschool children

Word count: 12904

The Development of Cardinal Extension: From Counting to Exact Equality

Imagine attending a lecture in an auditorium filled to capacity, where every seat is occupied by an eager student. After the lecture ends and everyone has left, you want to find out how many students attended. Is there a way to know? As numerate adults, we know that we can count the number of seats and then extend the resulting cardinal label (e.g., "eighty") to the number of attendees. Understanding this principle, sometimes called "cardinal extension" (Muldoon et al., 2003), involves two distinct abilities. First, it requires the non-linguistic ability to recognize that two sets have the same number of items if and only if their members can be placed in one-to-one correspondence, sometimes called "Hume's Principle" (Boolos, 1986; Decock, 2008; Frege, 1880, 1884; Heck, 2000). Second, it requires understanding that a particular number word can be applied to two different sets if and only if they have an equal number of items. Therefore, cardinal extension integrates both non-linguistic reasoning about exact equality and linguistic knowledge of how number words encode number.

How do children acquire this knowledge? According to one view, once children learn their first 1-2 number words, they quickly infer that all number words denote unique, exact, numerosities. Previous studies establish that, beginning sometime after the age of 2, children learn the meanings of the words 'one', 'two', and 'three' one at a time over the course of 1-2 years, during which time they are known as "subset-knowers" (since they know the meanings of only a subset of numbers, e.g., Wynn, 1990, 1992). According to Sarnecka and Gelman (2004), knowledge of these first few number word meanings might be sufficient to support an inference that all number words denote unique, exact, numerosities, and therefore that sets which differ in number should receive different numerical labels, while equinumerous sets should receive the same labels. Further, they suggest that this should extend even to words for which these children don't yet have meanings. In support of this, they presented children who only knew words up to 'three' with a large set - e.g., "Look, there are five frogs", and found that when an item was added to the set and

children were asked "Now are there five or six?" they correctly chose the alternative label (similar results were also found for subtraction trials). Children also correctly reasoned that the number word should not change when a transformation does not change the quantity (e.g., when a set is simply rearranged). Other studies, however, have questioned these results, finding that children fail with highly similar tasks (Brooks et al., 2013; Condry & Spelke, 2008; Sarnecka & Wright, 2013). For example, when subset-knowers are shown two sets that appear equal in number, are told that one of them contains, e.g., 'six' things, and then see 1 item added to that set, they choose randomly between the two sets when asked to find the one with 'six' (e.g., Brooks et al., 2013; Condry & Spelke, 2008). Given such findings, some have argued that apparent successes may not demonstrate an understanding of how number words encode exact numerosities, but may instead reflect the use of a pragmatic inference compatible with the principle of contrast - i.e., that any referent that undergoes change along a meaning-relevant dimension should receive a new label (Brooks et al., 2013; see Izard et al., 2014 for a discussion).

According to an alternative hypothesis, children may not understand that number words denote unique, exact, numerosities until they understand how counting works, and are able to provide cardinal labels for any set they can count (Carey, 2004, 2009; Condry & Spelke, 2008; Sarnecka & Wright, 2013). Sometime after they learn the meanings of 'one', 'two', and 'three', children appear to discover that counting can be used to both construct large sets and label their cardinalities, at which point they are sometimes called "Cardinal Principle" knowers (or CP-knowers). For example, in Wynn's Give-a-Number task, when CP-knowers are given a pile of objects, like fish, and asked to give "five fish", they are typically able to count and give the requested number, unlike subset-knowers. According to some proposals, it is this ability to accurately count and construct large sets that establishes children's understanding of how number words represent cardinality, since mastery of counting requires establishing one-to-one correspondence between labels and counted objects, which in turn guarantees that any two counts to a particular number - like

five - will result in the same number (Carey, 2004, 2009). As evidence for this, previous studies have found that CP-knowers outperform subset-knowers in tests of cardinal extension. For example, when shown two sets that appear equal in number, CP-knowers often correctly judge that if a number word, e.g., 'five,' applies to one set, it should also apply to the other, while subset-knowers fail at the same task (Sarnecka & Gelman, 2004; Sarnecka & Wright, 2013). Similarly, when shown a set of items that is labeled with a number word, e.g., 'four turtles', and asked to distinguish between two sets to find a set with the same number word label, CP-knowers, but not subset-knowers, succeed (Slusser & Sarnecka, 2011). These findings suggest that learning to accurately count and construct large numbers is related to a qualitative change in children's understanding of how number words encode exact equality.

While CP-knowers are able to accurately count and label large sets, some have suggested that these abilities may actually reflect rote procedures (Davidson et al., 2012), and that many CP-knowers still don't understand that every number word denotes a unique cardinality, or that equinumerous sets should receive the same cardinal label. Instead, they have argued that this knowledge may emerge after children become CP-knowers (for discussion see Schneider et al., 2022). As evidence for this, although CP-knowers outperform subset-knowers on tests of cardinal extension, they rarely perform at ceiling, failing from 15% to 40% of the trials depending of the task (Sarnecka & Gelman, 2004; Sarnecka & Wright, 2013; Slusser & Sarnecka, 2011). In fact, multiple past studies report variability on tests of cardinal extension until up to 5 years of age. However, in many cases these studies have not classified children's knower levels, leaving open the role of counting knowledge in children's performance. For example, in one study, Muldoon et al. (2005) showed 3- and 4-year-olds a set of lions who rode their bicycles to school and then disappeared inside it. Children were first asked how many bicycles there were – which they were able to count. They were then asked how many lions there were, to probe whether they can infer the number of a hidden set (the lions) from the number of a visible set in

one-to-one correspondence (the bicycles), without having to count the lions. When the number of items was large (5 to 6 items), 3-year-olds extended the label of the visible set to the hidden set about 56% of the time, while 4-year-olds succeeded about 86% of the time (see also Sophian et al., 1995 for similar results with 5-year-olds[1]). In another study, Frydman and Bryant (1988) found that when two equal sets were generated through sharing, fewer than half of the 4-year-olds tested were able to infer the cardinality of one set by counting the other set. Similarly, Muldoon et al. (2003) asked 3.5 to 5-year-olds to construct a set equivalent to a set they had already counted, and to judge whether a new set that was one-too-many or one-too-few had the same number of items as the counted set. They found that roughly half of children succeeded in these tasks (ranging from 43% - 65% depending on the task).

Adding to doubt that children master cardinal extension when they become CP-knowers is the fact that such children often fail on tasks that test non-verbal understanding of one-to-one correspondence. First, classic Piagetian studies of conservation show some degree of difficulty in children well past the age of 5, though subsequent studies argued that such failures might be explained by alternative factors. For example, similar to the Brooks et al. (2013)'s pragmatic account of Sarnecka and Gelman (2004), some have argued that repeatedly asking children the same question about a referent set ("Which line has more?") may lead them to infer that a new response is required, and that they should therefore base their judgment on the property that changed (e.g., line length), rather than on number (McGarrigle & Donaldson, 1974; Mehler & Bever, 1967; Rose & Blank, 1974). However, other studies have nevertheless found continuing failures in children as old as 6 on tasks that reduce these demands (Piaget, 1965; Russac, 1978; Schneider et al., 2022). For example, using tasks similar to those used in tests of non-numerate groups like the

—————

[1] Children were more successful at this task when the sets were small (2-3 items) – 3-year-olds succeeded 87% of the time and 4-year-olds 91% of the time. However, this might be due to children subitizing the sets while they were visible, memorizing the number of items of the hidden sets, and reporting it once asked.

Piraha and Mundurucu (Gordon, 2004; Pica et al., 2004), Schneider et al. (2022) presented children with a row of fish and asked them to match that row. Although both subset-knowers and CP-knowers performed better than chance when tested with small sets (e.g., 3), only CP-knowers were better than chance for large sets, but many failed to provide exact matches, instead approximating. This was true even when children were explicitly asked to match the number, and when the task was presented as a sharing game where their goal was to ensure fairness in how fish were distributed to two penguins. Although some studies have shown precursors to one-to-one correspondence in younger children (Izard et al., 2014), here too success is limited to cases in which children can make judgments based on small sets[2].

To summarize, previous studies have debated when and how children acquire the two core components of cardinal extension - i.e., that two equal sets should receive the same label, and that two sets are equal only if their elements stand in one-to-one correspondence. Some argue that this knowledge emerges after children learn just 1-2 small number words, while others argue that it develops when children become CP-knowers, or even sometime after this. Critically, however, previous studies are limited in various ways that make it difficult to know how such knowledge actually arises. First, some studies test cardinal extension and find variability between the ages of 3 and 5, but do not classify children into knower levels, making it difficult to know how their performance relates to knowledge of counting (e.g., Frydman & Bryant, 1988; Muldoon et al., 2003, 2005; Sophian et al., 1995). Second, other studies have classified children as subset-knowers or CP-knowers and tested differences between these groups, but have not analyzed sources of

---

[2] In Izard et al. (2014), children continued searching when 3 puppets were presented on 3 branches, then were hidden in a box, and 2 were removed from the box. They also succeeded when 6 puppets were presented on 6 branches, and 5 came out. However they failed when 6 puppets were presented on 11 branches and 5 came out, suggesting that their success in the 5/6 case was driven by their ability to track a small number of empty branches, not their ability to track 5 objects.

individual differences between children within these groups (e.g., Sarnecka & Gelman, 2004; Sarnecka & Wright, 2013; Slusser & Sarnecka, 2011). Third, as we explain below, while cardinal extension requires two components (recognizing equinumerous sets and understanding that such sets share the same number label), previous studies typically test only one or the other, but rarely test both.

Crucially, most studies that attempt to test both children's reasoning about equinumerosity and how they use this knowledge to extend number labels (e.g., Sarnecka & Gelman, 2004) do not differentiate between reasoning about exact vs. approximate representations of number[3]. This is important, because children might represent two sets as "the same" using approximate representations of number, even if the sets violate one-to-one correspondence and differ by 1 item[4]. A large literature indicates that human perception of large sets is supported by representations in the Approximate Number System (ANS, Cordes et al., 2001; Dehaene, 1997, 2009; Feigenson et al., 2004; Whalen et al., 1999) such that sets that differ by large ratios - e.g., 2:1, 2:3 - are easy to discriminate, while sets with smaller ratios often are not. Although the acuity of this system improves with age (Halberda & Feigenson, 2008; Huntley-Fenner & Cannon, 2000), even adults often fail to reliably discriminate large sets that have a ratio less than 7:8. Meanwhile, accurate and adult-like set comparison for sets beyond the subitizable range with perceptually non-discriminable ratios requires precise representations of numerosities (see Feigenson et al., 2004 for a discussion). Consequently, in order to be sure that children extend number

---

[3] One study, Sophian et al. (1995), did include trials that could differentiate between an exact equality vs. approximate quantity strategy, but did not report separate trial performance.

[4] As already noted, one exception to this is Sarnecka and Gelman (2004), which found that subset-knowers who extend number words to two equal sets subsequently refuse to do so after an object is added to one of them. However, in addition to conflicting with results from other tasks, this study did not test whether children's change in judgment was due to a violation of one-to-one, or the more general pragmatic heuristics that "a change in number should reflect a change in number word", which is neutral with respect to whether - and how - children represent the sets as equal (Brooks et al., 2013).

words to sets that are exactly equal, it's critical to ensure that their extension is not based on approximate matches between sets. While some studies have attempted to do this by showing children two equal sets and then adding or subtracting 1 item from one set (e.g., Sarnecka & Gelman, 2004), in such studies children might assume that any change in number requires a change in label without knowing how to assess equality, and also without thinking that this principle of contrast is specific to number (see, e.g., Brooks et al., 2013, for evidence that children make identical patterns of inference for changes in an object's part-whole structure).

The present studies aimed to address the limitations of past reports. Adapting the animal-bicycle task from Muldoon et al. (2005) in Experiment 1, we presented subset-knowers and CP-knowers with two sets of animals in unequal numbers - 5 bunnies and 7 lions. Animals in each group carried items (e.g., the bunnies had blue hats and the lions had red hats), such that the number of items was exactly equal to the number of animals in each group. The animal sets were then hidden, and children were prompted to infer the number of one of the two hidden animal sets (e.g., the number of bunnies). We reasoned that if children believe that two equal sets deserve the same numerical label, they should count the correct visible objects to infer the number of hidden objects (e.g., counting the bunnies' hats to infer the number of bunnies), whereas if they do not understand this they should count the set corresponding to the second set of animals. In Experiment 2, we probed the nature of the numerical representations used by children who appeared to succeed at cardinal extension, by asking whether they reasoned about exact equality or approximate magnitudes. Finally, while the main focus of these studies was to explore differences in performance between CP- and subset-knowers and the representations children use (e.g., one-to-one correspondence vs. approximation), we also asked asked whether performance was related to individual differences in rote counting ability, which previous studies find is related to both early number knowledge (Chu et al., 2020; Schneider et al., 2021) and later mathematics achievement (Aunola et al., 2004;

Koponen et al., 2016; Zhang et al., 2014).

Experiment 1

Method

All materials, data, and analysis scripts for both Experiments 1 and 2 are available at https://osf.io/eswa4/?view_only=044c906224724abeaf448790fa909b07. Data were analyzed using R (Version 4.1.3, R Core Team, 2022) and the R-packages *lme4* (Bates et al., 2015), *car* (Fox & Weisberg, 2019), *emmeans* (Lenth, 2023), *ggplot2* (Wickham, 2016), *cowplot* (Wilke, 2024), *tidyverse* (Wickham et al., 2019), and *papaja* (Aust & Barth, 2022). The design and analysis of both experiments were preregistered, and we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. A preregistration for Experiment 1 is available at https://osf.io/3v2cn/?view_only=e24cf57016474841822b81a61f95cab2.

Participants

Eighty-four[5] children were recruited from preschools in the US and Canada, and a children's museum in the US. All participants spoke English as a primary language. Based on preregistered criteria, we excluded two participants who did not provide a response for more than one trial of the Cardinal Extension task, and two trials due to experimenter error. Our final sample included 82 children, with 38 subset-knowers (23F, 15M; $M_{age} = $ 3.55 [2.13; 5.20]; $SD_{age} = 0.76$) and 44 CP-knowers (25F, 19M; $M_{age} = 4.63$ [3.08; 5.95]; $SD_{age} = 0.65$). The sample size was determined with an *a priori* power analysis on G*Power version 3.1.9.6 (Faul et al., 2009), and is adequate to achieve 80% power for detecting a medium effect, at a significance criterion of $\alpha = 0.05$ for a linear regression with 3 predictors (Age, Knower Level, and Set Size). This study was approved by [anonymized].

─────

[5] We recruited four more CP-knowers than preregistered due to an initial knower-level coding error.

Material & Procedure

Give-N.   Participants were given a titrated Give-N task (following the procedure in Wynn, 1992) to classify them as either subset-knowers or CP-knowers. Participants were shown a box with fish and a plate, and were asked to "put N fish on the plate." All participants started with 'five', and received increasingly larger numbers (N+1, up to 'six') if they succeeded and smaller numbers (N-1) if they failed. If the child gave the wrong number, they were prompted once to "count to make sure" and were allowed to fix their response. We recorded the largest number for which the child gave correct responses on at least two out of three trials. We excluded children who could not construct sets of 'one' on at least two out of three trials. Children who succeeded on sets of 'five' or 'six' on at least two out of three trials were designated as CP-knowers, and those who only succeeded on smaller numbers (e.g., 'one', 'two', 'three', 'four') were designated as subset-knowers.

Cardinal Extension.   Materials were prepared and presented as a slidedeck with recorded audio descriptions. Participants were first introduced to the animals used in the task (lions and bunnies). They then saw two familiarization trials with three animals, in which each animal was associated with one item (e.g., each bunny had a bike). They were then asked to report the number of animals by pointing to the screen and counting to familiarize them with the expectations for critical trials.

In each critical trial (Figure 1), participants saw two unequal groups of animals (e.g., 5 bunnies and 7 lions), each on one side of the screen. Of these two groups, one served as the target set (whose number was queried by the experimenter), though participants were not informed which set was the target until the end of the trial. Initially, each animal was physically connected to one item (e.g., each animal wore a hat), such that animals and items stood in one-to-one correspondence (e.g., 5 bunnies – 5 blue hats vs. 7 lions – 7 red hats). Each group of animals stood with their items in a vertical line with equal spacing to facilitate element tracking and counting. The animals then put down their

items and disappeared into a school. Participants were then asked for the number of one set of animals (e.g., "How many bunnies are in the school?"). We reasoned that if children believe that two equal sets deserve the same numerical label, they should count the correct visible objects to infer the number of hidden objects (e.g., counting the bunnies' hats to infer the number of bunnies), whereas if they do not understand this they should count the set corresponding to the second set of animals. If participants gave a response without counting, they were encouraged to 'point and count' (the experimenter never specified what they should count)[6]. If counting resulted in a different response, we analyzed the final count. Participants were allowed one opportunity to fix a wrong response.

Participants saw six critical trials in total: three small-set trials where sets were less than 4 (1:2, 1:3, 2:3), and three large-set trials where sets were greater than 4 (5:7, 5:8, 6:7). Lions and bunnies were used for all trials, but they carried different kinds of items for each trial. Participants saw one out of six pseudo-randomized trial orders. Trials were counterbalanced for the target animals, side of set appearance, and which animal set is larger.

Highest Count. This task was included as a general proxy of counting experience, to allow us to differentiate between CP-knowers with different degrees of counting training. Participants were asked to "count as high as you can," beginning from one. The experimenter prompted them with 'one' (but no other number) if they failed to respond, and they were prompted once after they stopped to keep counting. Participants were prompted to stop once they reached 100 or after they had made a mistake, whichever came first. We recorded the highest number they reached without errors.
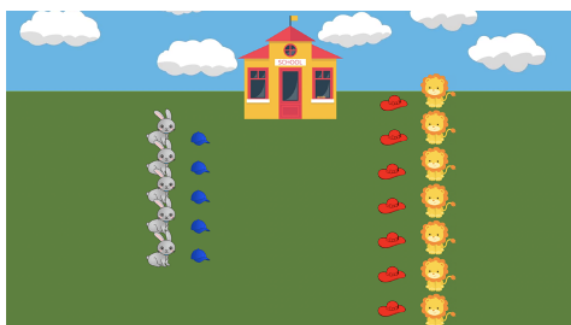
----

[6] Out of 490 total trials, there were 25 trials where CP-knowers (n=8) responded with a number but did not count (small sets: 14; large sets: 11), and 83 trials where subset-knowers (n=17) responded with a number but did not count (small sets: 43; large sets: 40)
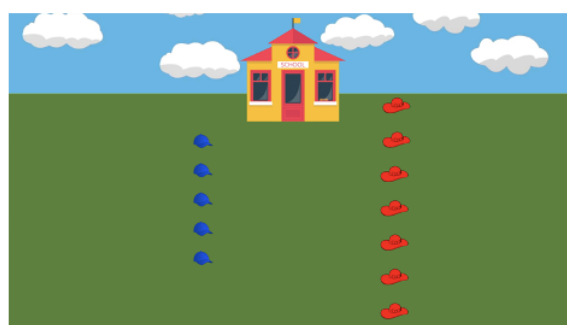
"Look! The animals are going to school! They're bringing their hats with them.
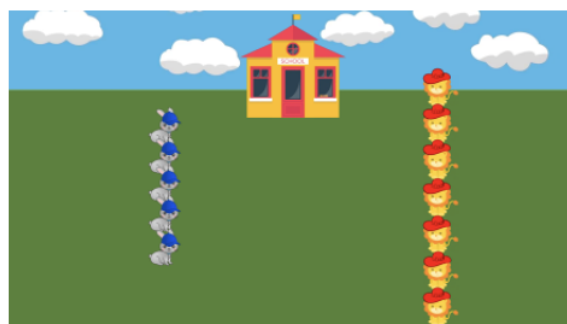
"Oh no! They can't bring their hats inside."

"Look! They're leaving their hats outside. Now they can enter the classroom."

"How many bunnies are in the school?"

Figure 1

Schematic of Experiment 1. Slides are presented in order, from left to right and top to bottom. The corresponding audio description is under each slide.

Results

Our primary question was whether CP-knowers were more likely to succeed atprovide correct responses on the Cardinal Extension task compared to subset-knowers across both small and large sets. To test this, we asked whether children knew that to find the number of animals in the school (e.g., bunnies), they should count the items that these animals left outside the school (and which were therefore equal in number) - what we describe below as Correct Set Choice. To identify children who selected the correct set, we coded two kinds of response as correct. First, we credited the child with a correct response if they counted the correct set. We counted these responses as correct even if the

attempted count resulted in an incorrect response in order to recognize that some children might have an understanding that two equal sets deserve the same label but nevertheless make procedural counting errors (e.g., in the case of some CP-knowers). Also, this allowed for the possibility, proposed by Sarnecka and Gelman (2004), that a subset-knower might know that two large sets should receive the same label if they are equal, but might not be able to say how many items the sets actually contain (because they can't yet count). Second, we also credited the child with a correct response if, despite not attempting to count, they nevertheless provided a correct numerical response. Here, we reasoned that such a response is most likely if the child has successfully identified the correct item set from which to infer the number of animals.

We also included a secondary analysis to explore how often children provided a correct numerical response overall. Although children might understand cardinal extension while still making counting errors, it is nevertheless interesting to know how often children were able to provide the correct numerical response and were therefore fully adult-like (e.g., by choosing the right side, and accurately counting the available items). For this measure, we credited children if they provided the correct numerical response whether or not they pointed to the screen as part of obtaining their answer. However, if children clearly counted the wrong set (or both sets) and arrived at the correct answer through a counting error, they were not credited with a correct response. We will refer to this measure as 'Correct Numerical Response' in the following analysis.

First, we consider the 'Correct Set Choice' measure. Overall, CP-knowers were more likely than subset-knowers to choose the correct set to count for both small and large sets (see Table 1 for statistical results). Two-tailed one-sample t-tests found that only CP-knowers ($M = 0.88$, $SD = 0.32$) performed better than chance at identifying the correct item set to infer the number of hidden animals ($t(43) = 12.13$, $p < .001$), and that they did so for both small and large sets when examined separately ($ps < .001$). Meanwhile, subset-knowers ($M = 0.49$, $SD = 0.50$) performed at chance ($t(37) = -0.23$, $p = .817$), and

failed to identify the correct set for both small and large sets ($ps > .05$). In a *post hoc* analysis suggested by an anonymous review, we found that subset-knowers also performed at chance ($ps > .05$) when the analysis was restricted to sets within their knower level (e.g., 2-knower only on trials involving sets of up to 2 items). Although CP-knowers performed above chance as a group, they nevertheless exhibited substantial variability (Figure 2A).
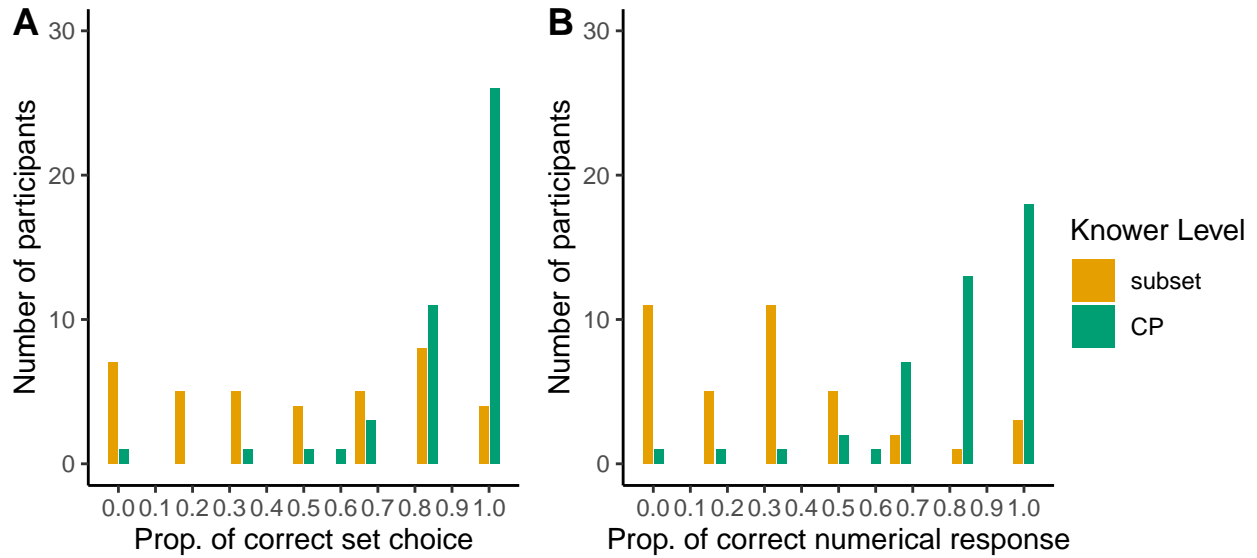
Figure 2

A) Number of participants against proportion of correct set choice. B) Number of participants against proportion of correct numerical response.

In order to further analyze the effect of CP knowledge on cardinal extension performance as measured by Correct Set Choice, we constructed a generalized linear mixed-effects models (GLMM) with the formula Correct Set Choice (0/1) ~ Age (z-scored) + Set Size (Small / Large) + (1 | Participant) + (1 | Trial), and tested whether adding children's Knower Level (CP / Subset) and Knower Level * Set Size interaction improved the fit of the model. All models included by-subject and by-item random intercepts. Our base model showed a significant effect of age ($\beta = 1.67$, 95% CI = [1.14, 2.19], $p < .001$), with older children performing better than younger children. However, set size (small vs large sets) did not further explain variation in performance ($p = .504$). Exploratory models

Table 1

Results of t-tests comparing cardinal extension performance between subset- and CP-knowers in Experiment 1.

Cardinal extension performance	subset		CP				
	M	SD	M	SD	t	df	p
Correct Set Choice							
Small sets	0.51	0.50	0.89	0.50	5.11	57.50	< .001
Large sets	0.46	0.50	0.88	0.50	5.18	60.45	< .001
Total	0.49	0.50	0.88	0.50	6.07	58.32	< .001
Correct Numerical Response							
Small sets	0.48	0.50	0.89	0.50	5.47	57.53	< .001
Large sets	0.16	0.37	0.74	0.37	9.61	86.14	< .001
Total	0.32	0.47	0.81	0.47	8.17	69.58	< .001

Note. Performance for each trial is coded as follows. For 'Correct Set Choice', 1: attempted to count the correct set (regardless of correct or incorrect numerical response), or provided the correct numerical response without counting, 0: counted the wrong set or both sets, or provided the wrong numerical response without counting. For 'Correct Numerical Response,' 1: provided the correct numerical response (by counting the correct set or without counting at all), 0: provided the wrong numerical response, or counted the wrong set or both sets (even if the right numerical response was reached through a counting error). Results were qualitatively the same when analysis was limited to sets within individual subset-knowers' knower level (in this case, the 'Large sets' comparison was not conducted).
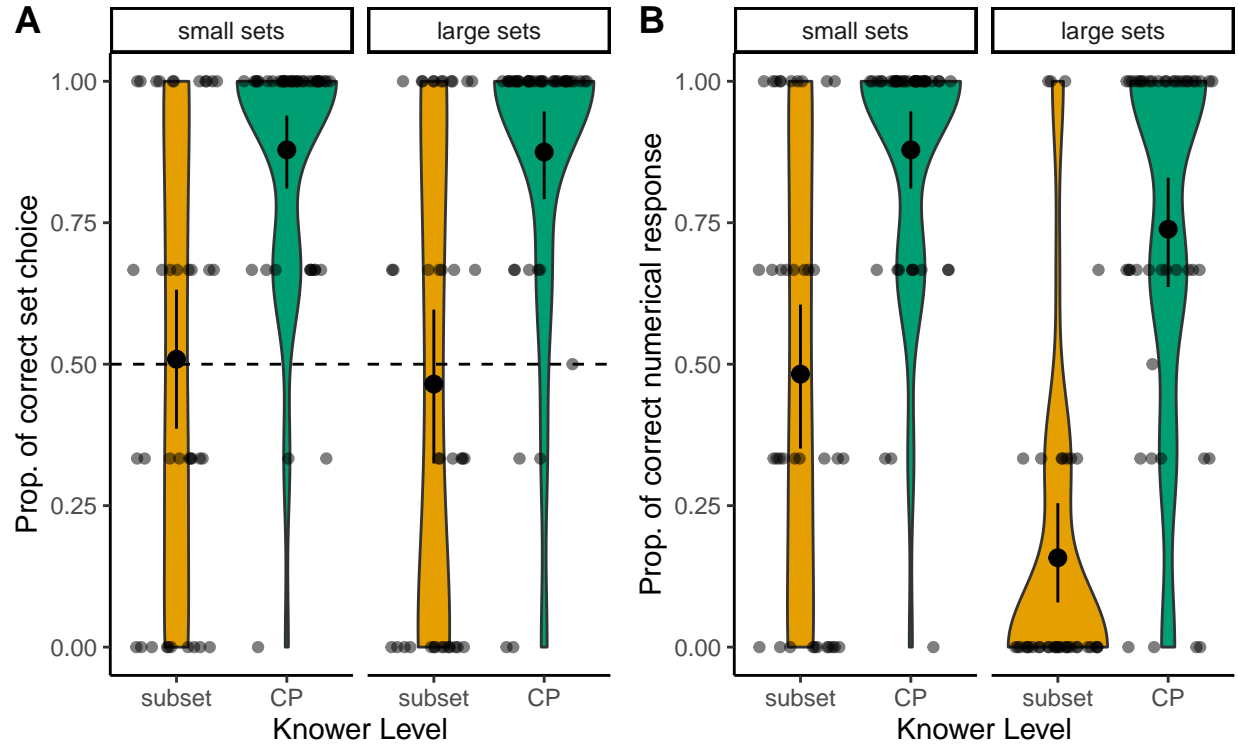
Figure 3

Each dot represents a participant. The width of the shaded area of violin plots represents the
proportion of the data located there. A) Proportion of correct item set choice by participant
against knower level. Horizontal dashed line indicates performance at chance = 0.50. B)
Proportion of correct numerical response by participant against knower level.

analyzing subset-knowers and CP-knowers separately showed that this age effect was
driven by only subset-knowers ($\beta = 1.56$, 95% CI = [0.85, 2.26], $p < .001$). Age did not
predict performance in the CP-knower group ($p = .210$). Adding Knower Level as a
predictor significantly improved the fit of the model ($\chi^2(1) = 6.98$, $p = .008$, $\Delta AICc =$
-4.93, $\Delta BIC = -0.79$), and revealed that CP-knowers were significantly more likely than
subset-knowers to select the correct set, even when controlling for age ($\beta_{CP} = 1.43$, 95% CI
= [0.37, 2.50], $p = .008$). This model also showed a significant effect of age ($\beta = 1.18$, 95%
CI = [0.61, 1.75], $p < .001$) but not of set size ($p = .501$)[7]. These main effects are

───────

[7] As suggested by an anonymous reviewer, we conducted a *post hoc* analysis with the same models, but

illustrated in Figure 3A. Adding Knower Level * Set Size interaction did not further improve the model ($\chi^2(1) = 0.19$, $p = .663$, $\Delta AICc = 1.87$, $\Delta BIC = 6.00$).

We next asked how often participants gave the correct numerical response. Here we expected subset-knowers would fail for large sets (since they generally can't count accurately), and to sometimes succeed for smaller ones, and for CP-knowers to perform better in both cases. Compatible with these expectations, a Welch two-tailed two-sample t-test found that CP-knowers provided the correct number of the hidden animal set more than subset-knowers in both small and large sets (Table 1). We again observed variability in CP-knowers' performance (Figure 2B). Twenty-five of the 44 CP-knowers (56.82%) gave a numerical response that was incorrect on at least one trial, resulting in a total of 46 out of 262 trials with an incorrect response. Some of these incorrect responses were due to procedural counting errors: on 18 of these 46 trials (13 CP-knowers) they counted the correct item set but made a mistake in their count (even though we gave them one chance to fix their wrong count). On 13 trials (11 CP-knowers), they counted the wrong item set, or both item sets, which pointed to a lack of cardinal extension understanding. Finally, on the remaining 15 trials (6 CP-knowers), CP-knowers provided the wrong numerical response without pointing to items in either visible set. This last behavior is compatible either with overconfidence in their ability to accurately estimate a set size, an attempt to count without pointing, or a failure to understand that the target number could be obtained by counting one of the visible sets.

———

removed any trials with sets beyond a particular subset-knower's knower level. Results were qualitatively similar. One difference was that Knower Level no longer surpassed the threshold of $p < .05$ ($p = .056$), and adding Knower Level similarly no longer improved the base mode ($p = .063$). Considering that in this analysis, CP-knowers still performed significantly better than subset-knowers (see t-tests reported above), we suspect this discrepancy in the regression models is due to the great difference in number of data points from CP-knowers and subset-knowers, and does not indicate a lack of difference between CP- and subset-knowers. Additionally, in this case we did not conduct any analysis involving the Knower Level * Set Size interaction, because the subset-knowers would only contribute small trials and no large trials.

To further explore the effect of CP-knower status on Correct Numerical Response, we constructed another set of GLMMs with the same fixed and random effect structure as the first analytical approach to predict children's ability to infer the correct number of animals. Specifically, our base model had the formula Correct Numerical Response (0/1) ~ Age (z-scored) + Set Size (Small / Large) + (1 | Participant) + (1 | Trial), and we tested whether adding children's Knower Level (CP / Subset) and Knower Level * Set Size interaction improved the fit of the model. The model that best explained the data included Age, Knower Level, Set Size, and Knower Level * Set Size interaction as predictors (compared against base model with only Age and Set Size: $\chi^2(2) = 20.34$, $p < .001$, $\Delta AICc$ = -16.23, ΔBIC = -7.95; compared against model with Age, Set Size and Knower Level: $\chi^2(1) = 4.77$, $p = .029$, $\Delta AICc$ = -2.71, ΔBIC = 1.43). CP-knowers were significantly better at determining the correct number of animals ($\beta_{CP} = 1.53$, 95% CI = [0.27, 2.78], $p = .017$), as were older children ($\beta = 1.15$, 95% CI = [0.55, 1.76], $p < .001$). We also found an effect of Set Size with children more likely to provide the correct numerical response in trials with small sets compared to those with large sets ($\beta_{large} =$ -2.59, 95% CI = [-3.66, -1.52], $p < .001$)[8]. Additionally, the Knower Level * Set Size interaction effect was significant, where subset-knowers showed a larger difference in performance between small and large trials compared to CP-knowers ($\beta_{CP*large} = 1.28$, 95% CI = [0.11, 2.45], $p = .032$). These effects are illustrated in Figure 3B.

As an exploratory analysis, we investigated whether children's Highest Count explained further variability in our data, by adding it as a fixed effect to the best-performing models from the two analysis approaches above. Seventy-eightchildren (out of 82) provided a Highest Count. As expected, CP-knowers ($M_{HC} = 22.72$ [2; 100];

─────────

[8] As suggested by an anonymous reviewer, we conducted a *post hoc* analysis with the same models, but removed any trials with sets beyond a particular subset-knower's knower level. Results were qualitatively similar. In this case we did not conduct any analysis involving the Knower Level * Set Size interaction, because the subset-knowers would only contribute small trials and no large trials.

$SD_{HC} = 22.54$) generally performed better in this task compared to subset-knowers (M_{HC} = 7.24 [1; 14]; $SD_{HC} = 3.84$). After excluding the four children who did not provide a Highest Count (all subset-knowers), we found that adding Highest Count did not explain additional variability for either Correct Set ($\chi^2(1) = 0.06$, $p = .805$, $\Delta AICc = 2.00$, ΔBIC = 6.08) or Correct Numerical Response ($\chi^2(1) = 0.30$, $p = .584$, $\Delta AICc = 1.77$, $\Delta BIC =$ 5.84). When we limited this analysis to only CP-knowers, Highest Count also did not explain additional variation in the observed performance (ps > .05). In all models, Highest Count showed low multicollinearity with other predictors (Variance Inflation Factor < 2, following the threshold established by James et al., 2013)[9].

Discussion

Experiment 1 found three main results. First, CP-knowers often chose the correct set to determine how many animals were hidden, compatible with some preliminary understanding of cardinal extension, while subset-knowers chose at chance. Second, although many CP-knowers succeeded at this task, there was nevertheless a fair amount of variability in their performance. Third, we found that many children failed to provide a correct numerical label, which could reflect a failure to understand cardinal extension, a lack of counting knowledge (e.g., in subset-knowers), procedural counting errors, or a decision to not attempt to count during the task. Together, these results suggest that children do not acquire an understanding of cardinal extension prior to the CP-stage, and therefore that knowledge of just 1-2 number words is not sufficient to support this ability. Similarly, due to the variability in performance even among CP-knowers, the data suggest that full mastery of cardinal extension likely emerges sometime after the CP transition, when children learn to accurately count large sets.

Amidst some variability in their performance, many CP-knowers nevertheless appeared to have a strong mastery of cardinal extension in Experiment 1, suggesting that

---

[9] We thank a reviewer for suggesting this *post hoc* analysis.

this knowledge might emerge with the CP transition. However, there are two reasons to worry that Experiment 1 might overestimate the number of children who grasp this principle. First, in Experiment 1 - as in studies that use a similar design such as Muldoon et al. (2005) and Sophian et al. (1995) - children may have deployed a simple heuristic to choose the appropriate set that did not rely on understanding of cardinal extension. In particular, because there were two sets of animals, each with their own items, when children were asked, e.g., how many bunnies were in the school, they may have chosen to count the bunnies' items based purely on the association between these animals and their items, but not based on their equinumerosity (i.e., if there were fewer items than bunnies they may have still counted the items, based on this association). Second, although animals and their target items in Experiment 1 were exactly equal in number, this does not guarantee that children actually used exact equality as the basis for their judgments. Instead, as noted in the Introduction, it is possible that they chose which set to count based on its approximate match to the hidden set. We addressed these questions in Experiment 2.

Experiment 2

Experiment 1 found that many CP-knowers infer that sets which appear equal in number can be labeled by the same number word. However, the study left open how children arrive at this conclusion. One possibility is that they select the correct set by noticing a one-to-one correspondence between items and animals, and reasoning that because the two sets are equal they deserve the same label (e.g., exactly 5 bunnies and exactly 5 hats). Alternatively, they might compare only the approximate quantity of these sets (e.g., approximately 5 bunnies and approximately 5 hats). Finally, another possibility is that they might succeed by merely noting the association between the correct item set and the target animal set (e.g., the bunnies appeared on the left, therefore choose the item set on the left; or the bunnies have blue hats, therefore choose the blue hats), without attending to cardinality at all.

To probe whether children use exact or approximate representations to support cardinal extension, and to eliminate the possibility of using identity associations between animals and items, in Experiment 2 we paired one animal set with two item sets in varying ratios. One item set exhibited one-to-one correspondence with the animal set, and thus could be counted to infer the number of animals once hidden. The second item set either differed from the animal set in a perceptually discriminable ratio (e.g., 5 hats : 10 bunnies), or was off by one in quantity and was not discriminable from the target set according to their ratio (e.g., 9 hats : 10 bunnies). This design prevented children from attending to the item/animal identity as a strategy to solve the task. Additionally, it differentiated two set comparison strategies. If CP-knowers succeed in cardinal extension through reasoning about one-to-one correspondence and exact set equality, they should succeed in both the perceptually discriminable (5:10) and off-by-one (9:10) cases. However, if they only attend to approximate quantities, they should succeed in trials with discriminable ratios, but not in the off-by-one trials.

Method

All materials, data, and analysis code for the experiment are available at https://osf.io/eswa4/?view_only=044c906224724abeaf448790fa909b07. A preregistration is available at https://osf.io/zrsw2/?view_only=74096634b5f346748a57275fb1d73d37.

Participants

Eighty children were recruited from preschools and a children's museum in the US. Given the failure of subset-knowers in Experiment 1, all participants were CP-knowers who spoke English as a primary language. We excluded two participants who missed more than one trial of the Cardinal Extension task based on preregistered criteria. We also excluded 18 trials where participants started counting the animals before the prompt, to ensure we only analyzed numerical responses from participants who attempted to make the inference after the animal set was hidden. Our final sample included 78 CP-knowers (43F, 35M;

$M_{age} = 4.68$ [3.29; 5.98]; $SD_{age} = 0.71$). The sample size was determined with an *a priori* power analysis on G*Power version 3.1.9.6 (Faul et al., 2009), and is adequate to achieve 80% power for detecting a medium effect, at a significance criterion of $\alpha = 0.05$ for a linear regression with 2 predictors (Age and Trial Type). This study was approved by [anonymized].

Material & Procedure

All materials, data, and analysis scripts for the experiment are available at https://osf.io/eswa4/?view_only=044c906224724abeaf448790fa909b07.

Give-N. Participants were given a Give-N task and a Highest Count task following the procedure from Experiment 1. In order to be confident that the children in this study were CP-knowers, we used a more conservative criterion for CP status than in Experiment 1, and classified children as CP-knowers only if they succeeded at constructing sets of "six" two out of three times on the Give-N task. Children who did not meet this criterion did not proceed to the Cardinal Extension task.

Cardinal Extension. Materials and procedure were similar to Experiment 1, with any differences noted. In the familiarization phase, participants saw an additional trial with three animals (in this study, only bunnies), but only two of them had items and one of them did not. The bunny missing an item was highlighted to the participant ("This bunny doesn't have a carrot"). This was done to highlight the violation of one-to-one correspondence, to be sure that any child who failed at the task did so despite noticing violations of one-to-one. Participants were also asked to point and count the number of bunnies for this trial.

In each critical trial (Figure 4), a set of bunnies appeared at the bottom of the screen with two sets of items. One set of items (the target set) was exactly equal to the number of bunnies, and the other set (the distractor set) had fewer items (e.g., 11 bunnies with 11 hats and 5 water bottles). The audio description highlighted the violation of

one-to-one correspondence between the distractor set and the set of bunnies (e.g., 'Some of the bunnies don't have their water bottles'). The experimenter also gestured to the bunny or bunnies missing any item. The bunnies then put down each item in each set one at a time (e.g., each bunny put down their hat, then each of the bunnies with backpacks put down their backpack), to further emphasize the one-to-one correspondence between the bunnies and the target set, and reiterate the mismatch with the distractor set. Like in Experiment 1, the bunnies then disappeared into a building, leaving their items behind. Participants were then asked for the number of bunnies, following the same procedure as described in Experiment 1. Here, like Experiment 1, sometimes participants gave a response without counting, despite being encouraged to 'point and count'[10].

Participants saw nine trials in total: three small-set trials where sets were less than 4 (1:2, 1:3, 2:3), and six large-set trials where sets were greater than 4. Large-set trials included three with discriminable ratios (Large-DR: 5:10, 5:11, 5:12) where the ratio between the bunnies and the distractor item set was at least 2:1, and three with non-discriminable ratios (Large-NR: 9:10, 10:11, 11:12) where the distractor set had one fewer item than the number of bunnies. Participants saw one of four pseudo-randomized trial orders and item pairings. The trials were partially-counterbalanced for order and location of item sets.

Results

Our primary questions were (1) whether children chose the correct set to infer the number of bunnies across trials of different set sizes and ratios, and (2) whether children who did this drew on one-to-one correspondence or approximate number representations as the basis for their judgments. As in Experiment 1, we first examined whether children

———

[10] Out of 684 total trials, there were 103 trials ($N=21$) where children responded with a number but did not count (small trials: 32 trials; large trials with perceptually discriminable ratios: 35 trials; large trials with off-by-one difference: 36 trials)
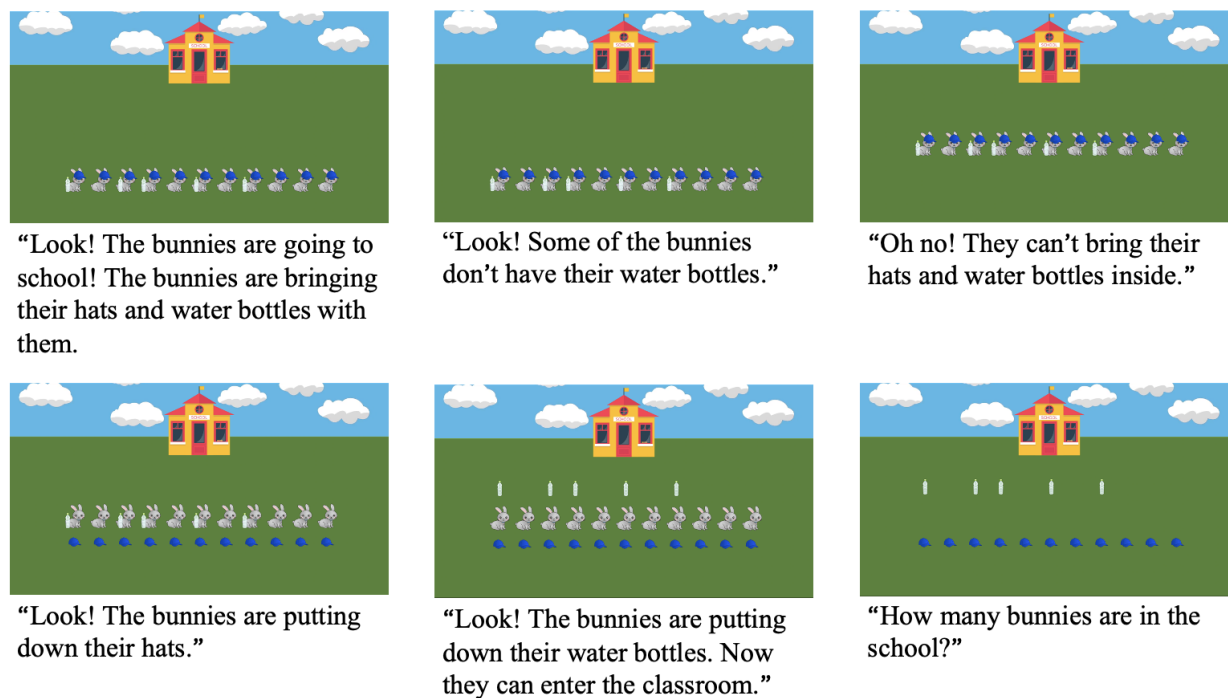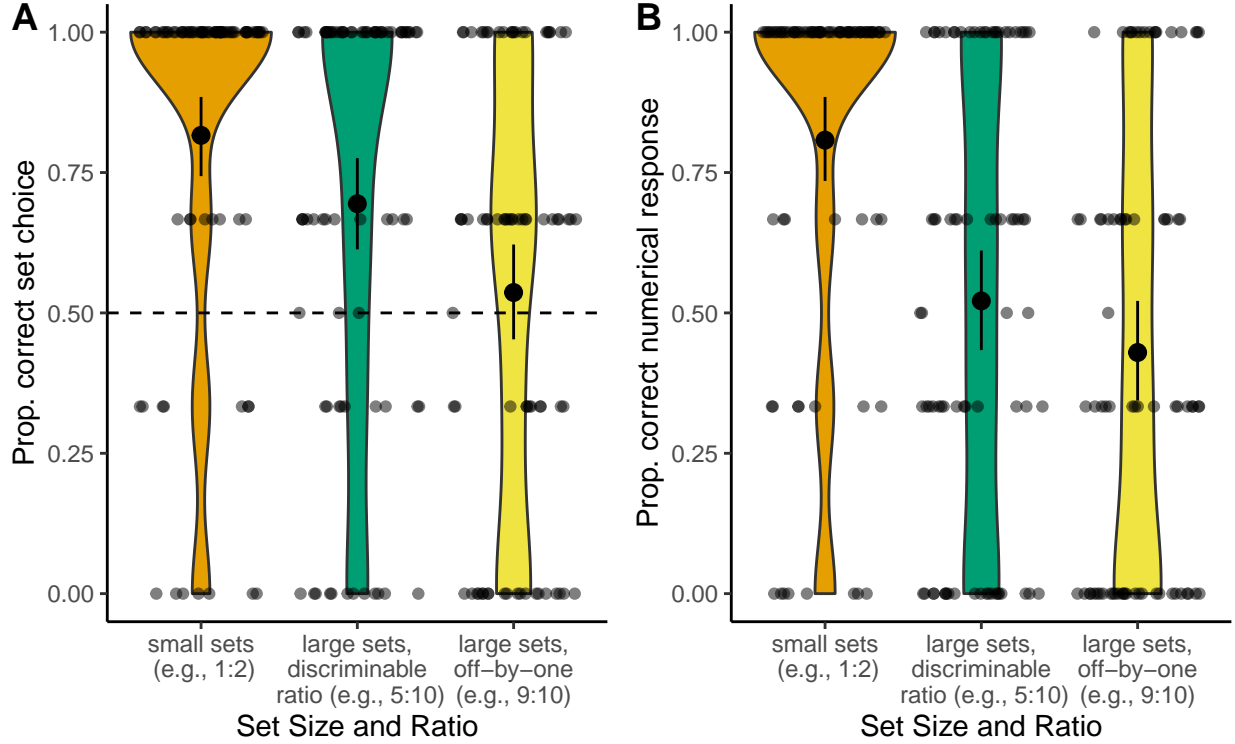
Figure 4

Schematic of Experiment 2. Slides are presented in order, from left to right and top to bottom. The corresponding audio description is under each slide.

made the Correct Set Choice. Two-tailed one-sample t-tests showed that overall performance was better than chance ($t(77) = 5.24$, $p < .001$). However, this was only true for small-set trials ($M = 0.81$, $SD = 0.39$; $t(77) = 8.47$, $p < .001$) and for large-set trials with perceptually-disciminable ratios ($M = 0.69$, $SD = 0.46$; $t(77) = 4.56$, $p < .001$). In large-set trials with an off-by-one difference ($M = 0.54$, $SD = 0.50$), children performed at chance overall ($t(77) = 0.84$, $p = .402$; Figure 5A). Performance in this experiment was even more variable than in Experiment 1 (Figure 6).

We next examined children's performance specifically on the off-by-one trials (Large-NR), which can only be solved correctly using one-to-one correspondence. Out of 78 children tested, only 21 (26.92%) succeeded in all trials, showing some evidence of using one-to-one correspondence in extending number words[11].

———

[11] On average, these 21 children were slightly older and had a greater Highest Count score ($M_{age} = 4.85$;

Figure 5

Each dot represents a participant. The width of the shaded area of violin plots represents the proportion of the data located there. A) Proportion of correct item set choice by participant against trial type. Horizontal dashed line indicates performance at chance = 0.50. B) Proportion of correct numerical response by participant against trial type.

To further analyze the effect of trial type on selection of the Correct Set Choice, we constructed a GLMM with the formula Correct Set Choice (0/1) ~ Age (z-scored) + (1 | Participant) + (1 | Trial), and tested whether adding Trial Type (Small sets / Large sets with perceptually discriminable ratios (Large-DR) / Large sets with an off-by-one difference (Large-NR)) as fixed effects improved the fit of the model. All models included by-subject and by-item random intercepts. We found no effect of age in our base model

$M_{HC} = 27.00$) compared to children who did not succeed in all trials ($M_{age} = 4.61$; $M_{HC} = 25.85$), though due to the small number of children we could not compare these values. Note that one of these 21 children contributed only 1 large trial with an off-by-one difference because they began counting the bunnies before the animals disappeared on the other 2 trials.
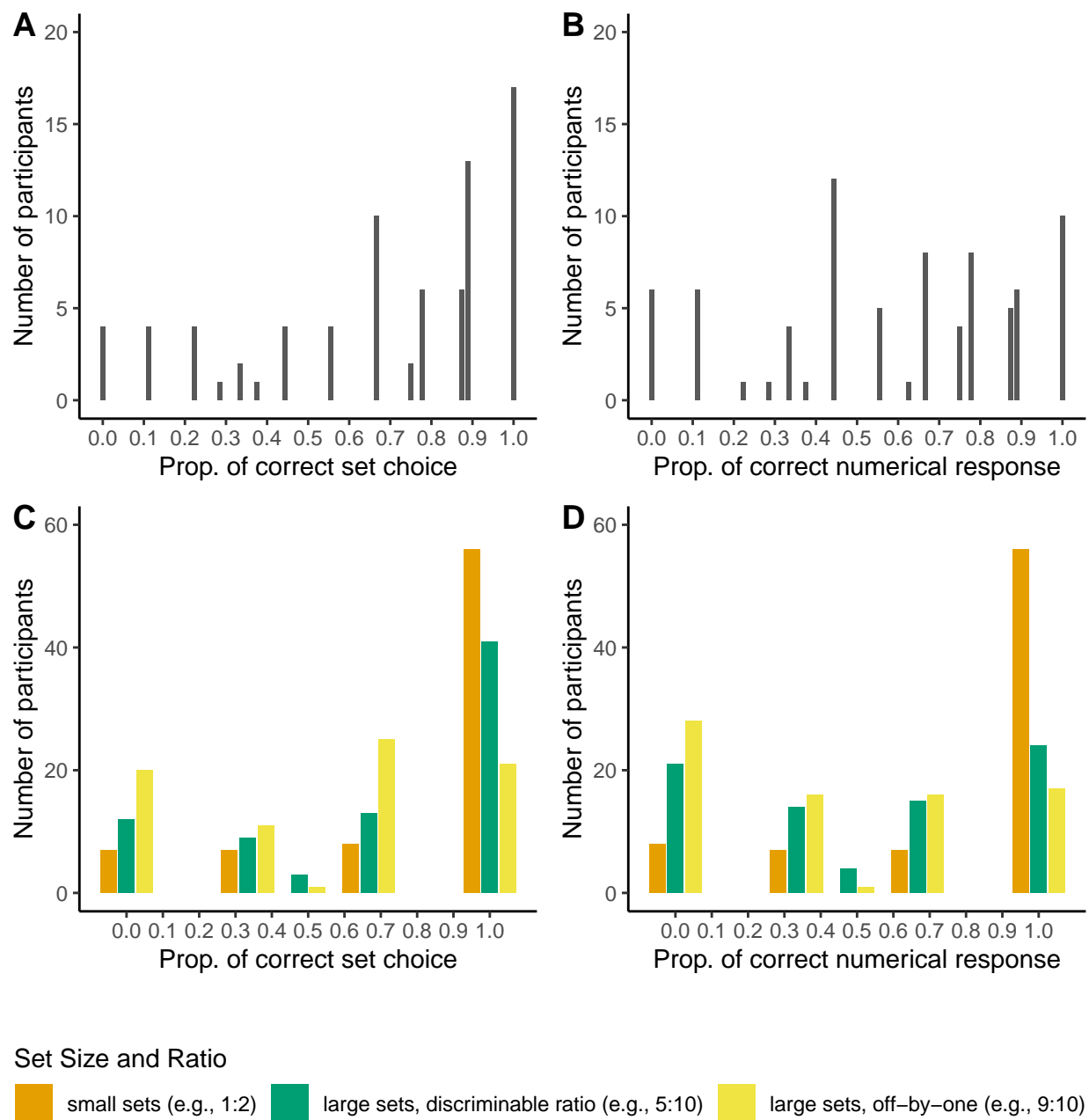
Figure 6

Variability in Experiment 2. All trials: A) Number of participants against proportion of correct item set choice. B) Number of participants against proportion of correct numerical response. By trial type: C) Number of participants against proportion of correct item set choice. D) Number of participants against proportion of correct numerical response.

Table 2

Distribution of responses in Large-DR and Large-NR trials.

	Number and Percentage of Trials per Trial Type	
Type of responses	Large trials with discriminable ratios	Large trials with an off-by-one difference
Correct	157 (69.47%)	123 (53.71%)
Incorrect, chose both sets	24 (10.62%)	25 (10.92%)
Incorrect, did not choose a set	32 (14.16%)	31 (13.54%)
Incorrect, chose distractor set	13 (5.75%)	50 (21.83%)

($p = .076$). When trial type was included as a fixed effect, we found a significant effect of trial type ($\chi^2(2) = 40.78$, $p < .001$), and still no effect of age ($p = .078$). The model with Trial Type explained significantly more variation in the observed data compared to the base model in a likelihood ratio test ($\chi^2(2) = 16.97$, $p < .001$, $\Delta AICc = $ -12.91, $\Delta BIC = $ -3.92). This finding replicated our finding from Experiment 1, which found that age did not explain cardinal extension performance in CP-knowers. *Post hoc* pairwise comparisons between the three trial types (Small, Large-DR, Large-NR) with Bonferroni correction found that children were more likely to provide correct responses in Small trials compared to Large-DR trials ($z = 3.09$, $p = .006$) and Large-NR trials ($z = 6.34$, $p < .001$). Success in Large-DR trials was also significantly higher than in Large-NR trials ($z = 3.75$, $p < .001$). When children fail to make the correct choice, they could make one of three error types: choosing both sets, not choosing either set, or choosing the wrong set (see Table 2). The most common error type in the Large-NR trials was choosing the wrong set, while the most common error type in the Large-DR trials was choosing neither set. Children chose both sets or neither at a similar rate in the Large-DR and Large-NR trials, but chose the wrong set more often in the Large-NR trials.

Similar to Experiment 1, we also conducted a follow-up analysis to explore how often participants gave the correct numerical response. On this measure, children succeeded in 80.35% of small trials, 52.65% of large trials with perceptually discriminable ratios, and 42.79% of large trials with an off-by-one difference (Figure 5B). Sixty-six of the 78 children tested (84.62%) gave a numerical response that was wrong on at least 1 trial, resulting in a total of 274 (out of 684) trials with the wrong numerical response. Out of these 274 trials, 65 trials (37 children) could be attributed to procedural error, where participants chose the correct set, but then counted wrongly. On 139 trials (47 children), participants counted the wrong set or both sets, which indicated failure in identifying the correct set for cardinal extension inference. Finally, on 70 trials (20 children), participants provided the wrong response but did not point to either set, which can either reflect overconfidence in estimating the number of items in the correct item set, or a lack of conceptual understanding about how the item set in one-to-one correspondence with the bunnies can be used to infer the number of bunnies. In comparison to the error pattern of CP-knowers' numerical response in Experiment 1, Experiment 2 found a larger proportion of errors that came from counting the wrong set or both sets, with this type of mistake accounting for more than half or the errors.

To further investigate the effect of different trial types on children's numerical response, we constructed a set of GLMMs with the same fixed and random effect structure as the Correct Set Choice analysis. However, we removed the by-trial random intercept due to overfitting. Specifically, our base model had the formula Correct Numerical Response (0/1) ~ Age (z-scored) + (1 | Participant), and we tested whether adding Trial Type (Small/Large-DR/Large-NR) improved the fit of the model. The model that best explained the data included both Age and Trial Type as predictors (compared against base model with only Age: $\chi^2(2) = 118.96$, $p < .001$, $\Delta AICc$ = -114.91, ΔBIC = -105.91). Older children were significantly better at providing the correct number of animals ($\beta = 0.54$, 95% CI = [0.03, 1.05], $p = .036$). Children's performance was also influenced by Trial

Type ($\chi^2(2) = 82.20$, $p < .001$), and *post hoc* pairwise comparisons with Bonferroni

correction showed the same pattern of success across trial types as the above analysis.

CP-knowers were more successful in Small trials compared to Large-DR trials ($z = 7.15$,

$p < .001$) and Large-NR trials ($z = 8.97$, $p < .001$), and success in Large-DR trials was also

significantly more likely than in Large-NR trials ($z = 2.84$, $p = .014$).

As an exploratory analysis, we added children's Highest Count as a fixed effect to

the best-performing models from the two analysis approaches above. Seventy-five children

(out of 78) provided a Highest Count, and performed similarly to the CP-knowers from

Experiment 1 ($M_{HC} = 26.04$ [3; 100]; $SD_{HC} = 20.04$). After excluding data from the three

children who did not provide a Highest Count, we found the same pattern as in

Experiment 1: Highest Count did not explain additional variability for either Correct Set

($\chi^2(1) = 1.92$, $p = .166$, $\Delta AICc = 0.12$, $\Delta BIC = 4.57$) or Correct Numerical Response

($\chi^2(1) = 1.94$, $p = .164$, $\Delta AICc = 0.10$, $\Delta BIC = 4.55$). Similar to Experiment 1, Highest

Count showed low multicollinearity with other predictors (Variance Inflation Factor < 2,

following the threshold established by James et al., 2013).

During our analysis of data for Experiment 2, we became concerned that children

might fail to select the correct set on large non-discriminable ratio (Large-NR) trials due to

failures of memory. Specifically, given that the sets are not perceptually discriminable,

after the bunnies disappeared into the school, children were required to recall which set

was missing one item in order to prefer the other set that was an exact match. Therefore,

children might fail at this task not because they lack an understanding of exact equality

but because they can't recall which set was exactly equal to the number of bunnies. To

explore this possibility, we tested an additional 23 CP-knowers (13F, 10M; $M_{age} = 4.83$

[3.78; 5.87]; $SD_{age} = 0.53$) on a reduced version of the task that included just 3 Large-NR

critical trials and asked them at the end of each trial which set had a missing element

("There was a bunny that was missing something. What was it missing, a [item 1] or [item

2]?"). We found that all 23 children remembered the missing item on all 3 trials. Critically,

despite succeeding at this memory check, these children nevertheless showed the same rate of failure at cardinal extension, and chose the correct set only 49.28% of the time. This provides strong evidence that even when children recognize that only one set exactly matches the target set, they often choose an approximate match anyway.

Discussion

Experiment 2 found that while children who were classified as CP-knowers succeeded on trials with small subitizable sets and large, perceptually discriminable sets, they generally failed when the distractor set and target set could not be discriminated using approximate number representations, and differed by just 1 item. A follow-up study found that this was not merely due to difficulties recalling which set was an exact match, and that children who were able to identify which set was missing an item nevertheless often made an approximate match regardless. These results suggest that, even when 3- to 6-year-old CP-knowers engage in cardinal extension, their dominant strategy is to extend number words to sets that are perceptually non-discriminable, even when one-to-one correspondence is violated. Additionally, we again found substantial variability in CP-knowers' understanding of cardinal extension, with only a few children appearing to truly understand the relationship between one-to-one correspondence, equinumerosity, and cardinal extension.

General Discussion

In two experiments, we investigated the role that counting knowledge plays in children's understanding that equal sets should be given the same numerical label. We found three main results. First, compatible with some previous studies (e.g., Sarnecka & Gelman, 2004; Sarnecka & Wright, 2013), Experiment 1 found that CP-knowers were more likely than subset-knowers to infer that two sets should receive the same number label only if they are numerically equal. In fact, subset-knowers performed at chance for both small and large sets. This provides evidence against the hypothesis (Sarnecka & Gelman, 2004)

that children understand that all number words have specific, exact meanings even before children become CP-knowers and can construct sets of large numbers. Second, Experiment 1 found that although many CP-knowers were able to choose the correct item set to infer the quantity of a hidden set, there was a fair degree of variability in CP-knowers' performance, suggesting that simply learning to count does not guarantee understanding of cardinal extension. Third, Experiment 2 provided further evidence that CP knowledge does not alone furnish understanding of cardinal extension. Although CP-knowers often correctly extended labels to equal sets when they were within the subitizable range or had perceptually discriminable ratios, most did not use violations of one-to-one correspondence as a cue to exact equality and often extended the same label to large sets that differed by 1 item. This is despite the fact that CP-knowers were able to notice violations of one-to-one correspondence in a follow-up study. Overall, we found that CP-knowers rarely used an adult-like strategy of using one-to-one correspondence to support cardinal extension. This suggests that CP-knowers initially extend number words to sets that are perceptually indiscriminable, without regard to whether they stand in one-to-one correspondence, and only later restrict cardinal extension to sets that are exactly equal.

These results of these experiments are compatible with the hypothesis that full understanding of cardinal extension - and thus how number words label specific, exact, numerosities - does not emerge until sometime after children become competent counters. These data are difficult to explain for theories that predict earlier learning and particularly with the hypothesis that this knowledge is acquired when children are subset-knowers. When subset-knowers - and some CP-knowers - were asked to say how many animals were hidden in the school, they often counted the items of the wrong animal (Experiment 1). In doing so, these children failed to understand that only items from the same animal type - that were equal to the number of hidden animals - could predict the number of animals that were in the school.

Our results also call into question the hypothesis that becoming a CP-knower

supplies adult-like knowledge of cardinal extension. According to some previous reports, becoming a CP-knower involves constructing an analogical mapping between the order of the count list and the set of ordered cardinalities created by increments of one individual, and therefore noticing that there is a one-to-one correspondence between number words and exact cardinal values, *inter alia* (Carey, 2004, 2009). This supports the inference that only sets in one-to-one correspondence are equinumerous and can be denoted by the same number word (Le Corre & Carey, 2007). This account posits that learning the cardinal principle entails understanding that numbers have exact meanings, and that number words can only be extended between sets in one-to-one correspondence. Against this, Experiment 2 found that CP-knowers performed at chance when tested with perceptually non-discriminable sets that clearly violated one-to-one correspondence. Taken together, our results suggest that even when CP-knowers seem to succeed in cardinal extension (as in Experiment 1), they extend number words based on their approximate match in magnitude, rather than on an exact match in number. Overall, this suggests that understanding of how number words encode exact equality continues to develop well after children master the counting procedures required to become a CP-knower.

While our results are broadly consistent with some previous findings, we also introduce several new findings to the literature, which may at first appear to conflict with some past studies. For example, our finding that CP-knowers extend cardinal labels to two sets even if they violate one-to-one correspondence seems, at first, to be at odds with previous studies that document an ability among CP-knowers to judge that sets that differ by just 1 item should be labeled by different number words (Sarnecka & Gelman, 2004; Sarnecka & Wright, 2013). However, like our study, these studies also found variability in CP-knowers' performance, suggesting that many CP-knowers fail to recognize that equinumerous sets should receive the same label - the first component of cardinal extension. Also, it's unclear in past studies whether children who extended the same label to two sets were indeed sensitive to exact equality - the second component of cardinal

extension. For example, in Sarnecka and Wright (2013) children had to judge whether two sets (5 vs. 6) were the same or different before being asked to provide numerical labels, and often received feedback when they failed to notice differences. Also, given how the items were arranged, children could use perceptual proxies like line length to approximate differences between the two sets, rather than using exact number. Similarly, as noted in the Introduction, although children in Sarnecka and Gelman (2004) judged that two previously equal sets should receive distinct numerical labels after one set had an item added or subtracted, children only needed to know the principle that "different numbers get different labels", or even "different referents get different labels", in keeping with the principle of contrast (see Brooks et al., 2013).

The central puzzle that flows from our findings is why some CP-knowers are able to use a one-to-one counting procedure to reliably count objects and construct large sets, but fail to use violations of one-to-one correspondence to inform cardinal extension. Neither age nor Highest Count seemed to explain the variability of performance in CP-knowers, suggesting that the kind of learning that leads to understanding of cardinal extension (including the use of one-to-one correspondence) is independent of rote training on the count list, but also that it involves more than just simple maturation. One possibility, proposed by Davidson et al. (2012), is that CP-knowers have acquired the ability to construct large sets as one of several rote counting procedures that they learn in early childhood [e.g., the 'how-to-count' principles described in Gelman and Gallistel (1978); including the one-to-one principle, the stable order principle, and the last-word principle, *inter alia*]. On this hypothesis, CP-knowers might know how to accurately count and construct large sets, but not understand the underlying conceptual principles that these rote behaviors reflect. That is, CP-knowers' understanding of counting may reflect an understanding of what counting is doing, but not what it is accomplishing (Heck, 2000). For example, when asked to 'give six fish,' these children might follow a procedure in this form: 'Begin counting from one, for each number word partition one item to a separate set,

stop counting at six, and give all counted objects.' This procedure results in an accurate set of six, but requires minimal conceptual understanding of the meaning of 'six' – for example, that 'six' denotes the same cardinality across any set constructed following the procedure or that 'six' denotes 'exactly six.' Lacking this understanding, these children might not realize that, if two sets stand in one-to-one correspondence, then counting one set indicates the cardinality of the second one. On this view, the "CP" in CP-knower may not stand for "cardinal principle" but instead for "cardinal procedure", since many CP-knowers appear to lack an understanding of how counting represents cardinality, but know how to deploy a procedure that nevertheless returns accurate cardinal labels.

However, alternative explanations for CP-knowers' limited understanding of cardinal extension also exist. One possibility, described by Schneider et al. (2022), is that very young children know that two sets are equal only if they stand in one-to-one correspondence, but they lack a reliable procedure for verifying this property for large sets - a critical prerequisite for adult-like cardinal extension. As noted by Schneider et al. (2022), studies of "parallel individuation" find that infants track small sets of hidden objects by assigning each a mental index, such that there is a one-to-one correspondence between these mental symbols and the things they represent (e.g., Feigenson & Carey, 2003). However, these symbols are represented in parallel and are therefore subject to a capacity limit, such that infants cannot track more than 3 objects at a time. What infants lack, therefore, is a sequential, memory-free, verification procedure for establishing one-to-one correspondence. Schneider et al. (2022) argued that counting provides just such a procedure and solves the "partition" problem (Gelman & Gallistel, 1978), by partitioning a set, one object at a time, into counted vs. uncounted objects. As each item is counted, it is tagged with a number word, such that both that word and the object it labels are removed from further consideration (e.g., by setting it aside) and a one-to-one correspondence is created between words and objects. Consequently, to arrive at an exact response, the child needs only to keep track of where they currently are in the count list,

and the next object they plan to count, making the process sequential and memory-free. One possibility, then, is that when children learn this one-to-one verification procedure, they extend the ability to sequentially map labels-to-objects to the problem of mapping objects-to-objects in sequence, a skill required for matching large sets.

Our study provides some evidence against the hypothesis laid out by Schneider et al. (2022), since in our studies, children did not require procedural knowledge to create one-to-one matches because these matches were created for them. Further, as shown in a follow-up study in Experiment 2, CP-knowers readily noticed violations of one-to-one correspondence when they arose. This raises the possibility that when children fail at set matching, they may truly fail to understand that sets are only equal if items in each are placed in one-to-one correspondence. Thus, it's unlikely that children in set matching studies have an adult-like appreciation of exact number, and fail to exactly match strictly due to missing procedure knowledge. That said, this doesn't rule out that learning procedures is what drives learning about exact number. For example, it may be that experience with counting provides children with evidence that adding or removing 1 item from a set changes the result of a count, and therefore that even non-perceptible differences in cardinality - like a difference of just 1 - are important to cardinal extension.

Compatible with this, there is independent evidence that procedural knowledge of counting generally precedes conceptual understanding from early in development (Fuson et al., 1983; Fuson, 1988), and that this continues past the CP-knowers stage, as Davidson et al. (2012) document. Also, previousresearch has found that children are more successful at comparing cardinalities or constructing sets of the same quantities when they can count sets, as opposed to placing items in one-to-one correspondence (Cowan & Daniels, 1989; Michie, 1984; Russac, 1978; Saxe, 1977). This suggests a progression whereby children become increasingly accurate at counting to establish cardinalities, with the ability to integrate this with reasoning about how one-to-one correspondence can be used to determine logical set equality developing afterwards. Furthermore, previous research has

shown that children's counting proficiency seems to continue to develop even after they become CP-knowers. For example, children who are classified as CP-knowers on the basis of their ability to accurately give sets of 5 and 6 often make errors when counting larger sets, such as 7, 8, or 9 (Krajcsi & Fintor, 2023; Wagner & Johnson, 2011). Also, CP-knowers vary in their ability to recognize mistakes in counting procedures Briars & Siegler (1984). Particularly relevant to the current discussion, Muldoon et al. (2003) found that children's ability to detect violations of counting procedures predicted their success in a cardinal extension task (though this might be explained by the presence of subset-knowers, since Muldoon et al., 2003 did not assess knower levels in this study). Perhaps as children gain more counting experience, they become better at deploying the counting procedures and noticing violations in counting. This might result in them realizing that violations of one-to-one between count words and items results in different numerical labels. More exposure to counting might also lead children to notice that adding or removing an item from a set results in a different cardinality. Given these possibilities, future studies should ask whether variability in cardinal extension performance among CP-knowers might be explained by differing abilities to deploy counting procedures, as well as metacognitive abilities about the counting principles.

References

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology*, *96*(4), 699. https://doi.org/https://doi.org/10.1037/0022-0663.96.4.699

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown.* https://github.com/crsh/papaja

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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

Briars, D., & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology*, *20*(4), 607. https://doi.org/10.1037/0012-1649.20.4.607

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. https://doi.org/10.1037/a0029384

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*, 59–68. https://doi.org/10.1162/001152604772746701

Carey, S. (2009). Where Our Number Concepts Come From. *The Journal of Philosophy*, *106*(4), 220–254. https://doi.org/10.5840/jphil2009106418

Chu, J., Cheung, P., Schneider, R. M., Sullivan, J., & Barner, D. (2020). Counting to infinity: Does learning the syntax of the count list predict knowledge that numbers are infinite? *Cognitive Science*, *44*(8), e12875. https://doi.org/https://doi.org/10.1111/cogs.12875

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), 22–38. https://doi.org/10.1037/0096-3445.137.1.22

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*, 698–707. https://doi.org/10.3758/BF03196206

Cowan, R., & Daniels, H. (1989). Children's use of counting and guidelines in judging relative number. *British Journal of Educational Psychology*, *59*(2), 200–210. https://doi.org/10.1111/j.2044-8279.1989.tb03091.x

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

Decock, L. (2008). Neo-fregeanism naturalized: The role of one-to-one correspondence in numerical cognition. *Behavioral and Brain Sciences*, *31*(6), 648–649. https://doi.org/10.1017/S0140525X08005645

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

Dehaene, S. (2009). Origins of mathematical intuitions: The case of arithmetic. *Annals of the New York Academy of Sciences*, *1156*(1), 232–259. https://doi.org/10.1111/j.1749-6632.2009.04469.x

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, *6*(5), 568–584. https://doi.org/10.1111/1467-7687.00313

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frege, G. (1880). *Booles rechnende logik und die begriffsschrift.*

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. https://doi.org/10.1016/0885-2014(88)90019-6

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

Fuson, K. C., Secada, W. G., & Hall, J. W. (1983). Matching, counting, and conservation of numerical equivalence. *Child Development*, 91–97. https://doi.org/10.2307/1129865

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

Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition*, *13*(3), 343–359. https://doi.org/10.1016/0010-0277(83)90014-8

Gelman, R., Meck, E., & Merkin, S. (1986). Young children's numerical competence. *Cognitive Development*, *1*(1), 1–29. https://doi.org/10.1016/S0885-2014(86)80021-1

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*(5695), 496–499. https://doi.org/10.1126/science.1094492

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457. https://doi.org/10.1037/a0012682

Heck, R. G. (2000). Cardinality, counting, and equinumerosity. *Notre Dame Journal of Formal Logic*, *41*(3), 187–209. https://doi.org/10.1305/ndjfl/1038336841

Huntley-Fenner, G., & Cannon, E. (2000). Preschoolers' magnitude comparisons are mediated by a preverbal analog mechanism. *Psychological Science*, *11*(2), 147–152. https://doi.org/10.1111/1467-9280.00230

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. https://doi.org/10.1016/j.cogpsych.2014.01.004

James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Koponen, T., Salmi, P., Torppa, M., Eklund, K., Aro, T., Aro, M., Poikkeus, A.-M., Lerkkanen, M.-K., & Nurmi, J.-E. (2016). Counting and rapid naming predict the fluency of arithmetic and reading skills. *Contemporary Educational Psychology*, *44*, 83–94. https://doi.org/https://doi.org/10.1016/j.cedpsych.2016.02.004

Krajcsi, A., & Fintor, E. (2023). A refined description of initial symbolic number acquisition. *Cognitive Development*, *65*, 101288. https://doi.org/10.1016/j.cogdev.2022.101288

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438. https://doi.org/10.1016/j.cognition.2006.10.005

Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology*, *52*(2), 130–169. https://doi.org/10.1016/j.cogpsych.2005.07.002

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means.* https://CRAN.R-project.org/package=emmeans

McGarrigle, J., & Donaldson, M. (1974). Conservation accidents. *Cognition*, *3*(4), 341–350. https://doi.org/10.1016/0010-0277(74)90003-1

Mehler, J., & Bever, T. G. (1967). Cognitive capacity of very young children. *Science*, *158*(3797), 141–142. https://doi.org/10.1126/science.158.3797.141

Michie, S. (1984). Number understanding in preschool children. *British Journal of Educational Psychology*, *54*(3), 245–253. https://doi.org/10.1111/j.2044-8279.1984.tb02588.x

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. https://doi.org/10.1016/j.ijer.2004.10.006

Muldoon, K., Lewis, C., & Towse, J. (2005). Because it's there! Why some children count, rather than infer numerical relationships. *Cognitive Development*, *20*(3), 472–491. https://doi.org/10.1016/j.cogdev.2005.05.008

Piaget, J. (1965). The stages of the intellectual development of the child. *Educational Psychology in Context: Readings for Future Teachers*, *63*(4), 98–106.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rose, S. A., & Blank, M. (1974). The potency of context in children's cognition: An illustration through conservation. *Child Development*, 499–502. https://doi.org/10.2307/1127977

Russac, R. (1978). The relation between two strategies of cardinal number: Correspondence and counting. *Child Development*, 728–735. https://doi.org/10.2307/1128241

Sarnecka, B., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition*, *92*(3), 329–352. https://doi.org/10.1016/j.cognition.2003.10.001

Sarnecka, B., & Wright, C. E. (2013). The Idea of an Exact Number: Children's Understanding of Cardinality and Equinumerosity. *Cognitive Science*, *37*(8), 1493–1506. https://doi.org/10.1111/cogs.12043

Saxe, G. B. (1977). A developmental analysis of notational counting. *Child Development*, 1512–1520. https://doi.org/10.2307/1128514

Schneider, R. M., Brockbank, E., Feiman, R., & Barner, D. (2022). Counting and the ontogenetic origins of exact equality. *Cognition*, *218*, 104952. https://doi.org/10.1016/j.cognition.2021.104952

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. https://doi.org/https://doi.org/10.1111/cdev.13524

Slusser, E. B., & Sarnecka, B. W. (2011). Find the picture of eight turtles: A link between children's counting and their knowledge of number word semantics. *Journal of Experimental Child Psychology*, *110*(1), 38–51. https://doi.org/10.1016/j.jecp.2011.03.006

Sophian, C., Wood, A. M., & Vong, K. I. (1995). Making numbers count: The early development of numerical inferences. *Developmental Psychology*, *31*(2), 263. https://doi.org/10.1037/0012-1649.31.2.263

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, *119*(1), 10–22. https://doi.org/10.1016/j.cognition.2010.11.014

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130–137. https://doi.org/10.1111/1467-9280.00120

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilke, C. O. (2024). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'.* https://CRAN.R-project.org/package=cowplot

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251. https://doi.org/10.1016/0010-0285(92)90008-P

Zhang, X., Koponen, T., Räsänen, P., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2014). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Development*, *85*(3), 1091–1107. https://doi.org/https://doi.org/10.1111/cdev.12173