COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13299

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13299

# Is Nonsymbolic Arithmetic Truly "Arithmetic"? Examining the Computational Capacity of the Approximate Number System in Young Children

Chen Cheng,[a] Melissa M. Kibbe[b]

[a]*Division of Social Science, Hong Kong University of Science and Technology*
[b]*Department of Psychological and Brain Sciences, Boston University*

Received 4 April 2022; received in revised form 9 January 2023; accepted 10 May 2023

**Abstract**

Young children with limited knowledge of formal mathematics can intuitively perform basic arithmetic-like operations over nonsymbolic, approximate representations of quantity. However, the algorithmic rules that guide such nonsymbolic operations are not entirely clear. We asked whether nonsymbolic arithmetic operations have a function-like structure, like symbolic arithmetic. Children ($n = 74$ 4- to -8-year-olds in Experiment 1; $n = 52$ 7- to 8-year-olds in Experiment 2) first solved two nonsymbolic arithmetic problems. We then showed children two unequal sets of objects, and asked children which of the two derived solutions should be added to the smaller of the two sets to make them "about the same." We hypothesized that, if nonsymbolic arithmetic follows similar function rules to symbolic arithmetic, then children should be able to use the solutions of nonsymbolic computations as inputs into another nonsymbolic problem. Contrary to this hypothesis, we found that children were unable to reliably do so, suggesting that these solutions may not operate as independent representations that can be used inputs into other nonsymbolic computations. These results suggest that nonsymbolic and symbolic arithmetic computations are algorithmically distinct, which may limit the extent to which children can leverage nonsymbolic arithmetic intuitions to acquire formal mathematics knowledge.

*Keywords:* Nonsymbolic arithmetic; Numerical cognition; Mathematical cognition; Approximate number system; Function arithmetic

Correspondence should be sent to Melissa M. Kibbe, Department of Psychological and Brain Sciences, Boston University, 64 Cummington Mall, Boston, 02215 MA, USA. E-mail: kibbe@bu.edu

## 1. Introduction

Marr (1982) noted that to understand any information processing system, including the human mind, we must define not only how the system *represents* information but the *algorithms* by which these representations can be processed by the system, since the algorithm constrains the representational format of its inputs and outputs. For example, while one can add two Arabic numerals together, and one can add two Roman numerals together, the algorithms by which one does so are not the same: place value addition can operate over Arabic numerals, but not Roman numerals. The study of representation and algorithm must go hand in hand in order to fully understand the computational capacity of any information processing system, from cash registers to minds.

A critical case study for the theoretical importance of studying both representation and algorithm in the human mind is the approximate number system (ANS). The ANS is a basic and universal cognitive system that enables us to quantify sets of items without language or formal symbols[1] (Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004). The ANS represents sets of individual items as a single magnitude that is noisy and imprecise (Dehaene, 1997; Gallistel & Gelman, 1992; Meck & Church, 1983; Wynn, 1995), and the discriminability of ANS representations depends on the ratio between the quantities they represent (Gallistel & Gelman, 2000; Lipton & Spelke, 2003; Pica, Lemer, Izard, & Dehaene, 2004; Xu, 2003; Xu & Spelke, 2000). The ANS is what allows us to get a rough sense of the size of a crowd, or to tell whether our sibling got more cereal than we did, without having to count. Crucially, the ANS is operational from infancy (Coubart, Izard, Spelke, Marie, & Streri, 2014; de Hevia, Izard, Coubart, Spelke, & Streri, 2014; Libertus & Brannon, 2009), and therefore, its emergence is not dependent on language acquisition or formal tuition. Its representational precision increases with development (Halberda & Feigenson, 2008).

From very early in development, humans can manipulate ANS representations in response to real-world changes to visible arrays of items. Infants and children can add to or subtract from quantities represented by the ANS (Booth & Siegler, 2008; Barth et al., 2006; McCrink & Wynn, 2004; Gilmore & Spelke, 2008; Kibbe & Feigenson, 2015, 2017; see Christodoulou et al., 2017, for a meta-analysis) and children can scale ANS representations up or down by a factor of, for example, 2 or 4 (McCrink & Spelke, 2010, 2016; McCrink & Wynn, 2007; McCrink, Shafto, & Barth, 2017). For example, after viewing two separate sets of items hidden sequentially behind an occluder, infants and children can generate an approximate representation of the total sum (e.g., McCrink & Wynn, 2004, 2009; Park, Bermudez, Roberts, & Brannon, 2016). Children's success in these tasks is dependent on the precision of their ANS representations, providing evidence that children are indeed using the ANS, and not symbolically mediated representations (like counting).

Children's untutored capacity to manipulate ANS representations has been taken as evidence of an early, pre-symbolic capacity for *arithmetic* computation that precedes formal mathematics education (e.g., Carey, 2001; Dehaene, 1997; Wynn, 1992). Indeed, before they have formal mathematics instruction, children appear to be able to manipulate nonsymbolic representations of quantity in a way that bears at least surface similarity to arithmetic with numerals–adding, subtracting, multiplying, and dividing representations of sets of items.

However, the algorithmic rules that guide such nonsymbolic manipulation processes are not entirely clear. One wedge into this problem is to examine whether the computational structure of the human capacity for nonsymbolic manipulations of ANS representations bears algorithmic similarity to the computational rules of symbolic arithmetic. True symbolic arithmetic is defined by *function rules* (Dedekind & Beman, 1901). These rules hold that (1) operations (like addition or subtraction) are functions performed over independent numerals, and (2) the outputs of these operations are themselves independent numerals that can be used as input into other arithmetic operations. For example, to solve the symbolic arithmetic problem $8+7 = \_$, one performs an operation (addition) over two independent numerals (8 and 7), and that operation specifies how these numerals should be combined. The result is another numeral (the output of the operation, 15), which is independent of the operation from which it was derived and can be used as input into another operation (e.g., 15 could now be used as an operand in a subtraction problem).

Function arithmetic is powerful because it allows for the principled *combination* of independent numerals. It is an open question whether the nonsymbolic manipulation of ANS representations affords similar computational power. That is, while children's early nonsymbolic capacities may bear a surface resemblance to symbolic arithmetic, it is unclear whether they bear *algorithmic* similarity to symbolic arithmetic. Understanding the algorithms that support nonsymbolic ANS computation is crucial for understanding the operational structure and computational capacity of our early quantificational abilities.

Previous work has revealed both differences and similarities between nonsymbolic and symbolic quantificational abilities in children and adults. The representational formats of symbolic and nonsymbolic quantity are different in several aspects. Symbolic representations are exact, while ANS representations are noisy and imprecise; symbolic and ANS representations of quantity are coded differently in the brain (Bulthé et al., 2014; Cohen Kadosh et al., 2011; Lyons et al., 2015); and different cognitive precursors (e.g., visuospatial skills and phonological awareness) are found to uniquely predict children's performance on symbolic and nonsymbolic ANS tasks (Yang et al., 2020; Zhang & Lin, 2015). Yet, symbolic and nonsymbolic operations also share some surface similarities. There is overlap in the brain regions that are activated when adults and children solve nonsymbolic and symbolic arithmetic problems (Butterworth & Walsh, 2011; Dehaene & Cohen, 2007; Dehaene, Piazza, Pinel, & Cohen, 2003; Lussier & Cantlon, 2016; Piazza, Pinel, Le Bihan, & Dehaene, 2007), and some evidence suggests that nonsymbolic numerical skills are related to symbolic math performance in children and adults (Lourenco et al., 2012; Chu, vanMarle, & Geary, 2015; DeWind & Brannon, 2012; Gilmore et al., 2010; Halberda, Mazzocco, & Feigenson, 2008; Libertus, Odic, & Halberda, 2012; Mazzocco, Feigenson, & Halberda, 2011a; Piazza et al., 2010; Mazzocco, Feigenson, & Halberda, 2011b; Starr, Libertus, & Brannon, 2013; Wang, Halberda, & Feigenson, 2021; Wang, Odic, Halberda, & Feigenson, 2016; Wong, 2020; see Chen & Li, 2014 and Fazio et al., 2014 for meta-analyses), although other studies have failed to find such a relationship (see Szkudlarek, Park, & Brannon, 2021; Bugden, Szkudlarek, & Brannon, 2021). And children can use approximate representations of quantity to support performance on symbolic tasks for which they are not yet adept (e.g., Gilmore et al., 2007), suggesting some transference between nonsymbolic and symbolic arithmetic abilities.

Nevertheless, the algorithms that underlie nonsymbolic arithmetic-like computation are not well understood.

How is nonsymbolic arithmetic computation accomplished over ANS representations? Despite surface-level similarities, nonsymbolic "arithmetic" could be accomplished in a way that does not require computation over inputs to produce an output. Take the example of nonsymbolic "addition." When viewing the sequential hiding of two sets of objects, participants could maintain a running total, updating a single magnitude in response to each new visual encounter with an array. Such a process is not a function, because it does not require the principled operation over independent input(s) to yield a specific, independent output. Instead, such a process could be thought of as a visual updating process: a single ANS representation whose magnitude is updated in response to visual input, metaphorically akin to increasing or decreasing the amount of liquid in a single cup. The outcome of such a process is, therefore, dependent upon the context in which it was initially formed (i.e., it is a modified form of the initial representation). This updating process could be accomplished in the same way ANS representations are initially formed (see, e.g., Meck & Church, 1983, Gallistel & Gelman, 2000; Inglis & Gilmore, 2013; Whalen, Gallistel, & Gelman, 1999; Wynn, 1998), and can explain the results of the majority of nonsymbolic arithmetic studies. For example, "subtraction" can be accomplished by decrementing a single ANS representation in response to visual input, and "multiplication" or "division" could be accomplished by scaling up or scaling down a single ANS representation in response to visual input.

Alternatively, nonsymbolic arithmetic may be accomplished in a way that more closely parallels the *function rules* of symbolic arithmetic computations, in which independent representations are input into a function that produces some output, which itself is an independent representation that can be further used in additional computations. For example, when performing a nonsymbolic "addition" task, children could combine their (separate) representations of the first set and second set using a mental operation, producing an output ANS representation that represents their sum. This process is algorithmically distinct from updating a single ANS representation, because it is a function that takes ANS representations as input and outputs a separate ANS representation that is the (algorithmically defined) combination of the inputs.

There is some initial evidence that at least one criterion for function arithmetic may be met in children's nonsymbolic arithmetic abilities. Kibbe and Feigenson (2015, 2017) showed young children, with limited experience with formal math, problems like $5+x = 17$ presented nonsymbolically: children viewed an initial quantity, observed a "magic cup" adding an unknown quantity, and then observed the final quantity after the contents of the cup were added. Children were then asked to choose which of two visible quantities matched the quantity that was added by the magic cup–essentially requiring children to "solve for x" in a nonsymbolic unknown-addend problem. Solving for an unknown nonsymbolic addend cannot be accomplished by incrementing or decrementing a single ANS representation. Instead, to succeed, children had to hold two separate ANS representations in working memory (the initial quantity before the cup was added and the final quantity after the cup was added) and perform an operation over these representations (e.g., taking their difference) to derive a solution (the quantity in the cup). Kibbe and Feigenson (2015, 2017) found that 4- to

6-year-old children successfully solved these kinds of problems, and follow-up studies ruled out lower-level perceptual strategies or symbolically mediated strategies that could account for children's success.

Kibbe and Feigenson's (2015, 2017) results suggest the possibility that nonsymbolic arithmetic operations may partly satisfy one of the rules of symbolic function arithmetic: children with little formal arithmetic training could perform a computation over two independent ANS representations to derive their difference, suggesting that nonsymbolic arithmetic operations may be performed over two separate ANS representations. However, the true test of whether nonsymbolic arithmetic computations abide by arithmetic function rules requires evidence that the output of a nonsymbolic arithmetic computation is an ANS representation that can function *independently of the context in which it arose*, and that nonsymbolic arithmetic computations *can be performed over the outputs* of such computations.

If children can use the solutions of nonsymbolic arithmetic computations as operands in further nonsymbolic computations, it would suggest that nonsymbolic arithmetic computation may have a function-like structure that allows for the principled combination of inputs, and, therefore, can take as input any ANS representation, even one that arose from an arithmetic computation. Here, we make the distinction between "computationally derived" ANS representations–referring to ANS representations that result from mental computations over ANS representations (i.e., the "outputs" of ANS computations)–and "visually derived" ANS representations–referring to ANS representations that arise from visual inputs (i.e., a set of items in the world that gives rise to a representation of quantity in the mind). If nonsymbolic arithmetic computations yield independent ANS representations as outputs, and these *computationally derived outputs* are treated computationally identically to purely *visually derived* ANS representations, that would suggest a combinatorial, function-like structure to nonsymbolic ANS computations which can operate over representations in the mind that are not necessarily grounded in the world. Alternatively, if computationally derived ANS representations cannot be readily used in further computations, it would suggest a more limited computational capacity for nonsymbolic arithmetic.

In two experiments, we investigated the functional capacity of the ANS by asking whether 4- to 8-year-old children could use the solutions to nonsymbolic arithmetic computations over ANS representations as operands in another nonsymbolic computation. We asked children to solve two nonsymbolic unknown-addend problems (as in Kibbe & Feigenson, 2015, Experiment 5), which required children to perform an arithmetic-like operation over two separate ANS representations, obtaining their difference. Because it is not possible to accomplish the nonsymbolic unknown-addend computation by incrementing or decrementing a single ANS representation in response to visual input, we reasoned that this problem format would be a strong test case, requiring children to combine two ANS representations to derive a third representation (the solution). We then showed children two sets containing different quantities of objects, and asked children which of the two solutions should be added to the smaller set of objects to make the two sets "about the same." Thus, the task required children to solve two nonsymbolic arithmetic problems, and to then use the remembered solutions to solve another nonsymbolic arithmetic problem with those solutions (i.e., deciding whether adding one solution or the other solution to the smaller set will balance the unequal sets).

We also included several additional measures to gain insights into potential sources of children's success or failure at the task, including examining children's ability to balance unequal sets by selecting between two *visible* quantities (thus requiring children to use *visually derived* ANS representations rather than *computationally derived* ANS representations to perform a similar nonsymbolic balancing operation) and examining the precision of children's representations of the computed addends by asking children to compare their representations of the unknown addends to visible quantities. We hypothesized that, if nonsymbolic arithmetic computations conform to arithmetic function rules, children should be able to use their representations of the solutions to the two nonsymbolic unknown-addend problems to balance the unequal sets. We also hypothesized that this ability may be limited by the precision of children's representations of the unknown addends.

## 2. Experiment 1

### 2.1. Method

#### 2.1.1. Participants

Participants were 74 4- to 8-year-old children (mean age: 6.3 years; range: 4.01–9.02 years). To obtain the sample size for this experiment, we conducted a power analysis based on the results of a small exploratory study (see Supplement for exploratory study details). The sample was powered for a binomial comparison against chance assuming a small effect using G*Power Version 3.1 ($g = .166$, $1\text{-}\beta = .8$, $\alpha = .05$, suggested sample size = 72). This sample was also sufficiently large to reliably detect a correlation between age and test trial performance, based on our exploratory study results (p H1 = .39, p H0 = 0, $1\text{-}\beta = .8$, $\alpha = .05$, suggested $n = 49$).

We recruited child participants from the greater Boston area through public birth records and family events. Participants were reported by their caregivers as female ($n = 37$) or male ($n = 37$). Further demographic information was collected through an optional demographic form. Among 27 returned forms, participants were identified by their caregivers as Asian (1), Asian/Black (1), Asian/Brazilian (1), Asian/Indian (1), Asian/White (1), and White (19). Two children were identified as Hispanic/Latinx. All families who completed the forms reported that at least one caregiver had a college degree or higher. The study procedures were approved by the Boston University Charles River Campus Institutional Review Board.

#### 2.1.2. Apparatus and stimuli

Children completed the study online via Zoom videoconferencing software (see Supplement for details on the Zoom setup and the devices families used in Experiments 1 and 2). The study stimuli were presented in Keynote presentation software by the experimenter using the screen-sharing function in Zoom. Stimuli for Experiments 1 and 2 are available at https://osf.io/yavqg/?view_only.

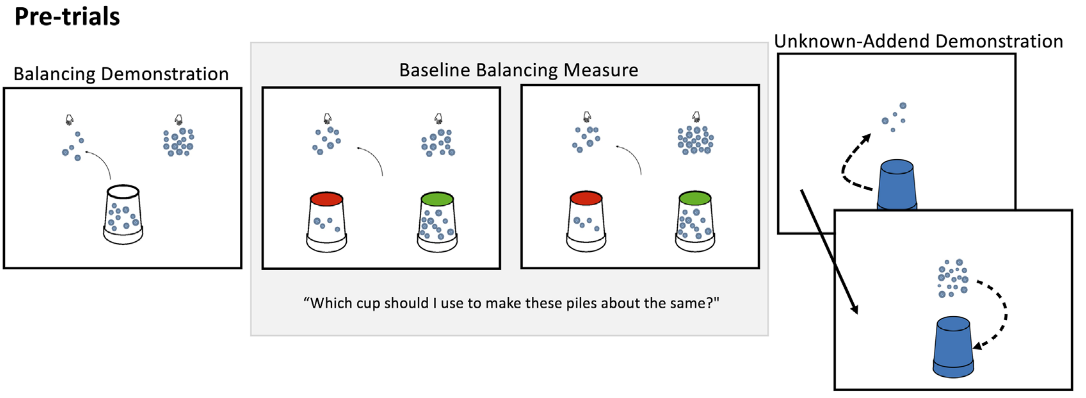


Fig. 1. Schematic of the four Pre-trials in Experiments 1 and 2.

### 2.1.3. *Procedure*

Children completed a series of four Pre-trials, two Unknown-Addend trials, a Test trial, and a series of Post-test Precision trials designed to assess the precision of children's representations of the solutions to the unknown-addend problems.

*2.1.3.1. Pre-trials:* The four Pre-trials (see Fig. 1) were designed to orient children to the kinds of events they would see in the experiment, and to gain a baseline measure of children's ability to perform a nonsymbolic balancing operation. The full details of these trials can be found in the Supplement. The first Pre-trial showed children that two unequal sets of buttons[2] could be balanced by adding the (visible) buttons contained inside a transparent cup. The second and third Pre-trials were Balancing Baseline trials: to obtain a baseline measure of children's ability to balance unequal sets, children were asked to choose which of two visible sets of buttons should be added to the smaller of two unequal sets of buttons to make them "about the same."[3] In the final Pre-trial, children were shown that an opaque cup could add buttons to an array. Children saw a set of buttons, and then saw an opaque cup move over the buttons, leaving behind its contents. We then highlighted the buttons that the cup left behind. Children were not asked to solve for the unknown addend in this trial; instead, this trial served to familiarize children with the idea that opaque cups could add more objects to sets of objects.

*2.1.3.2. Unknown-Addend trials:* The experimenter introduced children to two animated characters and their cups by saying "I want to introduce you to my friends. This is Gator [Gator jumped to attract attention] and his cup [the cup jiggled to attract attention]. He has some buttons inside his cup. This is Cheetah [Cheetah jumped to attract attention] and his cup [the cup jiggled to attract attention]. He also has some buttons inside his cup." Then, the experimenter said, "But I don't know how many buttons they have in each of their cups, can you help me figure it out?" Both characters and their cups were then removed from the screen.

The experimenter then proceeded to demonstrate Gator's and Cheetah's cups adding to sets of buttons. Whether Gator's cup or Cheetah's cup was demonstrated first was counterbalanced across children. Here, for convenience, we describe the condition in which Gator's cup was demonstrated first.

Gator and his cup first appeared on the lower left side of the screen. The experimenter said. "Here comes Gator again. He already has some buttons inside his cup [the cup jiggled]." A pile of nine buttons appeared in the center of the screen, and the experimenter said, "If I put a pile of buttons here, Gator's cup is going to come and add more buttons to this pile. Like this!" Gator's cup then completely covered the visible set, and moved back to its original position, revealing a set of 29 buttons. She then prompted children to examine the set by saying, "Did it work? See the buttons now?"

Gator and his cup remained on the lower left side of the screen, and Cheetah and his cup appeared. The experimenter said, "Now here comes Cheetah and his cup. He has also got some buttons inside his cup. But his cup is different from Gator's." A set of nine buttons appeared in the center of the screen, and the experimenter said, "If I put a pile of buttons here in the center of the screen, Cheetah's cup is going to come and add some more to this pile, and it's going to add a *different* number of buttons to this pile than Gator's cup just did. Like this!" Cheetah's cup then completely covered the visible set and then moved back to its original position, revealing a set of 17 buttons. The experimenter then prompted children to examine the set by saying, "Did it work? See the buttons now?" The visible set was then removed from the screen, while Gator, Cheetah, and their cups remained visible. Fig. 2 shows a schematic of the Unknown-Addend trials.

The quantities in the Unknown-Addend trials were chosen such that the final quantity (after the cup was added; 17 or 29) was sufficiently discriminable from the starting quantity (before the cup was added; 9) by at least a roughly 2:1 ratio (see Kibbe & Feigenson, 2015, 2017), to ensure that children could distinguish the difference between the two quantities. The variability of the sizes of the buttons and the pseudorandomized positions of the buttons within the arrays meant that it was difficult to determine at a glance, after the cup was added, which buttons were present before the addition and which were added. We randomized the positions of the buttons and limited the visible duration of each set of buttons (starting and final quantities) to roughly 3–4 s to make it very difficult for children to count or use other symbolic strategies.

*2.1.3.3. Test trial:* In the single Test trial, the experimenter said, "I want to show you some buttons." Two piles of buttons appeared sequentially at the top of the screen. The experimenter told children, "Here is one pile of buttons, and here is the other pile of buttons." Two animated hands then appeared pointing to the sets. The experimenter said, "See these piles of buttons? They have different numbers of buttons. But I want to make them about the same." Then, Gator and Cheetah jumped, and the experimenter said, "Whose cup do we have to use on this pile [an arrow pointed to the smaller set], so these two piles [the two animated hands pointing to the sets blinked] will be about the same?"

The smaller set always contained five buttons. Children were randomly assigned to one of the two Test conditions: a +Cheetah condition, in which the larger set contained 13 buttons,

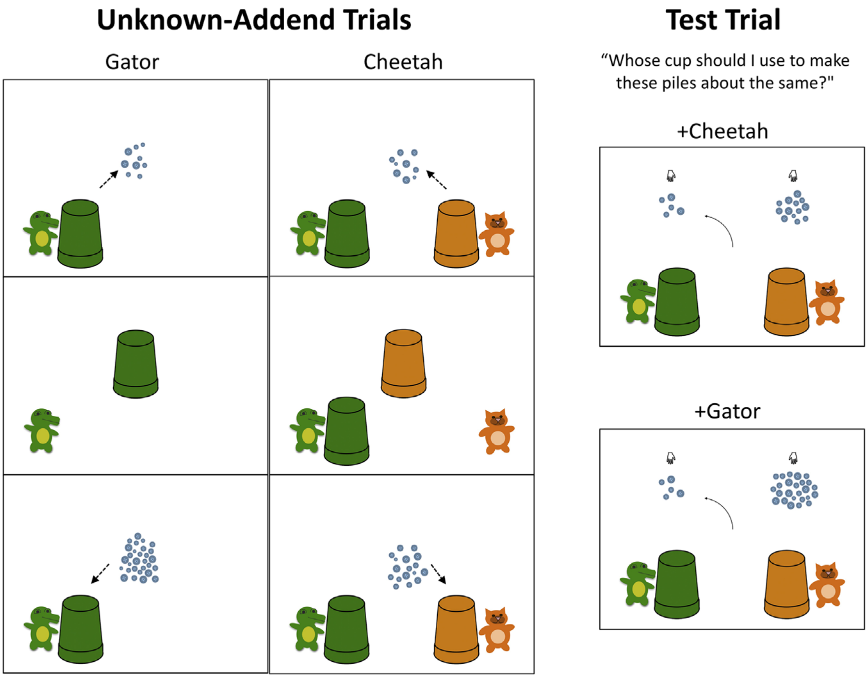


Fig. 2. Schematic of the Unknown-Addend trials (left panel) and Test trial (right panel).

or a +Gator condition, in which the larger set contained 25 buttons. Thus, whether the correct solution was the cup containing the larger quantity (+Gator) or the smaller quantity (+Cheetah) was counterbalanced across children. See Fig. 2 for a schematic of the Test trial.

The quantities in the Test trials (5 and 13 in the +Cheetah condition, or 5 and 25 in the +Gator condition) were chosen to allow us to detect whether children were using an alternative strategy that did not require attending to the components of the unknown-addend problems. Specifically, if children ignored the starting quantity and unknown addend action completely and attended only to the final quantities after the cups were added, and then attempted to use those final quantities as addends in the balancing problem, children in the +Gator condition would be more likely to choose Cheetah's cup (final quantity = 17, and 5+17 = 23, close to 25) than Gator's cup (final quantity = 29, and 5+29 = 34, larger than 25), and children in the +Cheetah would be more likely to choose at chance (since either final quantity, when added to the smaller set, would yield a quantity much larger than 13).

*2.1.3.4. Post-Test Precision trials:* After children completed the Test trial, they completed a series of 11 trials to assess the precision of their representations of the quantities in Gator's and Cheetah's cup (see Fig. 3). In the first Post-test trial, the experimenter showed both Gator, Cheetah and their cups and asked "Whose cup adds more buttons?" In the next 10 trials, the experimenter asked children to compare Gator's and Cheetah's cups to visible sets (five trials each, blocked). On each trial in the Gator block, the experimenter showed Gator

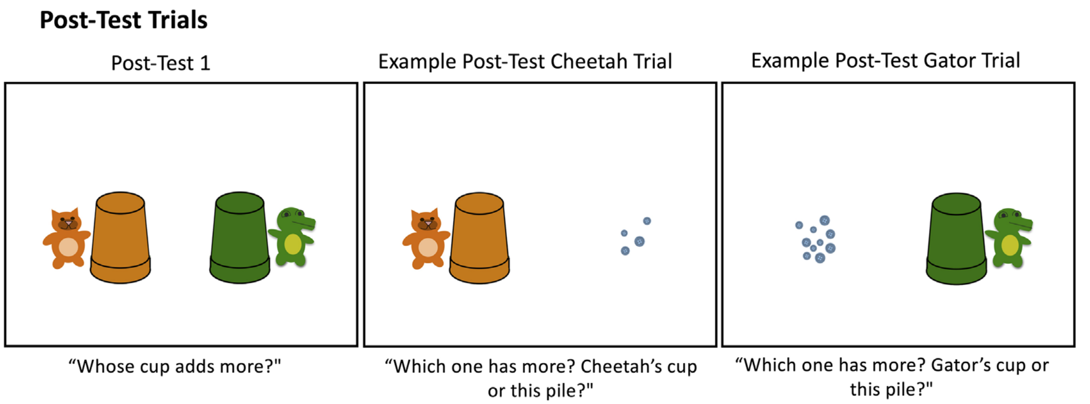


Fig. 3. Schematic of the first Post-Test trial, and schematic examples of Post-Test Precision trial comparisons for Cheetah's cup and Gator's cup. Children completed a total of 10 of these comparison trials (five for each cup).

and his cup on the left, a set of buttons on the right (either 10, 15, 25, 30, or 40 buttons) and asked, "Which one has more?" On each trial in the Cheetah block, the experimenter showed Cheetah and his cup on the right, a set of buttons on the left (either 4, 6, 12, 16, or 24 buttons), and asked "Which one has more?" The order of the blocks matched the order in which the cups were presented in the Unknown-Addend trials (e.g., if Cheetah's cup was demonstrated first in the Unknown-Addend trials, children completed the Cheetah block first in the Post-test).

### 2.2. Results

Data for Experiments 1 and 2 can be obtained at https://osf.io/yavqg/?view_only.

We found that 53/74 children (72%) correctly chose the cup that would balance the two sets in the Test trials (binomial test $p < .001$; $BF_{10} = 197.19$[4]). Children's pattern of responses was similar in the +Gator condition (25/36 [69%] chose correctly, binomial test $p = .029$, $BF_{10} = 4.07$) and the +Cheetah condition (28/38 [74%] chose correctly, binomial test $p = .005$, $BF_{10} = 16.47$), Fisher's exact test $p = .80$. While very few children older than 6 $\frac{1}{2}$ years responded incorrectly, a one-way ANOVA on children's ages with success on the Test trial as a factor was not statistically significant ($F(1, 72) = 3.81$, $p = .055$, $\eta^2 = .05$). These results are illustrated in Fig. 4, top panel.

We next investigated other potential contributors to children's success or failure in the Test trials. In the second and third Pre-test trials, children were asked to balance two unequal sets by selecting which of two visible quantities should be added to the smaller of the two sets. We asked whether children's ability to balance sets using fully visible quantities was related to their success in the Test trial. Across the two Pre-test Baseline Balancing trials, children chose the correct quantities to balance the two sets at rates significantly above chance (mean proportion correct = .68, SD = .36; Wilcoxon Signed Rank test asym. $p < .001$), and children who succeeded in the Test trial (M = .71, SD = .35) did not significantly differ from children who failed in the Test trial (M = .60, SD = .37) ($F(1, 72) = 1.51$ $p = .223$, $\eta^2 = .021$) in their

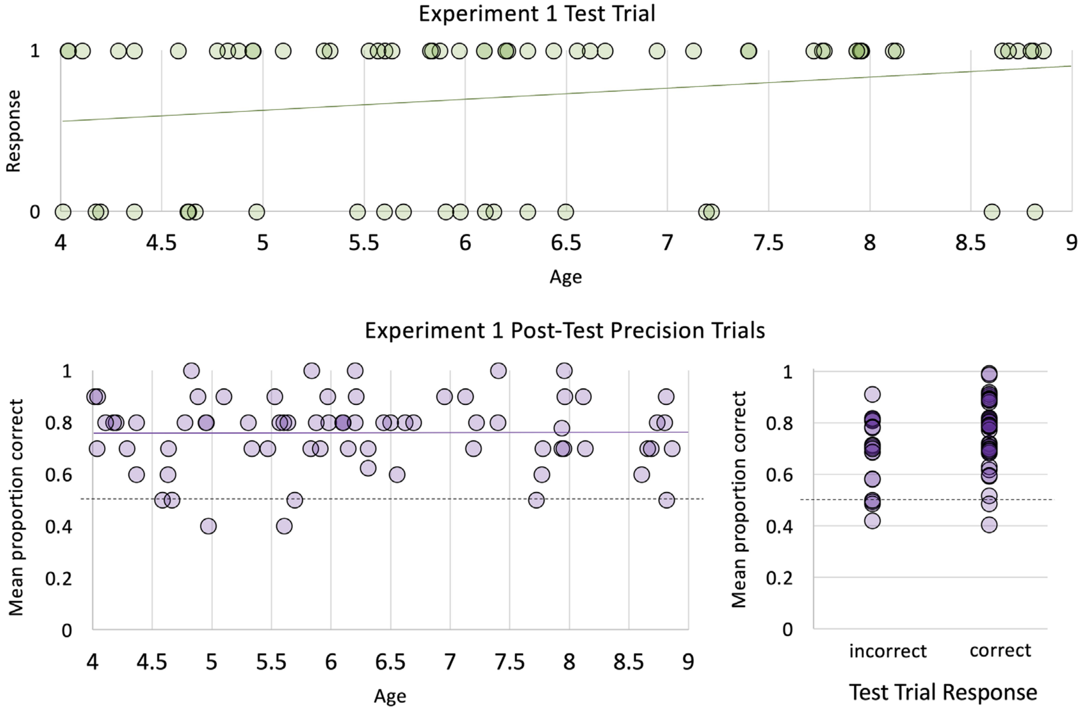


Fig. 4. The top panel shows individual children's responses on the Test trial as a function of age in Experiment 1. The bottom panels show individual children's mean proportion correct in the post-test precision trials as a function of age (left panel) and Test trial response (right panel) in Experiment 1. The solid lines represent the line of best fit, and the black dashed line represents chance-level performance (.5).

performance on the Pre-test Baseline Balancing trials, suggesting that children's Test trial performance was not significantly limited by their ability to execute the required balancing computation. Further, children were as successful in the Test trial (in which the balancing operation needed to be performed using derived solutions) as they were in the two Pre-test Balancing trials (in which the balancing operation could be performed using visible arrays; $\chi^2 = .57$, asym. $p = .751$).

We next examined the precision of children's representations of the solutions to the two unknown-addend problems, and whether the precision of their solutions related to their Test trial performance. In the first Post-test trial, children correctly responded that Gator's cup was larger than Cheetah's cup at rates significantly above chance (58/74 children, 78%, binomial $p < .001$, $BF_{10} = 30{,}978.46$). Analysis of children's responses to the remaining comparison trials (conducted on 735 total trials; five trials were excluded from analysis because children declined to respond on those trials) revealed that children correctly selected which was the larger quantity (the quantity in the cup or the visible quantity) at rates significantly above chance (mean proportion correct = .76, SD = .14, one-sample $t(73) = 16.30$, $p < .001$, $BF_{10} > 100{,}000$). Children's responses were significantly above chance both for Gator's addend (mean proportion correct = .80, SD = .17, one sample $t(73) = 15.25$, $p < .001$, $BF_{10} > 100{,}000$) and Cheetah's addend (mean proportion correct = .72, SD = .19, one sample $t(73) = 10.09$, $p < .001$, $BF_{10} > 100{,}000$), although their representations of Gator's addend were somewhat more precise than their representation of Cheetah's addend (paired samples $t(73) = -2.84$, $p = .006$), with Bayes factor yielding anecdotal support for the alternative ($BF_{10} = 3.99$). We speculate that this may be due to the fact that the final quantity after Gator's cup was added was more discriminable from the initial quantity than the final quantity after Cheetah's cup was added, which may have facilitated more precise representations of Gator's addend. Children's overall performance on Post-Test Precision trials was not correlated with age ($r = .01$, $p = .94$, $BF_{10} = .092$; see Fig. 4, bottom left panel).

To investigate whether the precision of children's representations of the addends impacted their Test trial performance, we conducted a repeated measures ANOVA on children's mean performance on the Post-Test Precision trials with Addend (Gator or Cheetah) as a within-participants factor and Test Success (succeed or fail) as a between-participants factor. We found a small main effect of Addend (reflecting children's slightly higher precision for Gator's addend; $F(1, 72 = 3.97$, $p = .050$; $\eta^2{}_p = .052$). We also observed a significant main effect of Test Success ($F(1, 72) = 6.97$, $p = .010$, $\eta^2{}_p = .088$) and no Test Success X Addend interaction ($F(1, 72) = 1.81$, $p = .183$, $\eta^2{}_p = .024$); children who represented the addends with higher precision were more successful in the Test trial. To investigate further, we ran a series of pairwise tests on each of the 10 Post-test Precision trials, comparing children who responded correctly versus incorrectly in the Test trial, with alpha set to .005 to correct for 10 comparisons. We found a difference only in the Post-Test Precision trial in which Gator's cup (quantity = 20) was compared to a set of 25 items (Mann−Whitney U = 342.50, $p = .003$; for the results of all comparisons, see Table S1 in the Supplement). On this trial, children who responded correctly in the Test trial also tended to respond correctly that the set of 25 was *more* than Gator's cup (mean proportion correct = .62). However, children who responded incorrectly in the Test trial were more likely to respond that the set of 25 was *less* than Gator's cup (mean proportion correct = .24), suggesting they *overestimated* the quantity in Gator's cup (although, critically, these same children still correctly responded that Gator's cup was *smaller than* the visible set of 30, mean proportion correct = .71, suggesting they were not likely to be simply using the final quantity after Gator's cup added (29) as their representation of Gator's quantity).

## 2.3. Discussion

The results of Experiment 1 suggest that children ages 4–8 were able to solve for two unknown addends in problems presented nonsymbolically, providing the first direct replication (to our knowledge) of Kibbe and Feigenson's (2015) Experiment 5, in which 4- to 6-year-old children solved for two unknown addends. Our study also extended this finding by providing the first data on the precision of children's representations of solved unknown addends. We found that their representations of those addends were overall fairly precise, and that they were able to compare their representations to a range of visible quantities with fairly high accuracy.

Further, children were successful at choosing which of the solved addends should be used to balance unequal sets, and children with more precise representations of the addends were overall more successful at doing so. These results suggest that children may be able to use the output(s) of nonsymbolic arithmetic computations as inputs into a new nonsymbolic arithmetic computation, and that this ability may be limited by the precision of the outputs.

Children were unlikely to be using symbolically mediated strategies to succeed in the task. First, children's solutions to the unknown-addend problems were noisy (as evidenced by their performance on the Post-test Precision trials), a key signature of the ANS. Second, children had limited time to view the sets, and did not show evidence of counting (i.e., children did not count out loud or point to the screen during the study), and previous work has suggested that children of this age struggle with the symbolic forms of unknown-addend problems while succeeding with nonsymbolic forms (e.g., Kibbe & Feigenson, 2015; Sherman & Bisanz, 2009; although we did not specifically test children's symbolic arithmetic abilities in our task). However, it is possible that children could have used the ANS to solve the unknown-addend problems in the Unknown-Addend trials, but then converted those to rough symbolic estimates (e.g., "Cheetah had about 8") and used those rough estimates in the Test trial to balance the unequal sets.

The design of Experiment 1 also leaves open another possible explanation of children's success at the balancing task in the Test trial. While children were indeed able to solve for the quantity inside each cup (as evidenced by their performance on the Post-Test Precision trials), children did not necessarily have to use those representations as inputs into a new arithmetic problem in the Test trial. Instead, children may have used a strategy of focusing only on the outcomes produced by the cups in the Unknown-Addend trials. Specifically, because each Unknown-Addend trial started with the same quantity (9), children observed that Cheetah's cup changed the starting quantity by a certain magnitude to the final quantity 17, and that Gator's cup changed the starting quantity by a different, larger magnitude to the final quantity 29. In the Test trial, children saw either 5 and 13 buttons, or 5 and 25 buttons, and had to decide whose cup should be added to the smaller quantity to make the sets about the same. To succeed, children could choose which of the two cups produced an outcome in the Unknown-Addend trials that was similar in quantity to the larger of two sets in the Test trial. Using this strategy, a child who is deciding which cup should be added to 5 to make it about the same as 13 would correctly choose Cheetah's cup, not because they added their representation of Cheetah's quantity to 5, but because the final quantity that Cheetah's cup produced in the Unknown-Addend trial (17) is closer to 13 than the final quantity that Gator's cup produced in the Unknown-Addend trial (29). Such a strategy would not, therefore, require children to operate directly with the solutions to the problems they solved in the Unknown-Addend trials.

In Experiment 2, we investigated this possibility. The overall structure of the task was similar, except that in Unknown-Addend trials, the starting quantities on each trial were chosen so that the final quantities after the cups were added would be the same. This meant that, while the cups continued to add different quantities, the outcomes of two Unknown-Addend trials were identical. If children can use the outputs of ANS computations as inputs into new computations, children should again succeed in the Test trial. However, if children in Experiment 1 were using a simpler strategy of using the outcomes of the Unknown-Addend trials to solve the Test trial, children should have more difficulty in Experiment 2, since the outcomes in both Unknown-Addend trials were the same. We tested a new sample of the oldest children in our age range (7- to 8-year-olds), since these children demonstrated the clearest success in Experiment 1.

## 3. Experiment 2

### 3.1. Participants

Participants were 52 7- to 8-year-old children (mean age = 8.01 years; range = 6.93–9.01 years). While the effect size for Test trial correct responses for 7- to 8-year-olds in Experiment 1 was fairly large (20/24 7- to 8-year-olds responded correctly in Experiment 1, $g = .33$), the sample size in Experiment 2 was powered to detect a more conservative effect ($g = .2$, alpha = .05, 1-beta = .8, suggested $n = 49$). Two additional children participated but were excluded from analyses due to experimenter error (1) or declining to complete all study procedures (1).

The recruitment procedure was the same as in Experiment 1. Participants were reported by their caregivers as female ($n = 26$) or male ($n = 26$). Thirty-six families completed the optional demographics form. Participants were identified by their caregivers as Asian (3), Asian/Black (1), Asian/Other (2), Asian/White (3), or White (26), and one declined to report. Two of those children identified as Hispanic/Latinx. All families who completed the forms reported that at least one caregiver in the household had a college degree or higher. The study procedures were approved by the Boston University Charles River Campus Institutional Review Board.

### 3.2. Apparatus, stimuli, and procedure

The apparatus, stimuli, and procedure were identical to Experiment 1, with the exception of the quantities used in the Unknown Addend trials. Specifically, the quantities added by Gator's cup (20) and Cheetah's cup (8) were the same as in Experiment 1, but unlike in Experiment 1 (in which the initial quantities in each demonstration were the same), in Experiment 2, the final quantities after the cups were added were the same. In the trial in which Gator's cup was demonstrated, children viewed a set of 10 buttons, saw Gator's cup move over the set, and then move back to its original position (leaving its buttons behind) to reveal the final set of 30 buttons. In the trial in which Cheetah's cup was demonstrated, children viewed a set of 22 buttons, saw Cheetah's cup move over the set, and then back to its original position (leaving its buttons behind) to reveal the final set of 30 buttons.

As in Experiment 1, children completed the four Pre-test trials, two Unknown-Addend trials, the Test trial, and then the series of 11 Post-Test Precision trials.

### 3.3. Results

We found that 31/52 children (60%) chose correctly in the Test trial, not significantly different from chance ($p = .212$), although the Bayes factor yielded only anecdotal support for the null hypothesis ($BF_{10} = .77$). Children's responses were similar in the +Gator condition

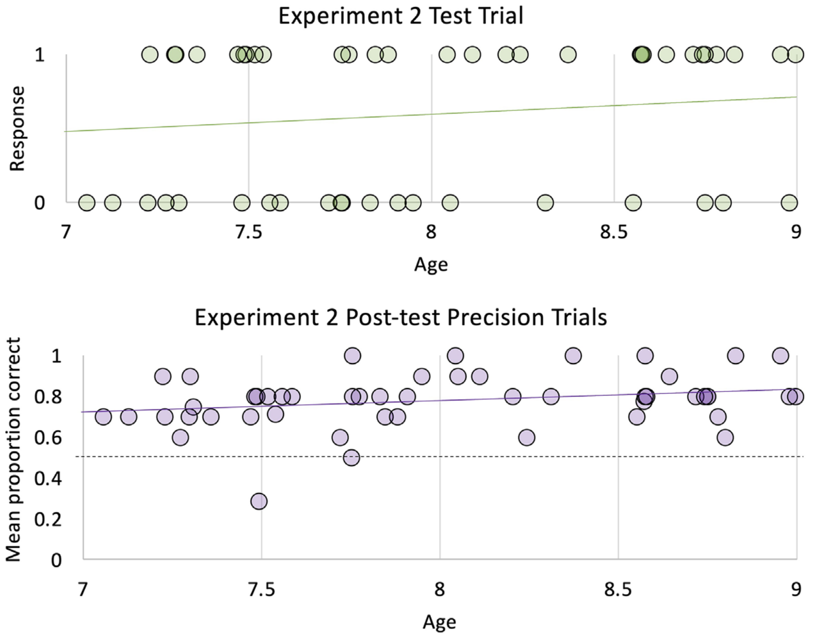


Fig. 5. The top panel shows individual children's responses on the Test trial as a function of age in Experiment 2. The bottom panels show individual children's mean proportion correct in the post-test precision trials as a function of age (left panel) in Experiment 2. The solid lines represent the line of best fit, and the black dashed line represents chance-level performance (.5).

(15/26 correct) and the +Cheetah condition (16/26 correct), Fisher's exact test $p = 1.0$. Children who succeeded in the Test trial were not significantly different in age than children who failed in the Test trial (succeeded in Test: M = 8.08 years, SD = .61; failed in Test: M = 7.90 years, SD = .61; ($F(1, 50) = 1.07$, $p = .307$, $\eta^2 = .021$). These results are summarized in Fig. 5.

Children overall were successful in the two Pre-trial Baseline Balancing trials, in which they were asked to choose which of two visible quantities should be used to balance unequal sets (mean proportion correct = .85, SD = .23; Wilcoxon Signed Rank test asym. $p < .001$), and children performed significantly worse in Test trials compared to the two Balancing Baseline trials ($\chi^2(2) = 11.47$, asym. $p = .003$), suggesting that they struggled more with the balancing operation when they were required to use derived solutions from two unknown-addend problems with equal final quantities.

We next examined the precision of children's representations of the unknown addends. Children successfully identified Gator's addend as larger than Cheetah's addend at rates significantly above chance (35/52 [67%] chose correctly, binomial $p = .018$; $BF_{10} = 5.48$). While an inspection of Fig. 5 shows a slight increase in precision with age, the correlation between age and overall precision was not statistically significant ($r = .26$, $p = .065$, $BF_{10} = .59$). Analysis of children's responses to the 10 comparison trials (conducted on 511 total trials; nine trials were excluded from analysis because children declined to respond on those trials) revealed that children overall correctly selected which was the larger quantity (the quantity in the cup or the visible quantity) at rates significantly above chance (mean proportion correct = .78, SD = .13, one sample $t(51) = 15.06$, $p < .001$, $BF_{10} > 100{,}000$; Fig. 5). Similar to Experiment 1, children's responses were significantly above chance both for Gator's addend (mean proportion correct = .82, SD = .16, one sample $t(51) = 14.85$, $p < .001$, $BF_{10} > 100{,}000$) and Cheetah's addend (mean proportion correct = .74, SD = .20, one sample $t(51) = 8.64$, $p < .001$, $BF_{10} > 100{,}000$), and children's representations of Gator's addend were slightly more precise than their representation of Cheetah's addend (paired samples $t(51) = -2.40$, $p = .02$), with Bayes factor again yielding only anecdotal support for the alternative ($BF_{10} = 1.58$).

A repeated measures ANOVA with Addend (Gator or Cheetah) as a within-participants factor and Test Success (succeed or fail) as a between-participants factor revealed a main effect of Addend (reflecting the slightly higher precision for Gator's addend vs. Cheetah's; $F(1, 50) = 6.89$, $p = .011$, $\eta^2_p = .121$), but no main effect of Test Success ($F(1, 50) = .70$, $p = .407$, $\eta^2_p = .014$) and no Test Success X Addend interaction ($F(1, 50) = 1.74$, $p = .193$, $\eta^2_p = .034$). The precision of children's representations of the unknown addends was not significantly different between children who succeeded versus failed in the Test trial.

### *3.3.1. Experiments 1 and 2 compared*

We compared children's data in Experiment 2 to the 7- to 8-year-olds' data ($n = 24$) from Experiment 1. Children were more successful in the Test trial in Experiment 1 (20/24 [83.33%] chose correctly) than in Experiment 2 ($\chi^2(1) = 4.19$, $p = .041$). Crucially, there was no significant difference in the precision of children's representations overall ($t(74) = .66$, $p = .509$, $BF_{10} = .22$), and no difference in the precision of their representations of Gator's addend ($t(74) = .38$, $p = .703$, $BF_{10} = .20$) or Cheetah's addend ($t(74) = .67$, $p = .506$, $BF_{10} = .23$) specifically, suggesting that the differences in Test trial responses across the two experiments are unlikely to be due to differences in precision of the solutions to the unknown addends.

## *3.4. Discussion*

Experiment 2 was designed to rule out a potential alternative strategy that children could use to succeed in the Test trial in Experiment 1 that did not require children to directly compute with the outputs of ANS operations. To that end, the only difference between Experiments 1 and 2 was the initial and final quantities used in the Unknown-Addend trials; the Pre-Test trials, the quantities of the unknown-addends, the Test trial, and the Post-Test precision trials were identical, and thus the cognitive demands of the task were equated between the two experiments. The results from Experiment 2 suggest that when the starting quantities in the two unknown-addend problems were different and the final quantities (after the cup was added) were the same, children still could successfully solve for both unknown addends. This result extends previous work (Kibbe & Feigenson, 2015) to show that children can solve for unknown addends in nonsymbolic problems, and can represent both with high fidelity and compare their magnitudes to visible quantities, even when the problems included different quantities as their starting addend. However, children were not readily able to select which addend should be used to balance unequal sets. Children's failure in the Test trial did not stem from difficulty with the balancing operation itself, since children were able to successfully choose which of two visible quantities should be used to balance unequal sets in the Pre-test Balancing trials. Children's failure in the Test trial also was not likely to be due to the precision of their representations of the unknown addends, since children represented the unknown addends with similar precision across Experiments 1 and 2. Further, these results suggest that children were not likely to be using symbolically mediated strategies in Experiment 1, converting the unknown-addends to symbolic estimates (e.g., "about 20") and using those estimates to balance the unequal sets, because such a strategy would also have led to success in Experiment 2.

Instead, these results suggest that children in Experiment 1 may have been relying on a strategy of focusing on the outcome of the unknown-addend operations (i.e., "how much did each cup make the starting quantity increase?") to decide which of the two cups should be used to balance the unequal sets, a strategy that was not possible in Experiment 2.

The Test trial in both Experiments 1 and 2 made significant demands on working memory: children had to hold the solutions to the unknown-addend problems in working memory, visually examine the two visible sets, and then perform one or more mental operations to decide which solution to choose to balance the two sets. In the Pre-Test Balancing trial (in which children succeeded), children did not have to hold as much information in working memory because the quantities that could be used to balance the sets were visible. Could children's failure in Experiment 2 be due to working memory limitations? There are several reasons to think not. First, children succeeded in the Test trial of Experiment 1, which made similar demands on working memory as the Test trial in Experiment 2. Even if children were using an alternative strategy in Experiment 1, that would still require children to hold two representations in mind (e.g., representations of the outcomes of the different Unknown-Addend demonstrations) and deploy them correctly to balance the two sets. Second, previous research has shown that children of the ages tested in Experiment 2 have sufficient working memory capacity to store at least four visual representations in working memory and to actively manipulate those representations (i.e., updating the remembered location of an object following a real-world change to its location; Cheng & Kibbe, 2022; Pailian et al., 2016). Furthermore, we have some evidence that children in our study likely had sufficient working memory capacity: a large subset of the children from Experiment 1 ($n = 64$ 4- to 7-year-olds) also completed a separate study after Experiment 1 that was designed to examine working memory strategy use. While these studies were not designed a priori to be compared, we were able to conduct an exploratory analysis examining children's working memory in relation to their Test trial performance in Experiment 1. The details of these exploratory results can be found in the Supplement. We found no significant differences in children's working memory performance between children who succeeded versus failed in the Test trial, and children's working memory capacity was overall quite high. Although children in Experiment 2 did not complete a working memory study, these children were older than the children in Experiment 1 and are, therefore, likely to have as much or more working memory capacity at their disposal (c.f., Cheng & Kibbe, 2022; Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Cowan et al., 2010;

Cowan, AuBuchon, Gilchrist, Ricker, & Saults, 2011; Pailian et al., 2016). Combined with the fact that the only difference we found between 7- and 8-year-olds children's performance in Experiments 1 and 2 was in the Test trial (and not in their ability to solve the Pre-Test Balancing trial, nor in the precision of their representations of the solved addends), we think that working memory limitations are unlikely to be the source of children's failure in the Test trials.

We discuss the implications for these results in the General Discussion.

## 4. General discussion

Previous work suggested that children's early nonsymbolic capacities may bear a surface resemblance to symbolic arithmetic, but the algorithms that support nonsymbolic ANS computation were less clear. To gain insights into the operational structure and computational capacity of our early quantificational abilities, we asked about the algorithmic structure of nonsymbolic arithmetic by examining whether it operates with function rules like symbolic arithmetic. In symbolic arithmetic, operations such as addition take as input independent numerals and produce as output another independent numeral. Since arithmetic operations can take as input any numeral, the solutions to arithmetic operations can serve as operands in new problems, making function arithmetic combinatorially powerful. Here, we asked whether nonsymbolic arithmetic computations obey such rules by testing children who have limited knowledge of the formal rules of symbolic arithmetic. In two experiments, children solved for unknown addends in two nonsymbolic arithmetic problems and were then asked to use those solutions as inputs into a balancing operation. Solving for an unknown addend requires the combination of two independent ANS representations to derive a previously unobserved solution, and cannot be accomplished by manipulating a single ANS representation, making it a strong test case to address whether computationally derived ANS representations can be used as inputs into further computations.

In Experiment 1, each of the unknown addends (Cheetah's and Gator's cups) were added to the same quantity (nine buttons) resulting in different final quantities (17 or 29 buttons). In Experiment 2, each of the unknown addends were added to different quantities (22 or 10 buttons) resulting in the same final quantity (30 buttons). To examine whether children could use the solutions to the addend-unknown problems in new computations, we showed children two unequal sets of objects, and asked them to choose whether Cheetah's or Gator's buttons should be added to the smaller set to make it about the same quantity as the larger set. We also measured children's ability to balance unequal sets using visible quantities and measured the precision of children's representations of the solutions to unknown addend problems.

We hypothesized that, if nonsymbolic arithmetic obeys similar function rules to symbolic arithmetic, then the outputs of nonsymbolic arithmetic computations (the solved addends) should be able to be used as inputs into another nonsymbolic arithmetic computation (the balancing operation). Contrary to this hypothesis, our results suggested that the solutions to nonsymbolic arithmetic computations may not operate as independent representations and may not readily be used as inputs into a new nonsymbolic computation. Children were able to select which cup should be used to balance the unequal sets in Experiment 1, but did not robustly do so in Experiment 2.

The contrast in results between Experiments 1 and 2 suggests that children in Experiment 1 were using a strategy in the Test trial that did not require them to operate directly with their solutions to the unknown addend problems. When the problems started with the same quantity, children could use a strategy of focusing on the final quantities in the unknown-addend problems to try to produce the correct outcome on the smaller set in the Test trial. That is, children could recall the degree to which Cheetah's and Gator's cup increased a single quantity (i.e., Cheetah's cup increased the initial quantity to ∼17, while Gator's cup increased the initial quantity to ∼29) and could then use that information in the Test trial to choose which cup might produce an outcome on the smaller set similar in quantity to the larger set. In Experiment 2, children could not use that strategy because the final quantities were identical. Indeed, even though children in both experiments could (1) balance unequal sets by selecting from two *visible* quantities, (2) estimate the quantity added by each cup in both experiments, and (3) maintain those representations in memory long enough to respond in the Post-test Precision trials, they did not readily use those representations directly to solve a new problem in the Test trial.

These results suggest that, while nonsymbolic arithmetic computations with ANS representations may bear a surface resemblance to symbolic arithmetic, they may be algorithmically distinct. Specifically, the outputs of nonsymbolic arithmetic computations may not be readily manipulated independently of the computational context in which they arose. This has potential implications for the format of representations that arise from *combinations of independent ANS representations*, as in the unknown-addend problems tested here. Our results suggest that computationally derived representations (i.e., representations that arise from the mental combination of ANS representations) are noisy, like the visually derived ANS representations (i.e., the starting quantity and the final quantity) that they arise from. Yet, these computationally derived ANS representations may not behave like visually derived ANS representations in computational contexts. This also means that nonsymbolic arithmetic computations may not be indifferent to the source of their inputs, and may be limited to operating over visually derived representations.

However, despite these limitations, children appear to be flexible in their ability to make use of past computations to help them solve a new problem, even if they may not readily use the outputs of those computations directly. Children in Experiment 1 were able to use what they observed in the Unknown-Addend trials to help them figure out how to solve the balancing operation in the Test trial. These results suggest that ANS computations are not completely quarantined from other computations, but rather may be drawn upon when the context of the new computation is similar enough.

Our results also contribute to our understanding of children's early capacity to solve unknown-addend problems (the symbolic forms of which are notoriously difficult for both children and adults; e.g., Booth, 1988; Kieran, 1992; Filloy & Rojano, 1989; Koedinger, Alibali & Nathan, 2008; Riley & Greeno, 1988; Tabachneck, Koedinger & Nathan, 1995). In both experiments, we successfully replicated and extended Kibbe and Feigenson's (2015, Experiment 5) finding that young children could concurrently solve two unknown-addend problems presented nonsymbolically. It is crucial to note that our method diverged from theirs in a few respects. First, we gave children only one demonstration of each unknown addend operating on a set, while Kibbe and Feigenson (2015) gave children three demonstrations of the unknown addend adding to different quantities. Our results suggest that children can solve "one-shot" nonsymbolic unknown-addend problems, just as previous work showed children can solve one-shot nonsymbolic addition, subtraction, or scaling problems (e.g., Barth et al., 2006; McCrink & Spelke, 2010), and imply a robust and stable operational capacity to solve unknown-addend problems in children. Second, our measure of children's success at solving for unknown addends diverged from Kibbe and Feigenson (2015). In their study, to test whether children had solved for the unknown addends, they showed children a single quantity and asked them to identify which addend it was, which did not allow them to examine the precision with which the solutions were represented. Because we asked children to compare their representations of the solved addends to a range of comparison quantities, we were able to observe that children represented these solutions with fairly high fidelity. Finally, our experiments included much larger samples than in Kibbe and Feigenson (2015; $n = 24$ 4- to 6-year-olds in Experiment 5). Like them, we also observed little developmental improvement in the ability to solve for unknown addends, although there was a slight suggestive trend in Experiment 2. Together with this previous work, our results shed additional light on children's capacity for solving nonsymbolic unknown-addend problems.

Our results also have potential implications for the extent to which the ANS can be leveraged to support the acquisition of formal arithmetic competence. While previous work has suggested a relationship between the precision of the ANS and some symbolic mathematics outcomes (e.g., Libertus et al., 2012; Mazzocco, Feigenson, & Halberda, 2011b; Wang et al., 2016), work that has explored the relationship between nonsymbolic and symbolic arithmetic competence has yielded mixed results (Budgen et al., 2021; Park & Brannon, 2013; Szkudlarek et al., 2016). Our results suggest that nonsymbolic arithmetic computations with the ANS may be algorithmically distinct from symbolic arithmetic computations, which may limit the extent to which children can draw on their early nonsymbolic intuitions to help them master the rules of symbolic arithmetic. However, we speculate that uncovering the algorithmic structure of the ANS can provide additional entry points for scholars who study the relationships between nonsymbolic and symbolic math. Further work is needed to better understand whether and to what extent the computational limits on nonsymbolic arithmetic impact (for better or for worse) the acquisition of formal arithmetic.

Our studies also have some important limitations. By design, we specifically looked at children's ability to use the outputs of unknown-addend operations as inputs into a completely different computation (a balancing operation). This design allowed for the strongest test case for our questions of interest since (1) unknown-addend operations require the combination of two independent ANS representations to derive the solution, and therefore, produce computationally derived ANS representations as outputs, and (2) children had to apply these outputs into a completely different operation, allowing us to examine the computational independence of the outputs of ANS operations. It is possible that the ANS may have some functional capacity, albeit more limited than the functional computational range of true arithmetic. For example, children may be able to use the outputs of unknown addend operations as the starting addend in another unknown-addend operation, leveraging the same computational context in order to do so. It is also possible that, with age and/or education, the computational range of the ANS expands. For example, adult observers may be effectively able to operate over computationally derived ANS representations in a variety of contexts. Further work should examine the algorithmic structure of a full range of ANS computations (including function-like operations like unknown-addend problems as well as other operations like summation or scaling), and their potential limits, across the lifespan. Finally, while we showed that children may not spontaneously use the outputs of ANS computations into further computations (suggesting an algorithmic limit on the ANS), our studies leave open the possibility that children may be able to *learn* to do so. An important question for future research concerns whether the limits on the computational ANS are fixed, or whether the computational structure of the ANS may be responsive to instruction or training.

## Acknowledgments

This research was supported by a National Science Foundation (BCS 1844155) grant awarded to M.M.K. We would like to thank NSF Living Laboratories at the Museum of Science Boston, Michaela Buckley, Keara Russell, Jiaqi Zhao, and Shiba Esfand for their assistance in data collection, and Derek E. Anderson and Aimee E. Stahl for helpful discussions.

## Open Research Badges

This article has earned Open Data and Open Materials badges. Data are available at https://osf.io/yavqg/ and materials are available at https://osf.io/yavqg/.

## Notes

1 We use the term “nonsymbolic” throughout to refer to ANS representations, to contrast these with formal, learned symbolic representations of quantity, such as number words or Arabic digits.
2 To highlight that the quantity of the sets (and not their continuous extent or density) was the crucial dimension of the stimuli, we varied the size of the buttons within each set, and used quantity-focused language (e.g., “how many”) throughout the experiment.
3 The experimenter described the sets as “about the same” rather than “equal,” since children of this age often have difficulty understanding the concept of mathematical equivalence (see McNeil, 2007) and since determining “equivalence” using the inherently noisy ANS may be impossible.
4 All Bayes factors were computed using the Jeffreys−Zellner−Siow prior and two-tailed test.

## References


Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*(3), 199–222. https://doi.org/10.1016/j.cognition.2004.09.011

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*(4), 1016–1031. https://doi.org/10.1111/j.1467-8624.2008.01173.x

Bugden, S., Szkudlarek, E., & Brannon, E. M. (2021). Approximate arithmetic training does not improve symbolic math in third and fourth grade children. *Trends in Neuroscience and Education*, *22*, 100149.

Bulthé, J., De Smedt, B., & de Beeck, H. O. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage*, *87*, 311–322.

Butterworth, B., & Walsh, V. (2011). Neural basis of mathematical cognition. *Current Biology*, *21*(16), R618-R621.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenisis. *Mind & Language*, *16*(1), 37–55.

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016

Cheng, C., & Kibbe, M. M. (2022). Development of updating in working memory in 4–7-year-old children. *Developmental Psychology*, *58*(5), 902–912. https://doi.org/10.1037/dev0001337

Christodoulou, J., Lac, A., & Moore, D. S. (2017). Babies and math: A meta-analysis of infants' simple arithmetic competence. *Developmental Psychology*, *53*(8), 1405. https://doi.org/10.1037/dev0000330

Chu, F. W., van Marle, K., & Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology*, *132*, 205–212. https://doi.org/10.1016/j.jecp.2015.01.006

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. *Developmental Science*, *17*(1), 11–22. https://doi.org/10.1111/desc.12108

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. S. (2011). Age differences in visual working memory capacity: Not based on encoding limitations. *Developmental Science*, *14*(5), 1066–1074. https://doi.org/10.1111/j.1467-7687.2011.01060.x

Cowan, N., Hismjatullina, A., AuBuchon, A. M., Saults, J. S., Horton, N., Leadbitter, K., & Towse, J. (2010). With development, list recall includes more chunks, not just larger ones. *Developmental Psychology*, *46*(5), 1119.

Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: When is feature binding difficult? *Developmental Psychology*, *42*(6), 1089–1102. https://doi.org/10.1037/0012-1649.42.6.1089

Dedekind, R., & Beman, W. W. (1901). *Essays on the theory of numbers: I. Continuity and irrational numbers, II. The nature and meaning of number*. Open Court Publishing Company.

Dehaene, S. (1997). *The Number Sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3-6), 487–506. https://doi.org/10.1080/02643290244000239

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Frontiers in Human Neuroscience*, *6*, 68. https://doi.org/10.3389/fnhum.2012.00068

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences*, *111*(13), 4809–4813. https://doi.org/10.1073/pnas.1323628111

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72. https://doi.org/10.1016/j.jecp.2014.01.013

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

Filloy, E., & Rojano, T. (1989). Solving equations: the transition from arithmetic to algebra. *For the Learning of Mathematics*, *9*(2), 19–25. https://www.jstor.org/stable/40247950

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1), 43–74. https://doi.org/10.1016/0010-0277(92)90050-R

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65. https://doi.org/10.1016/S1364-6613(99)01424-2

Gilmore, C. K., McCarthy, S. E., & Spelke, E. Geary, D. C., Hoard, S. (2007). Symbolic arithmetic knowledge without instruction. *Nature*, *447*(7144), 589–591. https://doi.org/10.1038/nature05850

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*(3), 394–406. https://doi.org/10.1016/j.cognition.2010.02.002

Gilmore, C. K., & Spelke, E. S. (2008). Children's understanding of the relationship between addition and subtraction. *Cognition*, *107*(3), 932–945. https://doi.org/10.1016/j.cognition.2007.12.007

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

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668. https://doi.org/10.1038/nature07246

Inglis, M., & Gilmore, C. (2013). Sampling from the mental number line: How are approximate number system representations formed? *Cognition*, *129*(1), 63–69. https://doi.org/10.1016/j.actpsy.2013.11.009

Kadosh, R. C., Bahrami, B., Walsh, V., Butterworth, B., Popescu, T., & Price, C. J. (2011). Specialization in the human brain: the case of numbers. *Frontiers in Human Neuroscience*, *5*, 62. https://doi.org/10.3389/fnhum.2011.00062

Kibbe, M. M., & Feigenson, L. (2015). Young children 'solve for x' using the Approximate Number System. *Developmental Science*, *18*(1), 38–49. https://doi.org/10.1111/desc.12177

Kibbe, M. M., & Feigenson, L. (2017). A dissociation between small and large numbers in young children's ability to "solve for x" in non-symbolic math problems. *Cognition*, *160*, 82–90. https://doi.org/10.1016/j.cognition.2016.12.006

Kieran, C. (1992). The learning and teaching of school algebra. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). New York: Macmillan.

Koedinger, K.R., Alibali, M.W., & Nathan, M.J. (2008). Trade-offs between grounded and abstract representations: evidence from algebra problem solving. *Cognitive Science*, *32*(2), 366–397. https://doi.org/10.1080/03640210701863933

Libertus, M. E., & Brannon, E. M. (2009). Behavioral and neural basis of number sense in infancy. *Current Directions in Psychological Science*, *18*(6), 346–351. https://doi.org/10.1111/j.1467-8721.2009.01665.x

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, *141*(3), 373–379. https://doi.org/10.1016/j.actpsy.2012.09.009

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense large-number discrimination in human infants. *Psychological Science*, *14*(5), 396–401. https://doi.org/10.1111/1467-9280.01453

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences*, *109*(46), 18737–18742. https://doi.org/10.1073/pnas.1207212109

Lussier, C. A., & Cantlon, J. F. (2017). Developmental bias for number words in the intraparietal sulcus. *Developmental Science*, *20*(3), e12385. https://doi.org/10.1111/desc.12385

Lyons, I. M., Nuerk, H.-C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology: General*, *144*(5), 1021–1035. https://doi.org/10.1037/xge0000094

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*(4), 1224–1237. https://doi.org/10.1111/j.1467-8624.2011.01608.x

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS One*, *6*(9), e23749. https://doi.org/10.1371/journal.pone.0023749

McCrink, K., Shafto, P., & Barth, H. (2017). The relationship between non-symbolic multiplication and division in childhood. *Quarterly Journal of Experimental Psychology*, *70*(4), 686–702. https://doi.org/10.1080/17470218.2016.1151060

McCrink, K., & Spelke, E. S. (2010). Core multiplication in childhood. *Cognition*, *116*(2), 204–216. https://doi.org/10.1016/j.cognition.2010.05.003

McCrink, K., & Spelke, E. S. (2016). Non-symbolic division in childhood. *Journal of Experimental Child Psychology*, *142*, 66–82. https://doi.org/10.1016/j.jecp.2015.09.015

McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science*, *15*(11), 776–781. https://doi.org/10.1111/j.0956-7976.2004.00755.x

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science*, *18*(8), 740–745. https://doi.org/10.1111/j.1467-9280.2007.01969.x

McCrink, K., & Wynn, K. (2009). Operational momentum in large-number addition and subtraction by 9-month-olds. *Journal of Experimental Child Psychology*, *103*(4), 400–408. https://doi.org/10.1016/j.jecp.2009.01.013

McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology*, *43*(3), 687. https://doi.org/10.1037/0012-1649.43.3.687

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320. https://doi.org/10.1037/0097-7403.9.3.320

Pailian, H., Libertus, M. E., Feigenson, L., & Halberda, J. (2016). Visual working memory capacity increases between ages 3 and 8 years, controlling for gains in attention, perception, and executive control. *Attention, Perception, & Psychophysics*, *78*, 1556–1573. https://doi.org/10.3758/s13414-016-1140-5

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, *24*(10), 2013–2019. https://doi.org/10.1177/0956797613482944

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child Psychology*, *152*, 278–293. https://doi.org/10.1016/j.jecp.2016.07.011

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41. https://doi.org/10.1016/j.cognition.2010.03.012

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*(2), 293–305. https://doi.org/10.1016/j.neuron.2006.11.022

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

Riley, M.S., & Greeno, J.G. (1988). Development of children's problem-solving ability in arithmetic. In H. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–200). New York: Academic Press.

Sherman, J., & Bisanz, J. (2009). Equivalence in symbolic and nonsymbolic contexts: Benefits of solving problems with manipulatives. *Journal of Educational Psychology*, *101*(1), 88. https://doi.org/10.1037/a0013156

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences*, *110*(45), 18116–18120. https://doi.org/10.1073/pnas.1302751110

Szkudlarek, E., Park, J., & Brannon, E. M. (2021). Failure to replicate the benefit of approximate arithmetic training for symbolic arithmetic fluency in adults. *Cognition*, *207*, 104521.

Tabachneck, H.T., Koedinger, K., & Nathan, M.J. (1995). A cognitive analysis of the task demands of early algebra. In Proceedings of the seventeenth annual conference of the Cognitive Science Society (pp. 397–402).

Wang, J., Halberda, J., & Feigenson, L. (2021). Emergence of the link between the approximate number system and symbolic math ability. *Child Development*, *92*(2), e186-e200. https://doi.org/10.1111/cdev.13454

Wang, J. J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, *147*, 82–99. https://doi.org/10.1016/j.jecp.2016.03.002

Wang, J. J., Odic, D., Halberda, J., & Feigenson, L. (2017). Better together: Multiple lines of evidence for a link between approximate and exact number representations: A reply to Merkley, Matejko, and Ansari. *Journal of Experimental Child Psychology*, *153*, 168–172. https://doi.org/10.1016/j.jecp.2016.09.005

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

Wong, T. T. Y. (2020). Are the acuities of magnitude representations of different types and ranges of numbers related? Testing the core assumption of the integrated theory of numerical development. *Cognitive Development*, *54*, 100888. https://doi.org/10.1016/j.cogdev.2020.100888

Wynn, K. (1992). Evidence against empiricist accounts of the origins of numerical knowledge. *Mind & Language*, *7*(4), 315–332.

Wynn, K. (1995). Infants possess a system of numerical knowledge. *Current Directions in Psychological Science*, *4*(6), 172–177. https://doi.org/10.1111/1467-8721.ep10772615

Wynn, K. (1998). Psychological foundations of number: Numerical competence in human infants. *Trends in Cognitive Sciences*, *2*(8), 296–303. https://doi.org/10.1016/S1364-6613(98)01203-0

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. https://doi.org/10.1016/S0010-0277(99)00066-9

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, *89*(1), B15–B25. https://doi.org/10.1016/S0010-0277(03)00050-7

Yang, X., Zhang, X., Huo, S., & Zhang, Y. (2020). Differential contributions of cognitive precursors to symbolic versus non-symbolic numeracy in young Chinese children. *Early Childhood Research Quarterly*, *53*, 208–216. https://doi.org/10.1016/j.ecresq.2020.04.003

Zhang, X., & Lin, D. (2015). Pathways to arithmetic: The role of visual-spatial and language skills in written arithmetic, arithmetic word problems, and nonsymbolic arithmetic. *Contemporary Educational Psychology*, *41*, 188–197. https://doi.org/10.1016/j.cedpsych.2015.01.005

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information