Available online at www.sciencedirect.com
SCIENCE DIRECT
COGNITION
Cognition 98 (2006) 199–222
www.elsevier.com/locate/COGNIT

Non-symbolic arithmetic in adults and young children

Hilary Barth[a,*], Kristen La Mont[a], Jennifer Lipton[a], Stanislas Dehaene[b], Nancy Kanwisher[c], Elizabeth Spelke[a]

[a] *Laboratory for Developmental Studies, Harvard University, 33 Kirkland St., Cambridge, MA 02138, USA*
[b] *Unité INSERM 562, Cognitive Neuroimaging, Service Hospitalier Frederic Joliot, CEA/DRM/DSV, 4 Place du general Leclerc, 91401 Orsay Cedex, France*
[c] *Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, Cambridge, MA, USA*

Received 30 July 2004; accepted 13 September 2004

Abstract

Five experiments investigated whether adults and preschool children can perform simple arithmetic calculations on non-symbolic numerosities. Previous research has demonstrated that human adults, human infants, and non-human animals can process numerical quantities through approximate representations of their magnitudes. Here we consider whether these non-symbolic numerical representations might serve as a building block of uniquely human, learned mathematics. Both adults and children with no training in arithmetic successfully performed approximate arithmetic on large sets of elements. Success at these tasks did not depend on non-numerical continuous quantities, modality-specific quantity information, the adoption of alternative non-arithmetic strategies, or learned symbolic arithmetic knowledge. Abstract numerical quantity representations therefore are computationally functional and may provide a foundation for formal mathematics.

Keywords: Symbolic arithmetic; Crossmodal addition; Non-symbolic numerosities

Numerical abilities observed in animals, human infants, and human adults share many characteristics, including a signature ratio limit on discriminability (Brannon & Terrace, 2002; Brannon, Wusthoff, Gallistel, & Gibbon, 2001; Dehaene, Dehaene-Lambertz,

* Corresponding author.
E-mail address: hbarth@alum.mit.edu (H. Barth).

0022-2860/$ - see front matter
doi:10.1016/j.cognition.2004.09.011

& Cohen, 1998; Whalen, Gallistel, & Gelman, 1999; Xu & Spelke, 2000). These shared characteristics suggest that a basic preverbal "number sense," or rough representation of mental magnitude, is present prior to the emergence of symbolic number representations. Evidence from studies on human children and adults further suggests that this mental magnitude representation supports the more complex symbolic numerical capabilities developed by humans alone (Dehaene et al., 1998; Gallistel & Gelman, 1992). It remains to be seen, however, whether approximate numerical representations can be used in arithmetical manipulations themselves.

Several lines of evidence lend credence to the hypothesis that the preverbal "number sense" provides a foundation for formal mathematics. First, many tasks that deal explicitly with exact symbolic numerosities automatically activate non-symbolic number representations (Dehaene, Dupoux, & Mehler, 1990; Moyer & Landauer, 1967). Second, human adults are able to perform rapid approximate arithmetic on Arabic digits, apparently utilizing a cerebral circuit distinct from that involved with the retrieval of exact numerical facts (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). Third, neurological patients with impairments to the non-verbal number sense show impaired symbolic calculation abilities (Dehaene et al., 1998). However, a key prediction of the hypothesis has not been conclusively tested: if non-symbolic representations of numerical quantity serve as the foundation for formal mathematics, then it should be possible to perform arithmetic operations on approximate number representations without engaging verbal/symbolic processes.

Suggestive evidence in support of this position comes from studies of non-human animals and of human infants. Untrained rhesus monkeys take account of simple addition and subtraction operations when they observe or choose between arrays containing small numbers of food objects (e.g. Hauser & Carey, 2003; Hauser, MacNeilage, & Ware, 1996; Sulkowski & Hauser, 2001), as do human infants in similar tasks (Feigenson, Carey, & Hauser, 2002; Kobayashi, Hiraki, Mugitani, & Hasegawa, 2004; Wynn, 1992). These studies, however, involved small numbers of objects, which likely are tracked by a system of parallel attention rather than by a system of enumeration (Carey, 2002; Simon, 1997). Preschool children also solve small-valued addition and subtraction problems under certain conditions (Jordan, Huttenlocher, & Levine, 1992; Zur & Gelman, 2004), but they likely use verbal counting to accomplish this task (Gelman, 1990).

Infants and animals also can discriminate between arrays of objects or sequences of sounds containing larger numbers of elements, given large enough comparison ratios (Hauser, Tsao, Garcia, & Spelke, 2003; Lipton & Spelke, 2003; Xu & Spelke, 2000), but it is not clear whether these ranges of numerosities may also be used in arithmetic operations. The foraging patterns of diverse animals depend on the rate of return of food, which has been claimed to require a complex combination of duration and number information (Leon & Gallistel, 1998). Trained animals' successful learning and generalization in number-based tasks also has been claimed to provide evidence for non-verbal addition (see Gallistel, 1990) and subtraction (Brannon et al., 2001). A large body of research on temporal processing, moreover, has been claimed to show that animals engage in arithmetic processing of durations (Gibbon, Malapani, Dale, & Gallistel; 1997; Meck, Church, & Gibbon, 1985). In all of these cases, however, the evidence for arithmetic operations is indirect and the claims have been disputed (e.g. Church & Broadbent, 1990; Dehaene, 2001). The existence of non-symbolic arithmetic capabilities therefore remains in question. Here we test directly for these capabilities in human adults and in preschool children with no relevant knowledge of symbolic arithmetic.

1. Experiment 1

In Experiment 1, adults performed numerical comparison and addition tasks on large sets of elements, presented either as visual arrays of dots or as a mixture of arrays of dots and sequences of tones. For the visual comparison task, subjects were presented with two arrays of dots in sequence and judged which array contained more elements. For the crossmodal comparison task, subjects were presented with an array of dots preceded or followed by a sequence of tones and judged whether the array or the sequence had more elements. For the visual addition task, subjects were presented with three visual arrays of dots and judged whether the third array had more or fewer elements than the sum of the elements in the first two arrays. For the crossmodal addition task, one of the first two visual arrays was replaced by a tone sequence, and subjects judged whether a third array or sequence had more or fewer elements than the sum of the preceding sets of elements.

1.1. Method

Participants. Fourteen adults aged 18–35 were paid to participate in the study. All had normal or corrected-to-normal hearing and vision.

Apparatus. Participants sat in a small darkened room at a distance of approximately 60 cm from the presentation screen. Visual stimuli were presented on a Sony Multiscan monitor by a Power Macintosh 8600 computer. Auditory stimuli were presented from the Macintosh's built-in speaker.

Stimuli. Each auditory sequence consisted of a series of brief tones of equal duration; tone duration and presentation rate varied across sequences (respectively, 20–60 ms tones and 7–11 tones/s). Each tone sequence ended with a high-pitched beep. Visual arrays presented black dots on a mid-gray background, distributed pseudorandomly such that dots did not touch or overlap. Dot size was constant within and variable across arrays (range, 0.2–0.6 cm diameter).

Design. Participants received 5 practice trials and 62 experimental trials in each of four counterbalanced stimulus conditions: Visual Comparison, Visual Addition, Crossmodal Comparison, and Crossmodal Addition. Numerosities ranged from 9 to 63. There were four possible comparison ratios included (collapsed across smaller:larger and larger:smaller ratios): 0.75, 0.8, 0.83, and 0.86 (in Comparison conditions, this was the ratio of the two arrays to be compared; in Addition conditions, this was the ratio of the sum of the first two arrays to the test array). Six of the 62 experimental trials could be answered correctly through strategies based on the numerosity of a single array (for example, judging the final array larger if it fell at the high end of the numerosity range; see Appendix A for further discussion of range-based and other alternative strategies). These 6 trials were excluded from the analysis. The numerosities used in the comparison tasks were matched to the sum and test arrays in the addition task.

Procedure. Visual comparison and addition tasks are depicted in Fig. 1A. For crossmodal comparison, one dot array was replaced by a tone sequence; participants determined whether the second set (whether it was a dot array or a tone sequence) contained more or fewer elements. For crossmodal addition, one of the addends was a tone sequence and one was a dot array. Half of the test sets were tone sequences; half were dot arrays. Participants pressed one button for “fewer” and another button for “more”. Feedback was provided throughout the experiment.

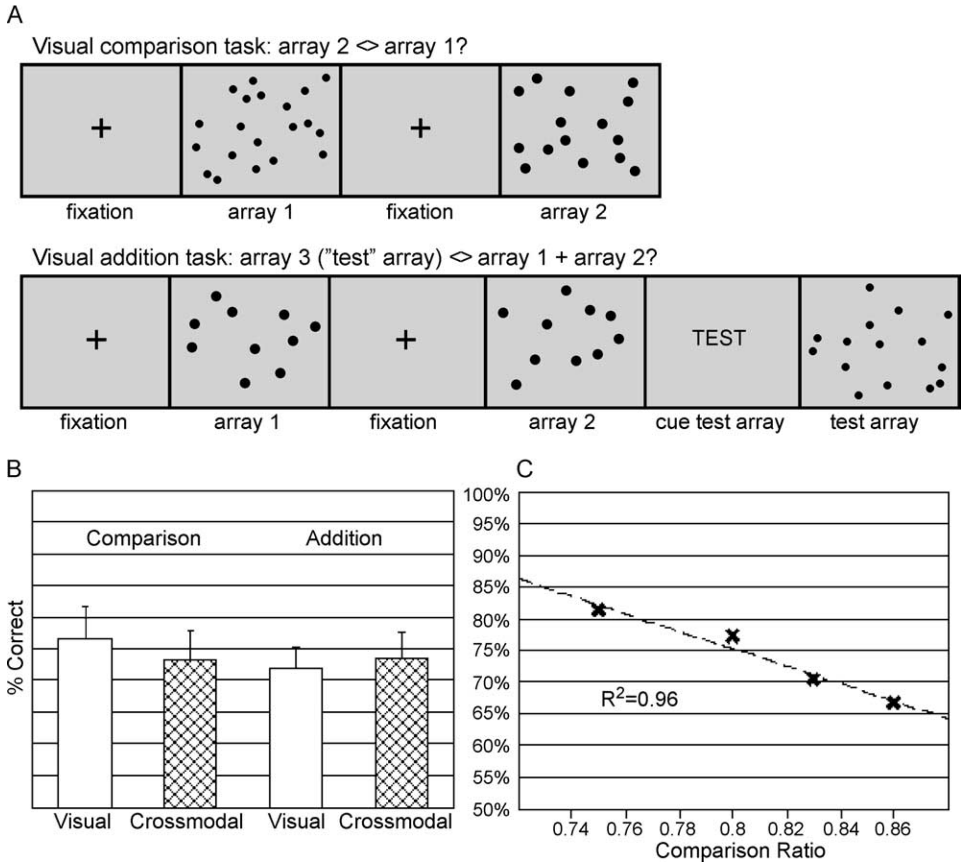

Fig. 1. Non-symbolic comparison and addition, within and across modality. (A) Schematic of visual comparison/addition tasks. (B) Accuracy for 4 conditions with 95% confidence intervals; all conditions > chance (all $t(13) > 10$, $P < 0.0001$). (C) Accuracy as a function of comparison ratio collapsed across conditions. An ANOVA demonstrated no main effects of operation or modality. There was a main effect of comparison ratio ($F(3, 39) = 22.18$, $P < 0.0005$). Accuracy decreased as the ratio of the compared quantities increased, a signature of the rough magnitude representation system. Reaction times (not shown) were as fast for addition as for comparison.

1.2. Results

Accuracy scores for all four conditions were better than chance (all $t(13) > 10$, $P < 0.0001$; see Fig. 1B). A 2 (Operation: addition vs. comparison) × 2 (Modality: visual vs. crossmodal) × 4 (Ratio) repeated-measures ANOVA demonstrated no main effects of Operation or Modality, but there was a main effect of Ratio ($F(3, 39) = 22.18$, $P < 0.0005$; Fig. 1C). Accuracy decreased as the ratio of the compared quantities approached 1, a signature of the rough magnitude representation system. Further analyses of subsets of trials revealed that performance was due neither to non-numerical features of the arrays nor to non-arithmetic strategies (see Appendix A).

1.3. Discussion

Experiment 1 provides evidence that adults can perform addition on non-symbolic number representations. Non-symbolic addition resulted in accuracy scores fully as good as those observed in the simpler comparison tasks. This study also provides additional evidence for the modality-independent nature of approximate large number representations: adults performed approximate arithmetic operations effectively even when both the modalities (visual or auditory) and the formats (spatial or temporal) of the addend sets differed (Barth, Kanwisher, & Spelke, 2003). Success in the across-modality conditions provides evidence that participants based their judgments on abstract *numerosity* rather than on continuous quantity variables that would not transfer across modalities (for further evidence see Appendix A). Accuracy depended on the ratios of the compared quantities: a signature of large approximate numerical representations (Fig. 1C).

2. Experiment 2

A second experiment tested adults' capacity for non-symbolic subtraction by comparing performance in comparison tasks to performance in subtraction tasks with visual arrays. For the subtraction task, subjects again were presented with three visual arrays of dots. They were asked to subtract the second array from the first and to compare this difference to the number of elements in the third array. For the comparison tasks, subjects compared two dot arrays, the first of which was equal in number to the difference in a corresponding subtraction problem and the second of which was identical to the comparison array in that problem.

2.1. Method

Participants. Eleven participants between the ages of 18 and 35 took part in the study for pay. All had normal or corrected-to-normal hearing and vision.

Apparatus and stimuli. These were the same as in Experiment 1.

Design. Participants received two blocks of each stimulus condition in counterbalanced order, 124 trials per condition total. Five practice trials preceded each block. Subtraction problems were reversed versions of Experiment 1 addition problems, arranged so that differences were always positive (e.g. for the addition problem "15 + 20 = (35) vs. 42," the corresponding subtraction problem was "35 − 15 = (20) vs. 24"). The magnitudes that were involved in the addition and subtraction problems therefore were the same, but the quantities to be compared differed. The comparison condition in this study used smaller numerosities, matched to the comparisons made in the subtraction problems. Of the 124 problems per condition, 14 were susceptible to range-based strategies and were omitted from the analysis (see Appendix A). Again there were four comparison ratios: 0.75, 0.8, 0.83, and 0.86.

Procedure. The procedure was as in the first experiment, except that subjects were instructed to "get a rough idea of the difference" of the first two arrays, and to determine whether the test array had fewer or more dots than this difference. Feedback was given throughout the experiment.

2.2. Results

Performance levels were well above chance and varied in relation to the comparison ratios, providing evidence that human adults can perform subtraction operations on approximate quantities. Accuracy scores for non-symbolic subtraction and comparison with dot arrays were 67% (chance = 50%, $t(10) = 9.111$, $P < 0.0001$) and 77% ($t(10) = 9.164$, $P < 0.0001$). A 2 (Operation) × 4 (Ratio) repeated-measures ANOVA demonstrated that accuracy was significantly worse for subtraction than for comparison ($F(1, 10) = 9.47$, $P < 0.02$). There was also a significant main effect of ratio ($F(3, 30) = 8.19$, $P < 0.0005$).

2.3. Discussion

Experiment 2 provides evidence that adults can subtract non-symbolic sets. As in the previous study, accuracy tended to decrease as the ratio of the compared quantities approached 1. In this experiment, however, subjects' accuracy scores were lower for subtraction relative to simple comparison, while there was no accuracy difference for addition vs. comparison in Experiment 1. This does not necessarily mean that the subtraction process is inherently more difficult than addition, however. Alternatively, the presentation of two sequential arrays with the instruction "subtract these!" may not be sufficient to defeat an automatic summation process that operates on contiguously presented quantities (LeFevre, Bisanz, & Mrkonjic, 1988; Olthof & Roberts, 2000; Stadler, Geary, & Hogan, 2001). Because such a process would likely interfere with subtraction, the subtraction deficit might disappear with a task that is more conducive to a subtractive process. Experiment 3 tested this possibility.

3. Experiment 3

To provide a direct comparison of addition and subtraction tasks and to present both tasks in a more natural context, we developed a non-symbolic arithmetic paradigm using animated sequences of events similar to those used in studies of human infants and non-human primates (Hauser et al., 1996; Sulkowski & Hauser, 2001; Wynn, 1992).

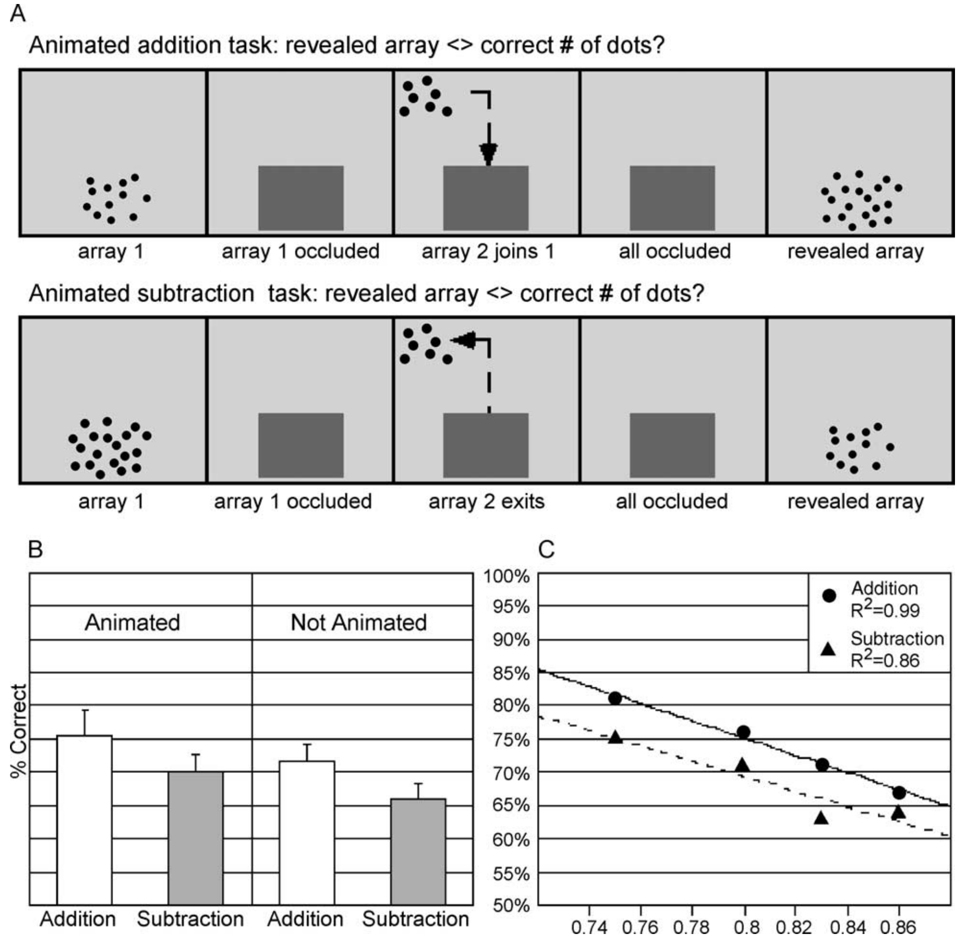

Fig. 2. Non-symbolic addition/subtraction tasks, with and without animation. (A) Schematic of animated versions of addition and subtraction tasks. (B) Accuracy for non-symbolic addition and subtraction with and without animation, with 95% confidence intervals, all >chance (all $t(16)>13$, $P<0.0001$). A mixed design ANOVA demonstrated a main effect of within-subjects factor Operation ($F(1, 32)=17$, $P<0.0005$) and of between-subjects factor Animation ($F(1, 32)=7.05$, $P<0.02$). (C) Accuracy as a function of comparison ratio for means of the addition and subtraction conditions; accuracy decreased as the ratio of the compared quantities increased.

For the addition task (Fig. 2A) we presented one visual quantity, occluded it, and then presented a second quantity that moved behind the occluder to join the first quantity. Then the occluder was removed to reveal the third quantity that differed from the correct sum by one of four ratios, and participants judged whether it was smaller or larger than the expected sum. For the subtraction task, one quantity was presented and occluded, a second smaller quantity moved out from behind the occluder, the occluder was removed to reveal the third quantity, and participants judged whether it was smaller or larger than the expected difference. One group of participants performed these new animated versions of the addition and subtraction tasks, and another group performed the addition and subtraction tasks with visual arrays from Experiments 1 and 2.

3.1. Method

Participants. Participants were 35 adults drawn from the same population, one of whom was excluded for performance more than 2 standard deviations below the mean.

Apparatus. The experimental apparatus was the same as in the previous studies.

Stimuli. The No Animation group used the dot arrays from Experiment 1, in which each array contained dots of the same size, but across arrays the dot sizes varied. Arrays in the Animation conditions occupied a smaller area than arrays in the No Animation conditions, potentially encouraging use of element density as a cue to numerosity, so element size was varied within arrays.

Design. Experiment 3 employed a mixed design (between-subjects factor Presentation Type and within-subjects factors Operation and Ratio). The No Animation and Animation groups (each $n=17$) each received both Addition and Subtraction conditions in blocked, counterbalanced order. The addition and subtraction problems were identical across groups and were the same as those used in the previous addition and subtraction studies, as were the discriminations to be made between the test arrays and the answers to those problems. The No Animation group received two blocks of the Visual Addition condition, and two blocks of the Visual Subtraction condition as described for the earlier studies.

Procedure. The No Animation conditions followed the procedures from the conditions in previous experiments using visual arrays. The presentation differed for the Animation group (see Fig. 2A), though the actual addition and subtraction problems were identical, as were the discriminations to be made between the test arrays and the answers to those problems. Timing was adjusted in order to keep the length of an Animation trial as similar as possible to the length of a No Animation trial. In the Animation Addition condition, the first addend array appeared in an imaginary rectangle in the bottom center portion of the display for ~400 ms. Then an opaque occluding rectangle appeared, concealing the array. After ~1000 ms, while the first array remained concealed, the second array appeared to travel from the top left portion of the display to the edge of the occluder and then to disappear behind it. The motion of the second array was not smooth in order to prevent tracking of the array, which might lead observers to attempt to count the dots. Also, the second array disappeared behind the occluder in a single step so that no cues to numerosity were available from the gradual occlusion of the array. The second array appeared, traveled across the display, and disappeared in ~800 ms. Then after a pause of ~1300 ms, the occluder disappeared, revealing the test array which remained for 400 ms. Subjects were told that they would see a representation of an addition problem; an initial quantity would be shown and covered, followed by the addition of a second quantity to the first, behind the occluder. The occluder would be removed, revealing a third quantity, which was an incorrect representation of the sum. The task was to determine whether the third revealed quantity was too small or too big to be the correct sum. The Animation Subtraction condition was analogous to Animation Addition but the quantities were the same as in the No Animation Subtraction case, and the second array appeared to move out from behind the concealing occluder to the edge of the display rather than moving from the edge of the display toward the occluder. In all conditions in Experiment 3, participants were *not* asked to make speeded responses, but were allowed to respond in whatever way seemed comfortable to them. Participants were told that they would probably do better on the task if they chose the answer that fit their first impression, and did not think too long about the task.

3.2. Results

Accuracy was above chance in all conditions (all ts(16) > 13, $P < 0.0001$), but subtraction performance was worse than addition for both the animated and the non-animated presentation groups (Fig. 2B and C). Overall accuracy scores were as follows: No Animation Addition 71%, No Animation Subtraction 66%, Animation Addition 75%, Animation Subtraction 70%. A mixed design 2 (Presentation type) × 2 (Operation: Addition vs. Subtraction) × 4 (Ratio) ANOVA demonstrated a main effect of the between-subjects factor Presentation Type ($F(1, 32) = 6.970$, $P < 0.02$), a main effect of the within-subjects factor Operation ($F(1, 32) = 17.592$, $P < 0.0005$), and a main effect of the within-subjects factor Ratio ($F(3, 96) = 31.154$, $P < 0.0005$). There were no significant interactions. Participants' accuracy was higher in the condition with animated displays, it again was higher for addition than for subtraction, and it again decreased as the ratio of the compared quantities approached 1. Further analyses revealed that performance did not depend on non-numerical variables correlated with number or on non-arithmetic strategies (Appendix A).

3.3. Discussion

Experiment 3 replicates and extends the findings of Experiments 1 and 2. It provides further evidence that adults can perform both addition and subtraction on visual arrays, that performance on both tasks shows the ratio signature of large approximate number representations, and that subtraction is performed at lower accuracy than addition when the sizes of the individual visual arrays are equated. Although animation of the addition and subtraction operations improved subjects' performance overall, it did not alter these patterns.

Subjects' difficulty with subtraction could stem from two sources. First, non-symbolic subtraction may be inherently more difficult than addition, as is the case with the symbolic versions of these operations (Fuson, 1984). Second, task performance may be limited by error in the representations of the numerosities of the operands. When approximate magnitude representations are subtracted, on this view, their variances should add, yielding a larger variance for the difference numerosity than the variance that would be associated with that numerosity as a sum. This means that the subtraction condition could have produced lower accuracy scores than the addition condition even if the subtraction operation itself is no more prone to error than addition. To test this possibility, we predicted error rates for each arithmetic task, on the assumption that the operations themselves are error-free and that the numerosity representations are characterized by *scalar variability*: the amount of noise in a representation is proportional to its magnitude (Gallistel & Gelman, 1992; Whalen et al., 1999).

Our behavioral data are fit quite accurately taking account only of the variability in the numerosity representations as determined by adults' performance on the numerical comparison tasks, the mathematical rules for combinations of variances, and one free parameter associated with the storage in working memory of the sum or difference (see Appendix B). This model provides a quantitative account of subjects' patterns of performance without positing any difference in the relative difficulties of the addition and subtraction operations.

4. Experiment 4

In Experiments 1–3, participants were adults with years of mathematical training. It is possible, therefore, that they assigned verbal numerical labels to each visual or auditory set and performed the addition and subtraction operations symbolically. Subjects' reports do not support this explanation.[1] Even if adults did not use symbolic arithmetic to solve the non-symbolic addition and subtraction problems in Experiments 1–3, however, we do not know what effects years of mathematical training might have had on the performance of these tasks. In Experiment 4, we tested young children with no relevant training in symbolic arithmetic to determine whether capacities for non-symbolic addition and subtraction are present prior to arithmetic instruction. Five-year-old children were presented with addition and comparison tasks adapted from the animated tasks used with adults in Experiment 3, accompanied by an engaging narration (Fig. 3A).

4.1. Method

Participants. Seventeen 5-year-old participants (range 5;0–5;7, mean 5;3.5) were recruited from the lists of participants in a cognitive development lab in the greater Boston area.

Apparatus. Children were tested in a small, quiet laboratory room. Stimuli were presented on a Power Mac G4 computer with a View Sonic GS790 color monitor and each session was videotaped.

Stimuli. In Comparison trials, children compared the numerosity of a blue dot array to the numerosity of a red dot array (Fig. 3A). Dots within an array were always the same size; dots could be the same size in both arrays or dot size could differ across arrays. The overall areas occupied by the blue and red arrays could also be the same or different across arrays. In addition trials, two arrays of blue dots were added and their (unseen) sum was

[1] No subject reported using a verbal strategy to perform the tasks. For a more direct test of this possibility, we presented a new set of participants with two identical blocks of animated addition trials. For one block they were instructed to choose their intuitive impressions of the right answers; for the other, to make a verbal estimate of the numerosity of each addend array and add those verbal estimates to get a sum. Reaction times were significantly slower when verbal estimates were required than when subjects were encouraged to use their approximate intuitions, and slower than in Experiments 1–3. This finding suggests that the participants in Experiments 1–3 did not engage in verbal estimation, comparison or arithmetic.

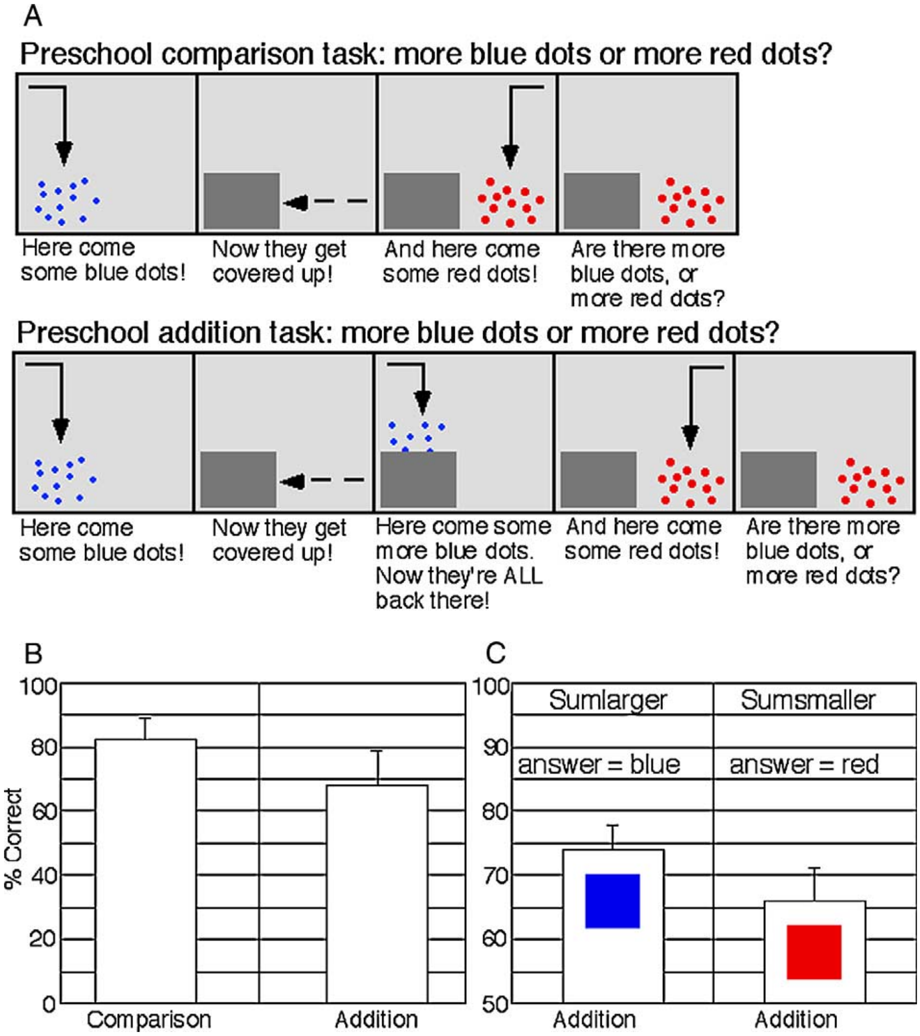

Fig. 3. Preschool comparison and addition. (A) Task schematic. (B) Comparison and addition data, with 95% confidence intervals. Performance in both conditions is >chance (both $t(16)>12$, $P<0.0001$). Comparison performance was better than addition (matched sample $t(16)=2.73$, $P<0.05$). (C) Addition performance by 5-year-olds in a second experiment, plotted separately for sum-larger and sum-smaller problems, reveal above-chance performance in both conditions ($t(17)=4.64$, $P<0.0003$ for sum-larger, $t(17)=2.95$, $P<0.01$ for sum-smaller). Because both addends were always smaller than the comparison array, success on sum-larger problems required genuine approximate addition.

compared to an array of red dots (Fig. 3A). Within a trial, all blue dots were the same size and blue arrays took up the same overall area. Sizes and areas could be the same or different across blue and red arrays. The varying dot sizes allowed us to test for children's use of strategies based on summed dot surface area rather than numerosity. Summed area and other non-numerical stimulus properties such as summed contour length have been found to influence infants' and young children's quantity judgments (Clearfield & Mix, 1999, 2001; Feigenson et al., 2002). There were three interleaved summed area conditions: when the red and blue arrays had dots of the same size, numerosity was positively correlated with summed dot area; when smaller dots accompanied the larger numerosity, summed area was equated across the two arrays; and when larger dots accompanied the larger numerosity, summed area was an exaggerated cue to numerosity. The arrays' overall sizes were varied so that we could examine the contributions of other non-numerical continuous quantity information (such as overall area occupied or dot density: see Appendix C for details of these continuous quantity controls).

Design. Each participant received one block of comparison problems followed by one block of addition problems, with 2 practice and 14 scored problems per block and with numerosities ranging from 6 to 60. On half the trials, the red dots exceeded the sum of the blue dots by a 5:3 ratio; on the remaining trials, the red dots were less than the array of blue dots by a 3:5 ratio. See Appendix C for a description of the controls for non-numerical properties and non-arithmetic strategies.

Procedure. In a comparison problem (see Fig. 3A), a set of blue dots appeared on the screen and was covered by a rectangular occluder. On half the trials, the dots were stationary and the occluder moved horizontally to cover them. On the other trials, the occluder was stationary and the dots appeared above it and moved downward behind it. Once the dots were occluded, an array of red dots moved onto the screen. In an addition problem (see Fig. 3A), the first addend array, a set of blue dots, suddenly appeared on the screen and was occluded as in the first type of comparison trial. Then the second addend array of blue dots appeared above the screen and moved behind it as in the second type of comparison trial. Finally, an array of red dots moved onto the screen. All events were narrated as in Fig. 3. After the final question, the child gave a verbal reply that the experimenter noted on the computer with a keypress.

4.2. Results

Children performed above chance for both conditions (Fig. 3B, both $t(16) > 12$, $P < 0.0001$). Accuracy was higher for comparison problems than for addition problems (matched sample $t(16) = 2.73$, $P < 0.05$). Performance did not depend on any of the tested non-numerical variables or non-arithmetic strategies (Appendix C).

4.3. Discussion

Experiment 4 provides evidence that preschool children can perform approximate comparison and addition of non-symbolic visual sets, and that they use number rather than the covarying continuous quantities of summed dot area, dot density, or overall display area to do so (see Appendix C). Because Experiment 4 used only dot arrays, however, it could not control for all non-numerical quantitative variables at once. In particular, the displays in Experiment 4 controlled for responses to the summed area of the display but not for responses to summed contour length. Because some researchers have suggested that contour length or circumference, rather than summed area, is used in these types of judgments (Clearfield & Mix, 1999, 2001), Experiment 5 tested this possibility.

5. Experiment 5

Experiment 5 replicated the findings of Experiment 4 with additional controls that allowed us to test for strategies based on summing the *circumferences* of the dots rather than the *number* of dots, and with a test of knowledge of relevant symbolic arithmetic facts.

Participants. Eighteen 5-year-old children (range 5;0–5;6, mean 5;1) were recruited from the same population as Experiment 4.

Method. The method was the same as in Experiment 4 except as follows. Children were given 4 practice comparison problems, 4 practice addition problems, and 24 test addition problems. Three conditions were intermixed to allow testing for strategies based on summing the *circumferences* of the dots rather than the *number* of dots (see Appendix C for details). Also, performance for trials in which the sum of the blue dots was *smaller* than the red array's numerosity was compared to performance on trials in which the blue sum was *larger*. In every trial, each blue addend array was smaller than the red array. If children simply answered by comparing a single blue addend array to the red array, or if children tended to answer "blue is larger" because there were always two blue arrays, performance would be poor on the "sum-larger" problems.

Experiment 5 included a post-test in which the children were given four symbolic addition problems. One of the problems was very easy ($5+5=10$ vs. 20) and served as a control for motivation and understanding of task demands; the other three problems were based on the dot problems and presented either the exact numerosities that appeared in those problems (e.g. $12+9=35$ vs. 21) or rounded versions of those numerosities (e.g. $20+10=30$ vs. 50). The symbolic problems were embedded in a simple story: "If your mom gave you 8 marshmallows, and then she gave you 13 more, how many would you have?" If children refused to produce the answer or produced an erroneous answer, a follow-up two-choice question was given: "Would it be more like 21 marshmallows or more like 35 marshmallows?"

5.1. Results

Addition task. Fig. 3C depicts the data plotted separately for problems in which the sum was larger or smaller than the comparison array. Performance was above chance in both conditions ($t(17)=4.64$, $P<0.0003$ for sum-larger, $t(17)=2.95$, $P<0.01$ for sum-smaller). A repeated-measures 2 (sum-larger vs. sum-smaller)$\times$3 (contour length control conditions) ANOVA showed no significant effects of either manipulation; accuracy levels were the same for the three conditions that tested for summing of dot circumference, and for the sum-larger vs. sum-smaller trials (all Fs<1, see Appendix C for details).

Symbolic post-test. For the easy problem, 58% of the children produced the correct answer spontaneously and 83% of the children were correct on the multiple choice test (binomial $P<0.05$), indicating that they understood the addition problems and were motivated to solve them. For the test problems, none of the children produced any correct answers, and performance on the two-choice test was 48%, not significantly different from chance (50%).

5.2. Discussion

The findings of Experiment 5 replicate and extend those of Experiment 4 while ruling out the possibility that children responded by simply comparing a single blue array to the red array. Because both addends were smaller than the comparison array for all problems, performance could not simply be based on comparison, suggesting a genuine operation of approximate addition. Participants also did not rely on continuous quantity information to make their judgments, as demonstrated by the continuous quantity test conditions (see Appendix C for further discussion of this and tests for additional alternative strategies that might have been adopted). Five-year-old children therefore are able to assess, represent, add, and compare discrete numerical quantities. The symbolic post-test confirmed that the children did not know either the exact or the approximate answers to the tested addition problems when the problems were presented symbolically. Their high accuracy on the non-symbolic problems therefore is not attributable to knowledge of symbolic addition.

6. General discussion

Three experiments provide evidence that adults can mentally add two arrays, or subtract one array from the other, and then compare the sum or difference to a third array. Addition was as accurate as comparison, was equally accurate when the two addends appeared in the same vs. different modalities and formats, and showed the ratio signature of large number representations. Subtraction produced a performance deficit relative to addition, but this difference may be due to variability in the magnitude representations rather than to the relative difficulties of the operations themselves.

Further evidence for non-symbolic arithmetic abilities comes from two experiments on five-year-old children with no relevant knowledge of symbolic arithmetic. Children showed above-chance performance on both visual comparison and visual addition tasks, although they performed at chance on symbolic versions of these tasks. Detailed trial-by-trial analyses showed that they did not use other strategies based on non-numerical stimulus properties or other operations. Children's performance did not result from the operation of attentional mechanisms that track small numbers of objects, nor did they adopt alternative non-addition strategies to complete the tasks. Previous investigations of non-symbolic addition capabilities have produced results that could be attributed to symbolic counting, attentional object tracking mechanisms, or other processes distinct from non-verbal addition. These findings provide the first unambiguous evidence that the capacity for non-symbolic addition emerges in humans prior to mathematics instruction.

In conclusion, these studies provide evidence that primitive approximate number representations enter into arithmetic operations, without influence from learned mathematics. Both adults and children without symbolic arithmetic training can use non-symbolic numerosities for arithmetic processing. These representations, shared by humans and other animals, therefore may play a key role in the development of formal mathematics (Geary, 1995; Spelke, 2000). Advances in understanding of non-symbolic numerical abilities may allow educators to harness this primitive number sense to enhance early mathematics instruction.

Acknowledgements

We thank Mary C. Potter and C.R. Gallistel for comments on an earlier version of this manuscript. This research was supported by National Institute of Health grant MH56037 to N. Kanwisher, National Institute of Health grant R37 HD23103 to E. Spelke, National Science Foundation ROLE grant REC-0087721 to N. Kanwisher and E. Spelke, a National Academy of Education/Spencer Foundation Postdoctoral Fellowship to H. Barth, and a McDonnell centennial fellowship to S. Dehaene.

Appendix A

A.1. Ruling out alternative strategies: adult studies

Though Experiment 4 and the preschool studies provide evidence against the use of verbal strategies in these tasks, there are alternative strategies that do not depend on verbal arithmetic knowledge. The addition and subtraction tasks, if carried out as instructed, involve a large working memory load. This may motivate participants to adopt strategies with relatively smaller demands on working memory. Some of these alternative non-verbal strategies may produce response patterns similar to those that would be observed if participants were indeed carrying out the desired addition and subtraction processes. For this reason, we conducted trial-by-trial analyses of our data to rule out the use of five likely alternative strategies.

A.2. Extremes of range strategy

Subjects may have attempted to determine the correct answer by judging that the test array is smaller [larger] than the sum when one of the addend arrays was especially large [small], or by responding based on the size of the test array alone. The addition and subtraction equations were designed so that this strategy would always yield the correct answer for only a small subset of trials (6 trials out of 62 per condition in Experiment 1, 14 out of 124 trials per condition in Experiment 2). These trials were excluded from our analyses to avoid potential inflation of accuracy scores. Experiments 3 and 4 used the same equations as 1 and 2 and their data were analyzed the same way.

A.3. Near/far strategy

In the addition condition, when there is little difference between the test array and the larger addend, the answer is likely to be "sum > test array." When there is a large difference between the test and the larger addend, the answer is likely to be "sum < test array." Therefore, participants might have simply judged the distance (near or far) between the test array and the largest addend. There is an analogous version of this strategy for the subtraction condition as well: when the first (larger) operand and the test array are close, the test array is larger than the sum; if the first operand and the test array are far, the test array is smaller than the sum. These near/far strategies usually make the same predictions as the addition or subtraction strategies that we ask subjects to use. Near/far strategies will fail for a subset of trials; to rule out the use of these strategies we examined performance for this subset.

For the addition conditions, the distance between the larger addend and *Z*, the test array, ranges from 5 to 29. For each subject, we tested the success rate for a subset of trials with intermediate distances, where the strategy should fail to make any clear prediction (distances from 15 to 19). For the subtraction conditions, the distance ranges from 3 to 27. Here we again looked at a subset of trials in which the near/far strategy gives no clear prediction (trials with an X–Z distance from 13 to 18). The use of the near/far strategy should lead to chance performance on these trial subsets; but in fact these trials produced better-than-chance performance. The results of the analyses follow.

Experiment; condition	Trials/subject	Accuracy (%)	*P* value	> Chance (50%)
1; visual addition	9	63	$P(13) < 0.03$	Yes
1; crossmodal addition	9	69	$P(13) < 0.003$	Yes
2; visual subtraction	28	74	$P(10) < 0.0001$	Yes
3; addition w/o animation	18	65	$P(16) < 0.0001$	Yes
3; subtraction w/o animation	28	75	$P(16) < 0.0001$	Yes
3; addition with animation	18	64	$P(16) < 0.0002$	Yes
3; subtraction with animation	28	76	$P(16) < 0.0001$	Yes

A.4. First operand strategy

Participants were asked to add the first two arrays and to compare this sum to the third array ($X+Y$ vs. Z). Because there is a substantial working memory demand involved in this judgment, participants may have adopted strategies that require fewer quantities to be held in working memory. They may have based their judgments on different comparisons such as $2X$ vs. Z instead. Analogously, for subtraction, instead of X–Y vs. Z, subjects might have used $X/2$ vs. Z. To test for the use of the $2X$ or $X/2$ judgments, we looked at the subsets of trials that could not be answered correctly based on these strategies. Our analyses showed that these subsets of trials produced above-chance performance.

Experiment; condition	Trials/subject	Accuracy (%)	*P* value	> Chance (50%)
1; visual addition	6	66	$P(13) < 0.03$	Yes
1; crossmodal addition	6	66	$P(13) < 0.0005$	Yes
2; visual subtraction	40	56	$P(10) > 0.05$	No*
3; addition w/o animation	12	68	$P(16) < 0.0001$	Yes

(*continued on next page*)

Experiment; condition	Trials/subject	Accuracy (%)	P value	>Chance (50%)
3; subtraction w/o animation	40	58	$P(16)<0.01$	Yes
3; addition with animation	12	74	$P(16)<0.0001$	Yes
3; subtraction with animation	40	70	$P(16)<0.0001$	Yes

[*]All of the problems in this subset of trials are "difference larger" problems. Therefore, tests against chance at 50% may be biased because many subjects show a tendency toward answering "test array larger" in subtraction conditions, so performance is typically worse on this type of "difference larger" trial. In particular, two of the participants in Experiment 2 showed an extremely strong "test array larger" bias. To test fairly for this strategy, it is necessary to find a good baseline against which to look at performance on these trials. We calculated a bias-corrected chance level against which to test performance; subjects answered "test array larger" for 57% of the trials in this experiment, so the bias-corrected chance level for our critical trial subset was 43%. Using the corrected chance level, performance on these trials is better than chance at 56%, $P(10)<0.0001$, Also, performance on corresponding trial subsets in the other subtraction experiments, whose participants were less biased toward answering "test array larger," was better than chance at the uncorrected 50% level. These findings provide evidence that participants do not perform the subtraction task by computing $X/2$ vs. Z.

A.5. Second operand strategy

Participants may have based their addition judgments on the comparison of $2Y$ vs. Z instead of $X+Y$ vs. Z. To test for the use of this strategy, we analyzed the subsets of trials that could not be answered correctly based on this strategy. Our analyses showed that these subsets of trials produced above-chance performance.

Experiment; condition	Trials/subject	Accuracy (%)	P value	>Chance (50%)
1; visual addition	6	63	$P(13)<0.02$	Yes
1; crossmodal addition	6	74	$P(13)<0.0003$	Yes
3; addition w/o animation	12	63	$P(16)<0.0004$	Yes
3; addition with animation	12	70	$P(16)<0.0001$	Yes

A.6. Larger operand strategy

Participants may have based their addition judgments on the comparison of 2(max X, Y) vs. Z instead of $X+Y$ vs. Z. To test for the use of the 2(max X, Y) vs. Z strategy, we analyzed the subsets of trials that could not be answered correctly based on this strategy. Our analyses showed that these subsets of trials produced above-chance performance.

Experiment; condition	Trials/subject	Accuracy (%)	P value	>Chance (50%)
1; visual addition	6	76	$P(13)<0.001$	Yes
1; crossmodal addition	6	80	$P(13)<0.0001$	Yes
3; addition w/o animation	12	79	$P(16)<0.0001$	Yes
3; addition with animation	12	74	$P(16)<0.001$	Yes

A.7. Summary

These trial-by-trial analyses provide strong evidence to support the claim that participants did perform these tasks by adding and subtracting numerical quantities, and that they did not rely on alternative strategies. A combination of these strategies is also unlikely to account for participants' success: many of the tested trials would have yielded incorrect answers for multiple alternative strategies.

Appendix B

B.1. Model of non-symbolic arithmetic performance

Here we introduce a simple mathematical model that explains in quantitative terms why and how performance in non-symbolic arithmetic is determined by the ratio of the numbers involved. In particular, the model can explain the lower performance in the subtraction task compared to the addition task without assuming a difference in the error associated with these two operations.

We assume that the numerosity of a set of n dots is represented internally by a Gaussian random variable N with mean n and with standard deviation wn, where w is the Weber fraction specifying the degree of precision of the representation (Gallistel, 1990; Gallistel & Gelman, 1992; similar though analytically less tractable results are obtained under the assumption of a logarithmic internal representation with fixed variability).

$$p[N \in (x, x + \mathrm{d}x)] = \frac{\exp\left(-\frac{1}{2}\left(\frac{x-n}{wn}\right)^2\right)}{\sqrt{2\pi}wn}\mathrm{d}x$$

Given this form of internal encoding, it is possible to derive mathematically the optimal response strategy and associated error rate for various numerical tasks. In the **comparison task**, optimal performance is achieved by computing the sign of an internal response criterion which is the difference of $N1$ and $N2$. This criterion, being the sum of two Gaussian random variables, is also a Gaussian random variable with mean $n1$–$n2$ and standard deviation $w\sqrt{n1^2 + n2^2}$. The error rate is the area under the tail of this Gaussian curve, or

$$p_{\text{comparison}} = \int_0^\infty \frac{\exp\left(-\frac{1}{2}\left(\frac{x+\mathrm{Abs}(n1-n2)}{w\sqrt{n1^2+n2^2}}\right)^2\right)}{\sqrt{2\pi}w\sqrt{n1^2+n2^2}}\mathrm{d}x = \frac{1}{2}\mathrm{erfc}\left(\frac{\mathrm{Abs}(n1-n2)}{\sqrt{2}\sqrt{n1^2+n2^2}w}\right)$$

where erfc(z) is the complementary error function, given by

$$\mathrm{erfc}(z) = 1 - \frac{2}{\sqrt{\pi}}\int_0^z \mathrm{e}^{-t^2}\,\mathrm{d}t$$

By letting $r = \frac{n2}{n1}$, one gets

$$p_{\text{comparison}} = \frac{1}{2}\mathrm{erfc}\left(\frac{\mathrm{Abs}(r-1)}{\sqrt{2}w\sqrt{r^2+1}}\right)$$

In the **addition task**, subjects are presented with three sets of numerosities $n1$, $n2$ and $n3$. A simple response criterion consists in computing the sign of $N1+N2-N3$. This value, being the sum of Gaussian random variables, is also a Gaussian random variable with mean $n1+n2-n3$ and standard deviation $w\sqrt{n1^2+n2^2+n3^2}$. The error rate is the area under the tail of this Gaussian curve, or

$$p_{\text{addition}} = \frac{1}{2}\,\text{erfc}\left(\frac{\text{Abs}(n1+n2-n3)}{\sqrt{2}w\sqrt{n1^2+n2^2+n3^2}}\right)$$

In general, this formula depends on the values of all three numbers $n1$, $n2$ and $n3$ rather than the mere ratio $r=\frac{n3}{n1+n2}$. However, in the case where $n1\cong n2$, where the sum is divided into two approximately equal numbers, the equation can be simplified as

$$p_{\text{addition}} = \frac{1}{2}\,\text{erfc}\left(\frac{\text{Abs}(r-1)}{\sqrt{2}w\sqrt{r^2+\frac{1}{2}}}\right)$$

More generally, letting $\alpha=\frac{n1}{n1+n2}$

$$p_{\text{addition}} = \frac{1}{2}\,\text{erfc}\left(\frac{\text{Abs}(r-1)}{\sqrt{2}w\sqrt{r^2+1+2\alpha(\alpha-1)}}\right)$$

Somewhat paradoxically, those equations indicate that, under an optimal response criterion, the error rate for addition is always smaller than for the corresponding comparison alone (because the term in α is always negative, $0\leq\alpha\leq1$). Intuitively, the superior performance in addition vs. comparison alone is due to the fact that, because the *variances* of the terms add, the *standard deviation* of a sum of scalar quantities can be proportionally smaller than the initial weber fraction of the operands. However, since no such effect is obtained in the human data, which indicate slightly higher error rates in the addition task than in the comparison task, we have to consider whether other sources of variance contribute to calculation performance in human subjects.

A simple hypothesis is that the temporary storage of the operation result contributes an additional variance term. It is natural to propose that this term is again scalar and has a variance proportional to the result of the operation ($\sigma=z(n1+n2)$ in the case of addition).

Under these conditions, the error probability becomes

$$p_{\text{addition}} = \frac{1}{2}\,\text{erfc}\left(\frac{\text{Abs}(n1+n2-n3)}{\sqrt{2}\sqrt{w^2(n1^2+n2^2+n3^2)+z^2(n1+n2)^2}}\right)$$

Or, with $r=\frac{n3}{n1+n2}$ and $\alpha=\frac{n1}{n1+n2}$

$$p_{\text{addition}} = \frac{1}{2}\,\text{erfc}\left(\frac{\text{Abs}(r-1)}{\sqrt{2}\sqrt{w^2(r^2+1+2\alpha(\alpha-1))+z^2}}\right)$$

The case of the **subtraction task** can be treated similarly. The corresponding error probability is

$$p_{\text{subtraction}} = \frac{1}{2}\,\text{erfc}\left(\frac{\text{Abs}(n1-n2-n3)}{\sqrt{2}\sqrt{w^2(n1^2+n2^2+n3^2)+z^2(n1-n2)^2}}\right)$$

Or, with $r=\frac{n3}{n1-n2}$ and $\alpha=\frac{n1}{n1-n2}$

$$p_{\text{subtraction}} = \frac{1}{2}\text{erfc}\left(\frac{\text{Abs}(r-1)}{\sqrt{2}\sqrt{w^2(r^2+1+2\alpha(\alpha-1))+z^2}}\right)$$

Given that $\alpha \geq 1$ (because $n1 \geq n2 \geq 0$), the term in α is strictly positive. This implies that the error rate in subtraction is always higher than that of the corresponding comparison problem alone, and also always higher than any corresponding addition problem with the same result. Intuitively, this is due to the greater magnitude of the operands in subtraction problems than in addition problems, which causes a greater uncertainty in the encoding of the operands.

The above equations were fitted to the data in two steps (see Fig. A1; data points represent behavioral data, lines show the predictions of the model). First, the error rates in comparison alone were fitted using non-linear regression with a single free parameter, the perceptual Weber fraction w. An excellent fit was achieved with $w=0.19$, a value comparable to that reported in other numerical perception or production paradigms (Whalen et al., 1999). Second, this value of w was applied to the equations for addition and subtraction error rates, which were fitted to the data again using a single free parameter, the calculation Weber fraction z. An excellent fit was obtained with $z=1.3w$. This simple model therefore provides a quantitative account of the higher error rates in subtraction than in addition without attributing any greater error to the subtraction operation itself.

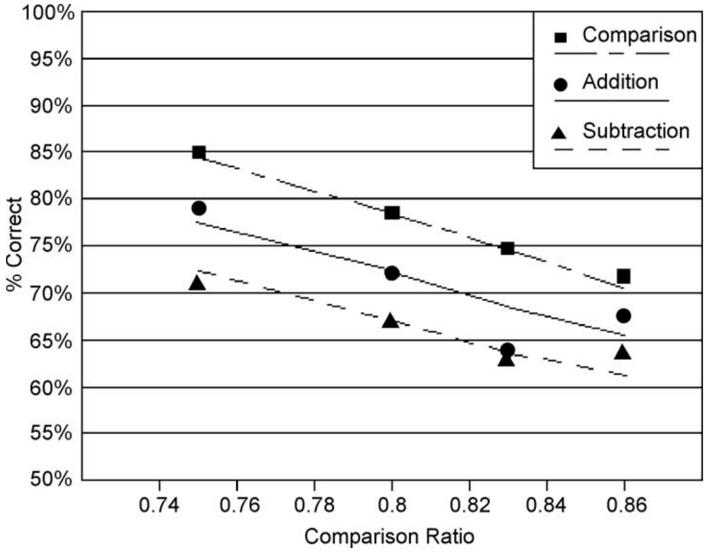

Fig. A1. The fit of the model to our data. Points represent behavioral data (accuracy scores); lines represent the model's predictions.

Appendix C

C.1. Ruling out alternative strategies: preschool studies

In Experiments 4 and 5 (dot array addition with 5-year-olds), the problems were much easier than those given to the adults and precluded use of the same methods to rule out the 5 alternative strategies. Nevertheless, our analyses suggest that children did not adopt any of these strategies.

C.1.1. Near/far strategy

For the 5-year-olds' addition problems, this distance-based strategy always predicted the correct answer. Also, for Experiment 4, the distance was always quite clearly either near or far; there were no "no clear prediction" trials to test as we did with the adults. For Experiment 5, however, there were trials that were near (distance 1, 2, or 5), far (distance 30 or 35), or medium (distance 10 or 14). Performance was no worse on the medium trials than on the near or far trials. Performance on the medium trials was above chance; 72%, $P(17) < 0.0004$. These findings suggest that 5-year-olds did not use this strategy.

C.1.2. 2X, 2Y, and 2(max X,Y) strategies

Because the comparison ratios are quite easy in the addition problems for 5-year-olds, there are no problems in which these strategies make different predictions from $X+Y$ vs. Z. Therefore, we cannot test subsets of the trials directly as we did with the adults. However, all of the problems in Experiments 4 and 5 have the same comparison ratio (sum:test = 3:5 or 5:3); this means that difficulty based on the $X+Y$ strategy should always be the same. If the subjects are using these other strategies, however, the difficulties would not always be the same because the $2X$:Z, $2Y$:Z, and 2(max X, Y):Z ratios vary. By looking at performance across these ratios, it should be possible to get an idea of whether these strategies are being used.

If subjects are simply using $X+Y$ vs. Z, the ratios associated with these other strategies should not matter and performance should not vary as the ratios vary. Because of the small numbers of trials involved, statistically significant effects of ratio are unlikely. Therefore, to avoid unfairly concluding that participants are adding, we will interpret any apparent non-significant effect of ratio in terms of the use of an alternative (non-addition) strategy.

The following descriptions of the data are qualitative and do not refer to significant effects. For Experiment 4, accuracy drops with increasing difficulty in the $2Y$ strategy ratios, while accuracy increases with difficulty in the $2X$ ratios. This pattern looks as though children might be using the $2Y$ strategy. But, for Experiment 5, the pattern goes in the opposite direction. Accuracy appears to increase as the $2Y$ ratios get more difficult, while accuracy decreases as the $2X$ ratios get more difficult. If anything, the Experiment 5 children appear to be using the $2X$ strategy. Therefore, either the two sets of 5-year-olds are using entirely different non-addition strategies that require non-symbolic multiplication instead (which seems unlikely), or we can fairly interpret these data as showing no effect of $2X$:Z or $2Y$:Z ratios. In neither study is there any particular direction apparent in the 2(max X, Y) ratios. These patterns suggest that the preschool participants performed the addition task by computing $X+Y$ vs. Z, as do adults.

C.2. Further testing for alternative strategies in 5-year-olds: continuous quantity

C.2.1. Experiment 4: Comparison and addition

Effects of summed surface area on children's judgments were assessed by presenting three types of problems. Of the 14 problems in each category, 6 used the same size dots for both arrays (so that numerosity was positively correlated with surface area), 4 used smaller dots for the larger numerosity (so that surface area was equated across the two arrays), and 4 used larger dots for the larger (so that surface area was an exaggerated cue to numerosity). Two of the problems in the 6-problem set were not repeated in the 4-problem sets. Statistical tests reported here were performed on the entire 14-problem set, except for the surface area manipulation test which included only the 12-problem subset omitting the two non-repeated problems. All tests were performed on both data sets, yielding the same results. Effects of overall array area and density on children's judgments were assessed by assigning each trial pseudorandomly to one of two groups. In the first group, the arrays with the larger numerosities were presented in a larger area on the screen, holding density roughly constant. In the second group, the array with the larger numerosity had a greater density, holding overall area constant.

Children displayed high performance levels (Fig. 3B). Performance on both operations was higher for the trials in which the larger numerosity was presented in a larger area (comparison $t(16)=3.27$, $P<0.002$; addition $t(16)=4.29$, $P<0.0006$), suggesting that children weighted area more heavily than density in making their judgments. Nevertheless, performance was above chance for both groups in both operations (all $t(16)>4$, $P<0.005$), demonstrating that area was not the sole determinant of children's choices. There was no effect of the surface area manipulation on performance ($F(2, 32)=0.54$, $P>0.5$); children did not respond based on summed dot surface area. We also analyzed performance separately for problems for which the sum was larger vs. smaller than the comparison array. Performance was above chance for both conditions (both $t(16)>5$, $P<0.0001$) which did not differ from each other (matched samples $t(16)=1.84$, $P>0.05$).

C.2.2. Experiment 5: Addition by preschool children

Three conditions were intermixed to allow testing for strategies based on summing the circumferences of the dots rather than the number of dots. In the first condition, all of the dots were the same size, so that summing the contour lengths of the dots would produce correct responses. In the second, when there were more red dots the red dots were also bigger (and the same for the blue dots). Again, summing the contour lengths of the dots would produce correct responses. In the third condition, the larger dots were associated with the smaller numerosity so that total red and blue contour lengths were equalized. Here, a strategy based on contour length would lead to chance performance. Accuracy levels were the same for the three conditions, demonstrating that children did not base their judgments on this continuous variable ($F(2, 34)<1$).

Strategies based on the overall area and density of the arrays were assessed by dividing the trials into 3 groups. As in Experiment 4, for the first group the arrays with the larger numerosities were presented in a larger area on the screen in order to hold density roughly constant, and for the second group, overall area was held constant. A third group was included in Experiment 5 in which the larger numerosity was presented in a smaller overall area. There were no differences among these three conditions ($F(2, 34) = 2.6$, $P > 0.05$). Children evidently based their responses on numerosity, not summed contour length, area, or density, and they compared the final array to the sum of the addends rather than to one or the other addend individually.

To assess whether children followed numerical range-based guessing strategies (for example, when the final comparison array is larger than average, guess that it exceeds the sum; when it is smaller than average, guess that it is less than the sum), we defined a subset of addition problems for which these strategies would yield chance performance and compared children's performance on that subset (mean 73%) to their performance on the rest of the problems (mean 67%). Both groups produced above-chance performance (both $t(17) > 4$, $P < 0.001$) and they did not differ from each other (matched samples $t(17) = 1.81$, $P > 0.05$). These findings provide further evidence that children engaged in a process of non-symbolic addition.

References

Barth, H., Kanwisher, N., & Spelke, E. S. (2003). The construction of large number representations in adults. *Cognition*, *86*, 201–221.

Brannon, E. M., & Terrace, H. S. (2002). The evolution and ontogeny of ordinal numerical ability. In M. Beckoff, C. Allen, & G. M. Burghardt (Eds.), *The cognitive animal* (pp. 197–204). Cambridge, MA: MIT Press.

Brannon, E. M., Wusthoff, C. J., Gallistel, C. R., & Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science*, *12*, 238–243.

Carey, S. (2002). Evidence for numerical abilities in young infants: A fatal flaw? *Developmental Science*, *5*, 202–205.

Church, R. M., & Broadbent, H. A. (1990). Alternate representations of time, number, and rate. *Cognition*, *37*, 55–81.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, *10*, 408–411.

Clearfield, M. W., & Mix, K. S. (2001). Amount versus number: Infants' use of area and contour length to discriminate small sets. *Journal of Cognition and Development*, *2*, 243–260.

Dehaene, S. (2001). Subtracting pigeons: Logarithmic or linear? *Psychological Science*, *12*, 244–246.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the human and animal brain. *Trends in Neurosciences*, *21*, 355–361.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 626–641.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974.

Feigenson, L., Carey, S., & Hauser, M. D. (2002). The representations underlying infants' choice of more: Object-files versus analog magnitudes. *Psychological Science*, *13*, 150–156.

Fuson, K. (1984). More complexities in subtraction. *Journal of Research in Mathematics Education*, *15*, 214–225.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

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

Geary, D. (1995). Reflections of evolution and culture in children's cognition. *American Psychologist*, *50*, 24–37.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate–inanimate distinction as examples. *Cognitive Science*, *14*, 79–106.

Gibbon, J., Malapani, C., Dale, C., & Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology*, *7*, 170–184.

Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: Examinations of content and format. *Cognitive Psychology*, *47*, 367–401.

Hauser, M. D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences USA*, *93*, 1514–1517.

Hauser, M. D., Tsao, F., Garcia, P., & Spelke, E. S. (2003). Evolutionary foundations of number: Spontaneous representations of numerical magnitudes by cotton-top tamarins. *Proceedings of the Royal Society, London*, *B270*, 1441–1446.

Jordan, N., Huttenlocher, J., & Levine, S. C. (1992). Differential calculation abilities in young children from middle- and low-income families. *Developmental Psychology*, *28*, 644–653.

Kobayashi, T., Hiraki, K., Mugitani, R., & Hasegawa, T. (2004). Baby arithmetic: One object plus one tone. *Cognition*, *91*, B23–B34.

LeFevre, J. A., Bisanz, J., & Mrkonjic, L. (1988). Cognitive arithmetic: Evidence for obligatory activation of arithmetic facts. *Memory and Cognition*, *16*, 45–53.

Leon, M. I., & Gallistel, C. R. (1998). Self-stimulating rats combine subjective reward magnitude and subjective reward rate multiplicatively. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 265–277.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large number discrimination in human infants. *Psychological Science*, *14*, 396–401.

Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*, 591–597.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520.

Olthof, A., & Roberts, W. A. (2000). Summation of symbols by pigeons (*Columba livia*): The importance of number and mass of reward items. *Journal of Comparative Psychology*, *114*, 158–166.

Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A 'non-numerical' account. *Cognitive Development*, *12*, 349–372.

Spelke, E. S. (2000). Core knowledge. *American Psychologist*, *55*, 1233–1243.

Stadler, M., Geary, D., & Hogan, M. (2001). Negative priming from activation of counting and addition knowledge. *Psychological Research*, *65*, 24–27.

Sulkowski, G. M., & Hauser, M. D. (2001). Can rhesus monkeys spontaneously subtract? *Cognition*, *79*, 239–262.

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

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*, 749–750.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, 1–11.

Zur, O., & Gelman, R. (2004). Young children can add and subtract by predicting and checking. *Early Childhood Research Quarterly*, *19*, 121–137.