

# Perception Precedes Computation: Can Familiarity Preferences Explain Apparent Calculation by Human Babies?

David S. Moore

Pitzer College and Claremont Graduate University

Laura A. Cocas

Pitzer College

Two studies of 5-month-old infants explored whether a phenomenon reported by K. Wynn (1992) reflects a familiarity preference instead of a mathematical competence. Experiment 1 was a conceptual replication of Wynn's study. When data were analyzed with the relatively liberal statistical approach used by Wynn, the original phenomenon was replicated. However, an analysis of variance revealed that girls and boys behaved in different ways, and that boys did not behave as Wynn would have predicted. Experiment 2 was identical to Experiment 1, with one exception that should not have influenced computation: Infants in this study were completely familiarized with the test displays before testing. Again, the data revealed an interaction involving sex: Boys tended to be influenced by their familiarity with the test displays, whereas girls tended to behave as Wynn would have predicted. These findings are discussed with reference to the literature on sex differences—specifically the finding that male infants are typically immature relative to their female peers—as well as to articles that have been critical or supportive of Wynn's conclusions.

*Keywords:* infants, calculation, mathematics, sex differences, familiarity preferences

The abilities to form and manipulate cognitive representations of quantity are undoubtedly adaptive and have been found to characterize human adults and the adults of several other types of animals, including monkeys (Hauser, Tsao, Garcia, & Spelke, 2003), dogs (West & Young, 2002), birds (Pepperberg, 1991), and rodents (Capaldi & Miller, 1988). Such findings have been taken as support for the argument that certain quantitative competencies—including detection of numerosity, counting, and simple arithmetic—are evolved forms of cognition that are innately specified (Bjorklund & Bering, 2002; Geary, 2000). Proponents of this idea have sought additional support for their claims in reports that 5-month-old human infants can “compute the precise results of simple additions and subtractions” (Wynn, 1992, p. 749). Although it now appears that infant behaviors that have been con-

troversially interpreted as calculation can be replicated (Simon, Hespos, & Rochat, 1995; Uller, Carey, Huntley-Fenner, & Klatt, 1999; but see also Vilette, 2002, or Wakely, Rivera, & Langer, 2000), understanding such behaviors as innate does nothing to illuminate their developmental origins (Lickliter & Berry, 1990; Moore, 2002). Instead, it is always advisable to seek an understanding of how various component processes interact to contribute to the development of a competence.

Several reinterpretations of the “calculation” phenomenon originally reported by Wynn (1992) have been offered in recent years; each explains Wynn's data in a way that does not require awarding sophisticated computational abilities to young infants (Cohen & Marks, 2002; Ferraro, 1993; Haith, 1998; Simon, 1997). The current experiments were undertaken to explore the hypothesis that Wynn's study can best be understood as having engaged a simple perceptual process, namely the tendency of infants to be influenced by recently perceived stimuli.

Wynn's (1992) original study relied on the presumption that an infant's expectations are revealed when looking time to one of two equally attractive test displays is increased following a pretest event; such an increase in fixation during test trials has been taken (by Baillargeon, Spelke, & Wasserman, 1985, among others) to indicate violation of an expectation that was formed as a result of witnessing the pretest event. The stimuli used in Wynn's study were real, three-dimensional objects manipulated on a puppet-theater-type stage. Wynn's participants had no reliable visual preference initially for one- versus two-object test displays. However, after seeing a pretest event involving either the addition of one object to one object ( $1 + 1$ ) or the subtraction of one object from two objects ( $2 - 1$ )—an event that was presumed to produce an expectation about the resulting number of objects to be seen in the test displays—participants no longer fixated the displays equally. Rather, as Wynn predicted, participants fixated test displays longer if they contained an “unexpected” number of objects. That is, participants fixated one object longer after seeing the  $1 +$

---

David S. Moore, Department of Psychology, Pitzer College; Department of Psychology, Claremont Graduate University. Laura A. Cocas, Department of Psychology, Pitzer College. Laura A. Cocas is now at the Department of Neuroscience and the Interdisciplinary Program in Neuroscience, Georgetown University.

These experiments were conducted while David S. Moore was supported by funds provided by various research and awards committees of Pitzer College (excluding a committee controlled by Stephen A. Naftilan). Portions of this article were presented at the biannual meeting of the Society for Research in Child Development, Washington, DC, April 1997.

We gratefully acknowledge the assistance of Heather Coleman, Miryam Longest, Nisha Gottfredson, Dovid Green, and several other Pitzer College undergraduates in the collection and analysis of the data, of J. Steven Reznick in the analyses of the data, and of David Lewkowicz in the interpretation of the data presented herein. Additionally, we are indebted to Dona Bailey and Mark Guterman, who contributed to the preparation of the text and the figures, respectively, in this article. Finally, we thank those parents who allowed their babies to serve as participants in these studies.

Correspondence concerning this article should be addressed to David S. Moore, Pitzer College, 1050 North Mills Avenue, Claremont, CA 91711. E-mail: dmoore@pitzer.edu

1 pretest event (presumably because they expected to see two objects after this event and were surprised to see only one), but they fixated two objects longer after seeing the 2 – 1 pretest event (presumably for analogous reasons).

Unfortunately, a close look at Wynn's (1992) study reveals a confound in its design. In the 1 + 1 pretest event, infants saw one object placed on the stage, then saw a screen raised that occluded their view of the object, and finally saw a second object moved across the stage to a location (presumably) behind the screen. In the 2 – 1 pretest event, infants saw two objects placed on the stage, then saw a screen raised that occluded their view of both objects, and finally saw a hand move behind the screen and emerge with an object that was subsequently removed from the stage. In both cases, the appearance of the display just prior to the introduction of the occluding screen was perfectly correlated with the appearance of the test display containing the supposedly unexpected number of objects. As a result, it is possible that infants spent more time fixating the displays that they did, not because these displays contained an unexpected number of objects, but rather because they were perceptually identical to what was seen just prior to the occlusion of the stage.

Several theorists (Bogartz, Shinsky, & Schilling, 2000; Bogartz, Shinsky, & Speaker, 1997; Cashon & Cohen, 2000; Schilling, 2000) have advanced similar arguments in critiques of other studies using the violation-of-expectation paradigm. Haith (1998) identified a general problem in interpreting the results of studies that measure infants' fixation when he noted that these sorts of paradigms were developed

to address sensory and perceptual questions, not questions of high-level cognitive processing. Many factors affect looking, including . . . familiarity, novelty, recency, predictability, and the time lapse between stimulus exposures . . . investigators who pursue high-level cognitive constructs must . . . fend off every possible perceptual interpretation of differences to entertain default cognitive interpretations. (p. 170)

Consequently, before concluding that infants are capable of computation, researchers must rule out the possibility that the behaviors observed in Wynn-style experiments reflect infants' prior experience viewing the test displays.

Given the extremely large number of studies published since the early 1960s that have indicated that infants prefer to look at novel stimuli, one might reasonably ask why infants would ever prefer familiar stimuli. In fact, Hunter, Ames, and Koopman (1983), after considering the data available up to that point, proposed a model that makes predictions based on factors such as infant age, stimulus complexity, and familiarization time. This model predicts that, all other factors being equal, less mature brains, more complex stimuli, or shorter familiarization times will contribute to familiarity preferences that result, most likely, from infants trying to extract more information from a display that has still not been completely processed. In contrast, more mature brains, simpler stimuli, or longer familiarization times give rise to novelty preferences, most likely because infants in such circumstances seek stimulation that is not yet boring to them. Recognizing this phenomenon, Haith (1998) concluded that data reported by Wynn (1995), rather than conclusively supporting her contention that infants can add and subtract, instead indicate that "there is clearly an effect of varying infants' opportunity to build a strong visual image of the objects" (p. 175).

Cohen and Marks (2002) conducted a series of studies that explored the possibility that infants tested in Wynn's (1992) paradigm preferentially fixate familiar displays. To distinguish between results supportive of Wynn's hypothesis and the familiarity hypothesis just outlined, these researchers attempted to replicate Wynn's findings using "a larger set of test events that includes outcomes with 0 and 3 objects as well as 1 and 2 objects" (p. 189). Such an approach seems reasonable at first, because Wynn's hypothesis predicts that infants' fixations of zero- or three-object test displays would be lengthy (because these would be computationally unexpected after 1 + 1 or 2 – 1 pretest events), whereas the familiarity hypothesis predicts that infants' fixations of these displays would be short (because they are unfamiliar). However, this approach is problematic inasmuch as it does not, in trials otherwise identical to those Wynn designed, directly manipulate any factors known to influence infants' preferences for familiar displays. In their Experiment 3, Cohen and Marks did familiarize infants with one or two objects prior to showing them test displays of zero, one, two, or three objects, but in this experiment, the infants were not exposed to any sort of addition or subtraction pretest events, rendering these conditions incomparable with Wynn's. Moreover, Cohen and Marks's data cannot be directly compared with Wynn's data because infants exposed to two test displays might be expected to behave differently than infants exposed to four test displays; as Carey (2002) commented, "Cohen and Marks' data [might] depend upon infants becoming overloaded by the complexity of the four-choice outcome" (p. 203).

Similarly, Wynn (2002) offered the following two critiques of Cohen and Marks's (2002) methodology:

1. Providing 5-month-old infants with many different test outcomes . . . may increase infants' attention to the perceptual features of, and perceptual differences between, test trials—differences which are superficially evident and therefore easier to process—over the conceptual differences between them (detection of which requires inferential processes which may be the first to suffer under conditions of information overload), . . .

2. In a paradigm [like Cohen and Marks's] in which the majority of test trials depict impossible outcomes . . . infants may quickly learn to "expect the unexpected" . . . [Cohen and Marks's] modification may therefore actually invalidate an expectancy-violation paradigm. (p. 208)

Consequently, the familiarity hypothesis can best be addressed with a study that utilizes only two outcomes—one of which is computationally possible and the other of which is not—and that directly manipulates infants' familiarity with the test displays. The current Experiment 2 is just such a study. The first step in addressing the familiarity hypothesis involves replicating Wynn's (1992) original findings.

## Experiment 1

### Method

*Participants.* New parents in the San Gabriel Valley of Southern California were contacted by mail, and interested parents returned a postcard indicating desire to participate in the study. The final sample consisted of 62 full-term, healthy 5-month-old infants, including 36 girls (mean age = 154 days) and 26 boys (mean age = 152 days). Of these 62 infants, 24 were firstborn, 27 were second born, and 11 were third or later born; the overwhelming majority (93.5%) were living in two-parent, nuclear families at the time of testing. Although specific information on the families' socioeconomic statuses, races, and ethnicities was not collected, the sample was diverse, including families of European American, Asian American,

Latino, and African American backgrounds. The families' diversity was also reflected in the fact that 27.6% of the parents had obtained at least a college education, and 31.7% had obtained no more than a high school education; the remaining 40.7% had educational experiences that fell between these extremes. An additional 75 participants were tested but were eliminated from statistical analyses; this group included 42 girls (out of 78 tested) and 33 boys (out of 59 tested), so male-generated and female-generated data were excluded from analysis in nearly identical proportions. Reasons for elimination from statistical analysis included not fixating the pretest event (described later) sufficiently on more than 50% of the trials (34 participants), fixating the screen for the maximum possible duration (i.e., ceiling) on every trial (2 participants), fussing uncontrollably or becoming distracted (19 participants), refusing to fixate orientation displays before trials and thereby preventing trial onsets (19 participants), or technical malfunction (1 participant). Forty participants were excluded from the addition group, and 35 participants were excluded from the subtraction group.<sup>1</sup> Each participant in the final sample was volunteered by a parent who was present with the infant throughout the procedure, and all infants were screened for a family history of epilepsy and seizure disorder (because our visual stimuli involved flashing lights).

**Design.** Each infant was initially exposed to two different baseline trials in succession. Subsequently, infants were randomly divided into two groups; infants in one of the groups ( $n = 31$ ) were exposed to a series of addition pretest events, whereas infants in the other group ( $n = 31$ ) were exposed to a series of subtraction pretest events. Boys and girls were distributed equally across these experimental conditions. Each participant saw a test display immediately after each pretest event.

**Equipment and stimuli.** The procedure was controlled by custom software run on an IBM-386 clone computer. The software assigned participants to groups, presented the stimuli, timed trials, and collected and stored looking-time data from an experimenter-operated joystick.

The stimuli were presented on a high-resolution computer monitor screen that subtended approximately 14° of visual angle. The stimuli were computer-generated, rectangular random checkerboards (Karmel, 1969) presented on a gray background (see Figure 1). Each checkerboard subtended approximately 1.5° × 2.0° of visual angle. Internal checks subtended approximately 20 min × 16 min of visual angle, and their initial color was either black or white, as determined by the computer's time-seeded random-number generator.

The baseline trials unfolded as follows. In one baseline trial, a single checkerboard emerged from the left side of the monitor screen and moved across the bottom of the screen to the right, coming to rest near the right edge of the screen (Final Position 1); this process took approximately 6 s. When the checkerboard reached this position, it began to flash (see later description). The other baseline trial was similar, except that it involved two checkerboards separated by gray background space subtending 1.5° degrees of visual angle. Both checkerboards emerged from the left side of the screen and moved together to the right for approximately 6 s until the rightmost checkerboard had reached Final Position 1; at this point, given the separation between the two checkerboards, the leftmost checkerboard was in Final Position 2. When the checkerboards ceased to move, both began to flash.

The "addition" pretest event involved a single checkerboard emerging from the left side of the screen and moving across the bottom of the screen to the right, coming to rest at Final Position 1; this process took 3 s. A red video curtain then descended rapidly from the top of the monitor screen, eventually covering the entire right half of the screen (including the checkerboard). Then, in a process that took approximately 5 s, a second, very similar checkerboard emerged from the left side of the screen and moved across the bottom of the screen to the right, following the first checkerboard's path. When this checkerboard encountered the left edge of the video curtain, it gradually appeared to move behind the curtain, disappearing. Then the curtain was raised rapidly to reveal the test stimuli, flashing. Wynn (1992) considered this type of event to represent the operation 1 + 1.

The "subtraction" pretest event was similar to the event just described, except it represented the operation 2 - 1 (Wynn, 1992). As in the preceding

operation, a checkerboard emerged from the left side of the screen and moved to the right, coming to rest at Final Position 1; this process took 3 s. In this operation, however, while the first checkerboard remained static at Final Position 1, a second checkerboard emerged from the left and followed the same path as the first checkerboard, coming to rest at Final Position 2, a process that required only 2 s. At this point, the red video curtain descended, covering the right half of the screen, including both of the checkerboards. Next, in a process that took approximately 5 s, the second checkerboard emerged from behind the left edge of the curtain and, reversing its original trajectory, moved smoothly to the left until it disappeared off of the left edge of the monitor screen. Once again, the curtain was then raised to reveal the flashing test stimuli.

The test displays were conceptualized by Wynn (1992) as possible results of the operations. In the current replication, they consisted either of one flashing checkerboard, located at Final Position 1 (the one-object test display), or of two flashing checkerboards, located at Final Positions 1 and 2, respectively (the two-object test display). Test displays were presented immediately after the conclusion of pretest events, when the curtain was raised. In an effort to increase attention to the displays, a checkerboard was made to flash by changing the color (black or white) of all like-colored sections of the checkerboard at random (but rapid) intervals. This was accomplished by having the computer randomly redetermine the color value of internal checks approximately nine times per second. As a consequence, at any given moment, a checkerboard could appear as it did initially, as a negative image of its initial appearance, or as an entirely black or entirely white square. This had the effect of making the checkerboards appear to flash off and on, producing a compelling but not moving stimulus. Figures 1 and 2 portray events in the addition and subtraction conditions, respectively.

**Apparatus and procedure.** Participants were tested individually at the Claremont Infant Study Laboratory. After arriving at the laboratory, caretakers filled out a questionnaire requesting general information about the child (including health information). Testing was then carried out in a dimly lit, 30 in. × 65 in. enclosure consisting of a front wall supporting a very large, black, fabric-covered panel, black curtains on two sides, and a rear wall painted black. During testing, participants were seated on a caretaker's lap approximately 1 m from the fabric-covered front panel, which contained a square hole at the infant's eye level; this hole revealed a standard 14-in. video graphics array computer monitor screen. The panel also contained one small hole (well above the monitor hole) through which an infant's fixations could be unobtrusively observed and measured and another hole through which a video camera could record the infant's behavior.

<sup>1</sup> The participant attrition rate in this study raised the concern that the participants successfully completing the procedure were different from those who did not in some important way that would render the results less generalizable. Thus, an attrition analysis was undertaken to determine which, if any, of a variety of participant variables could account for participant attrition. Participant variables analyzed included ages of the participants and the parents, birth order and weight, socioeconomic status, duration of prenatal health care, prenatal risk factor exposure, hours spent weekly in day care, parental ratings of quality of infant's previous night's sleep, and time elapsed since last feeding and diaper change. This analysis revealed two effects: on average, participants who completed the procedure had been fed more recently than those who did not complete the procedure,  $t(135) = -2.22, p < .028$ , and those who completed the procedure were slightly younger (153.5 days) than those who did not (156 days),  $t(135) = -2.10, p < .038$ . Thus, it seems that participants who did not complete the procedure did not do so primarily because they were hungry. To the extent that the age effect is reliable, the relatively simple experimental stimuli seem to have bored older infants faster than they bored younger infants. Neither effect indicates that participants who successfully completed the procedure were different from the entire sample of participating infants in any theoretically important way.

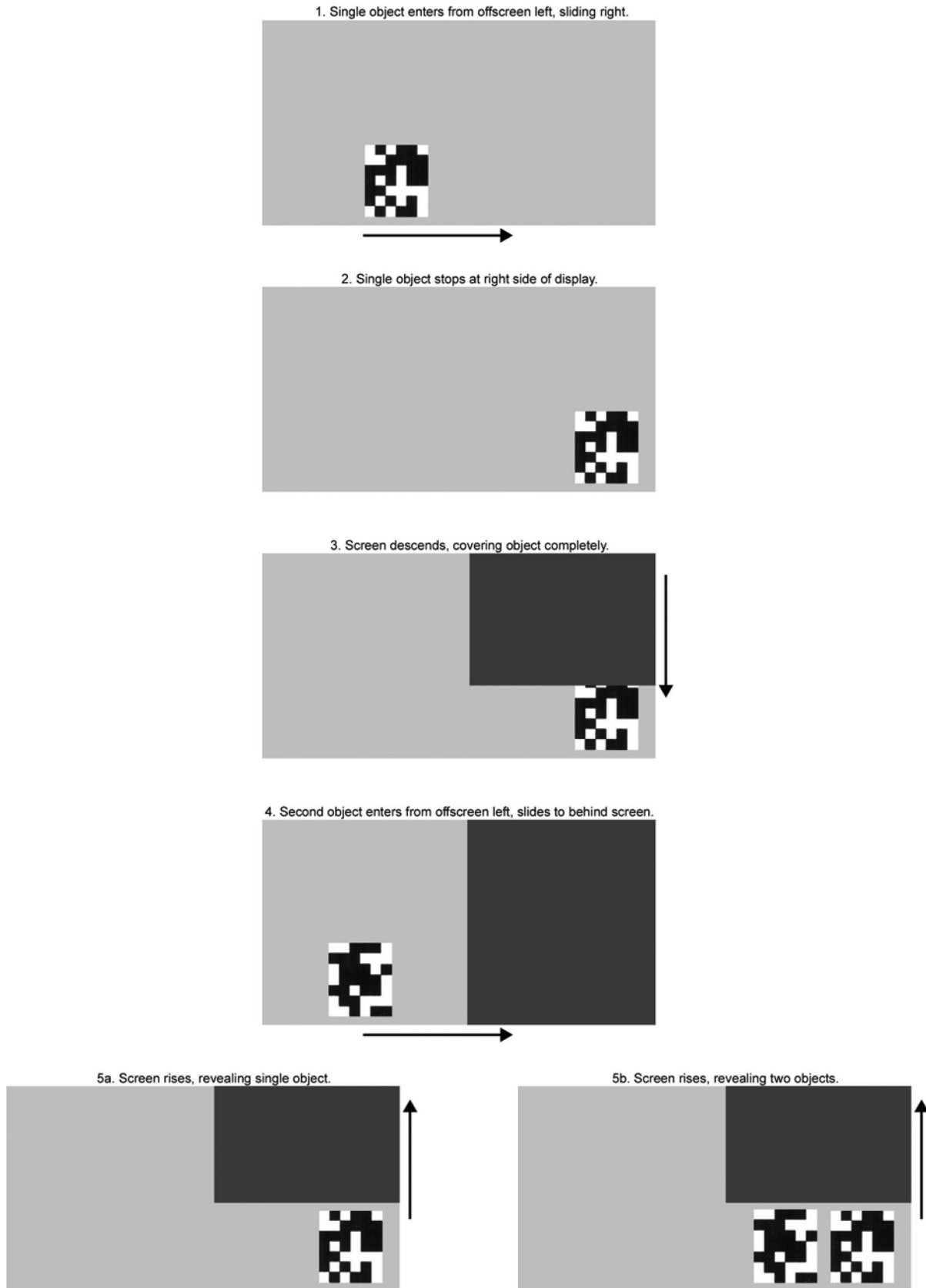


Figure 1. Graphic portrayal of incidents constituting pretest events in the addition condition.

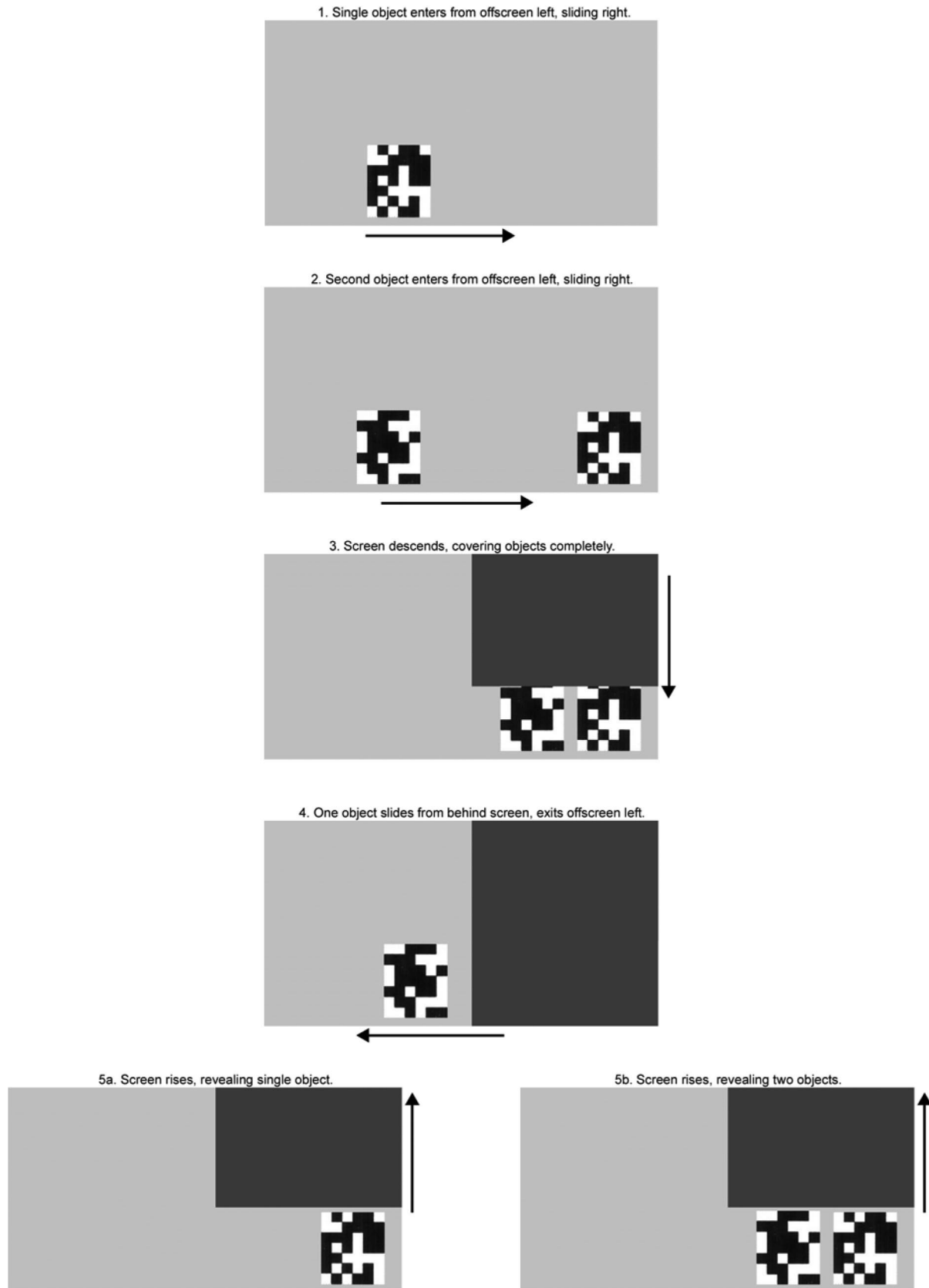


Figure 2. Graphic portrayal of incidents constituting pretest events in the subtraction condition.

Each caretaker was given information concerning the stimuli and procedure and was then asked to orient the infant toward the monitor. Caretakers were instructed to attempt to keep the infant in a quiet, alert state but to refrain from interfering with the infant's perception of the stimuli. Finally, caretakers were instructed to keep their eyes closed throughout the procedure, so as to remain unaware of the stimulus conditions; caretakers' compliance with this request was checked periodically and was observed to be 100%.

The experiment began once the infant was situated and calm. All trials began with the presentation of orientation displays consisting of variously colored ovals flashing in the center of the video screen; these displays served the purpose of initially attracting the infant's attention to the monitor screens. When the observer judged the infant to be attending to the display, a button was pushed to begin the trial.

Each infant first saw both baseline trials; the computer randomly determined which baseline trial was presented first. Fixation duration was recorded once the checkerboard(s) stopped moving and began to flash. Each baseline trial continued until the infant looked away from the screen for 2 s consecutively after having fixated the screen for a cumulative time of at least 2 s, or until the stimuli had flashed for 30 s without having been fixated for a cumulative time of 2 s. If the trial ended without the infant having fixated the stimulus for a cumulative time of at least 2 s, it was repeated (11 out of 62 participants experienced at least 1 repeated baseline trial—or using another metric, 15 out of 139 baseline trials were repeated—under this criterion).

Once both of the baseline displays had been adequately fixated by an infant, he or she saw six additional trials (test trials), each constructed of a pretest event (either addition or subtraction, depending on the group to which he or she had been randomly assigned) seamlessly followed by the presentation of a test display. The computer randomly determined which test display was presented in the first test trial; the type of test display seen in subsequent test trials was alternated until the conclusion of the experiment. Fixation duration was recorded only during presentation of the test displays (the limited processing capacity of the computer that controlled the protocol precluded the collection of looking-time data during the pretest events, as the processor was busy rendering the moving stimuli during these intervals). As in baseline trials, each test display was presented continuously until the infant looked away from the screen for 2 s consecutively after having fixated it for a cumulative time of at least 2 s, or until the stimuli had flashed for 30 s; if the test trial ended without the infant having fixated the test display for at least 2 s, it was repeated (9 test trials out of a total of 381 were repeated under this criterion). Comments about the infants' attention to the pretest events could be inserted into the fixation duration record, so that fixation of test events following unattended pretest events could be appropriately disregarded.

A trained observer recorded each infant's fixations during test trials online by operating a computer-read joystick to indicate fixation and fretting. The observer was unaware of the critical visual experiences of the participants; the distance to the participant prevented the observer from resolving any corneal image enough to allow discrimination of the stimulus conditions, and no other cues were available to the observer as to the number of objects being presented in a given test display. Studies using this experimental set-up and identical equipment have yielded reliability estimates (using two observers) that have ranged from  $r = .96$  to  $r = .98$ .

## Results

Test trials in which the observer deemed fixation of the pretest event to be inadequate were rejected and subsequently repeated; data from rejected trials were excluded from statistical analysis. Inadequate fixation was defined as either looking at the display for less than approximately two thirds of the total duration of the pretest event or not looking during the critical movements of the stimuli. (The critical movements all occurred in temporal proximity to the movements of the red video curtain. Because the observer could tell when the curtain was in place by the reddish glow it produced in the testing room, it was possible to ascertain whether the infant was fixating the display during the critical movements, even as it was impossible to tell if the movements

involved addition to or subtraction from the object or objects already in the display). Data from participants who had 50% or more of their test trials rejected for these reasons were not analyzed further.

Initial analyses were conducted as in Wynn's (1992) study. To compare fixation times during test trials, each participant's average duration of fixation for the two test displays was first computed, and then a preference score was calculated by subtracting each participant's average fixation of the two-object test display from his or her average fixation of the one-object test display; means and preference scores are printed in Table 1. The preference scores of participants in the two groups (addition or subtraction) were then compared with a one-tailed, between-subjects  $t$  test (conducting these analyses as Wynn, 1992, did, baseline comparisons were made with two-tailed  $t$  tests, whereas test comparisons were made with one-tailed  $t$  tests).<sup>2</sup> This analysis of fixation of the test displays revealed the difference that Wynn would have predicted in the two groups' preferences,  $t(60) = 1.75, p < .05$ ; both groups of infants spent more time fixating test displays containing a number of objects that would have been unexpected if they had formed expectations on the basis of the mathematical operations that some theorists think the pretest events used in this study portray. Specifically, infants exposed to the addition pretest event spent more time fixating one- than two-object test displays, whereas infants exposed to the subtraction pretest event spent more time fixating two- than one-object test displays. Such a difference was not detected across groups in baseline trials,  $t(60) = 0.76, p > .45$ .<sup>3</sup>

A more common way of analyzing data such as these is through an analysis of variance (ANOVA); among other advantages, such an approach resolves the problems associated with deciding whether one- or two-tailed  $t$  tests are more appropriate in situations like the current one. Thus, these data were also subjected to a four-way, 2 (trial type: baseline vs. test)  $\times$  2 (test display: one object vs. two objects)  $\times$  2 (group: addition vs. subtraction)  $\times$  2

<sup>2</sup> Wynn's use of one-tailed  $t$  tests in her 1992 letter to *Nature* remains controversial, because the data collected in most violation-of-expectation studies should arguably be evaluated with two-tailed tests. Two-tailed tests are advisable in these cases because there are invariably good theoretical reasons for predicting a preference for either of the two displays presented during test trials; in the current study, it would have been theoretically reasonable to predict that an infant in the addition condition would prefer one-object displays (because they represent the unexpected result of a putative mathematical operation) or to predict that an infant in this condition would prefer two-object displays (because of the perceptual novelty of such displays). Because the current experiment was an attempt to replicate Wynn (1992), it seemed only fair in this report to use the same statistical analyses she used.

<sup>3</sup> Wynn (1992) excluded from her analyses infants who had more than a 10-s preference for one of the test displays during baseline trials. Because application of this criterion to the current data set would have required exclusion of 11 participants (i.e., 17.7% of the available data), a univariate analysis was conducted on baseline difference scores. This analysis revealed that these scores were normally distributed (Shapiro-Wilk  $W = .93, p < .004$ ), and were centered around a mean of 1.13 s ( $SD = 7.98$  s). Using a very conservative definition of an outlying difference score (Tabachnick & Fidell, 1983), we excluded from a reanalysis of the data any participant whose baseline difference score was more than 2  $SD$  from the mean. The pattern of results generated by the remaining 57 participants was the same as that reported in the main text: Infants in the subtraction group preferred the two-object test display, whereas infants in the addition group preferred the one-object test display,  $t(55) = -1.65, p < .053$ . There continued to be no difference across groups in baseline fixation scores,  $t(55) = -0.52, p > .60$ .

(sex: male vs. female) mixed-design ANOVA, in which trial type and test display were repeated measures variables; raw fixation time was the dependent measure in this analysis. If, as Wynn predicted, participants' fixations of one- or two-object test displays are influenced by their experience seeing addition or subtraction pretest events, the expected result of this analysis would be a three-way interaction involving trial type, test display, and group (the use of trial type as a factor in this analysis simply allows individual infants' fixations during test trials to be evaluated in light of any preexisting baseline preferences for one- or two-object displays<sup>4</sup>). Such a three-way interaction would indicate that the addition and subtraction groups' preferences for the one- or two-object test displays changed from baseline to test in differing ways after they experienced the addition and subtraction pretest events. In fact, this analysis revealed only a four-way, Trial Type × Test Display × Group × Sex interaction,  $F(1, 58) = 4.62, p < .04$ .

Examination of the cell means contributing to this four-way interaction suggested that while the girls' means fell into the predicted pattern, the boys' did not. Because this sex difference was not predicted, post hoc tests were conducted, first on the boys' and then on the girls' data, considered independently. The dependent measure in these post hoc analyses was the difference between the participants' preference scores during the baseline trials and their preference scores during the test trials; these difference scores, the averages of which are depicted in Figure 3, represent how infants' preferences for the one- or two-object displays changed from baseline to test after seeing the addition or subtraction pretest events. The difference scores generated by the partic-

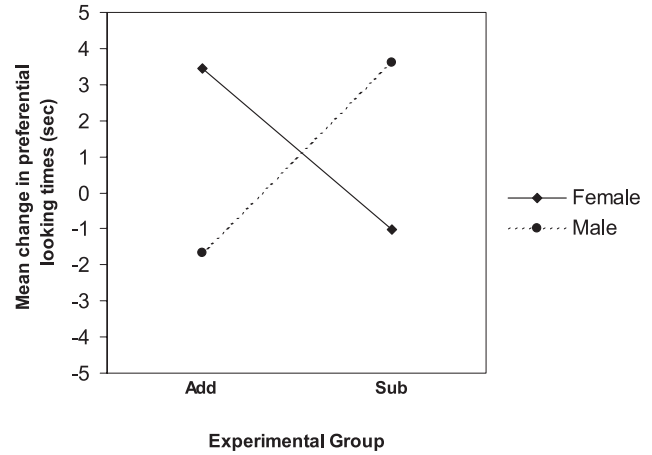


Figure 3. Experiment 1: Mean change from baseline to test in preference for one-object test displays as a function of experimental condition and participant sex. Add = addition; Sub = subtraction.

Table 1  
Experiment 1: Mean Fixation Times and Mean Preferences for One Object Over Two Objects (in Seconds)

Trial type and group	Mean fixation				Preference: one object	
	One object		Two objects		M	SD
	M	SD	M	SD		
All participants						
Baseline						
Addition	12.73	9.24	13.09	10.25	-0.36	8.01
Subtraction	13.51	10.07	15.41	9.92	-1.90	8.01
Test						
Addition	12.08	5.90	11.13	6.28	0.95	4.24
Subtraction	11.88	6.65	12.87	6.51	-0.99	4.41
Female participants						
Baseline						
Addition	10.15	6.61	11.54	8.76	-1.39	7.42
Subtraction	13.61	9.62	14.11	9.44	-0.50	6.40
Test						
Addition	11.52	6.36	9.45	5.46	2.07	3.84
Subtraction	12.10	7.26	13.61	7.06	-1.51	5.28
Male participants						
Baseline						
Addition	16.30	11.30	15.24	12.05	1.06	8.86
Subtraction	13.37	11.06	17.22	10.66	-3.85	9.76
Test						
Addition	12.86	5.34	13.47	6.79	-0.61	4.43
Subtraction	11.57	5.98	11.83	5.79	-0.26	2.85

ipants in the two groups (addition or subtraction) were compared with two-tailed, between-subjects *t* tests. The boys in the addition condition did not differ significantly from the boys in the subtraction condition on these scores,  $t(24) = -1.55, p > .13$ , two-tailed; likewise, the girls in the addition and subtraction conditions did not differ significantly from each other,  $t(34) = 1.51, p > .14$ , two-tailed. Consequently, when their data were considered in isolation, neither our male nor our female participants behaved in accordance with Wynn's (1992) predictions. Nonetheless, when the girls' data were considered in the context of the boys' data (as they were in the ANOVA reported above), the pattern of means that the girls produced can be seen to differ significantly from the pattern of means that the boys produced.

Specifically, female participants in the addition condition shifted from a slight baseline preference<sup>5</sup> for two-object displays

<sup>4</sup> No such preexisting baseline preferences were detected in this data set. Using baseline raw fixation times as the dependent data, we conducted an additional three-way, 2 (test display: one object vs. two objects) × 2 (group: addition vs. subtraction) × 2 (sex: male vs. female) mixed-design ANOVA, in which test display was the repeated measures variable. This analysis revealed no significant main effects or interactions.

<sup>5</sup> Both here and in the *Results* section following the description of Experiment 2, our use of the word *preference* should not be taken as indicating that the average fixations of infants in the particular experimental cells being discussed differed significantly from zero. On the contrary, significant statistics typically emerged from the analyses reported only when the behaviors of infants in one group were considered relative to the behaviors of infants in other groups, not when considered relative to zero (no preference). Consequently, these groups cannot be considered to have had reliable preferences for one display or the other. However, if one hypothetical group of infants looks at Display A insignificantly more than at Display B, whereas another hypothetical group looks at Display B insignificantly more than at Display A, a significant difference between the two groups can still be revealed in statistical analyses. Because it is not apparent how to clearly discuss the significantly different behaviors of these two hypothetical groups without referring to one group's preference for Display A relative to the other group's preference for Display B, we have used the word *preference* throughout our text for ease of exposition.

to a preference in test trials for one-object displays; in contrast, male participants in this condition shifted from a slight baseline preference for one-object displays to a very slight preference in test trials for two-object displays (in Figures 3 and 4, shifts from baseline to test toward a preference for one-object displays are indicated by positive data points, and shifts from baseline to test toward a preference for two-object displays are indicated by negative data points). Although infants in the subtraction condition, on average, always looked longer at two-object displays than at one-object displays—a finding that likely reflects infants' established tendency to look longer at displays with higher contour densities (Fantz & Fagan, 1975; Karmel, 1969)—female participants in this condition shifted from a very slight baseline preference for two-object displays to a stronger preference for such displays in test trials (reflecting *decreasing* looking at one-object displays, hence a negative data point in Figure 3); in contrast, male participants shifted from a baseline preference for two-object displays to an extremely weak preference for such displays in test trials (reflecting *increasing* looking at one-object displays, hence a positive data point in Figure 3). These preferences in individual cells are not statistically significant when considered in isolation, however the four-way interaction reported above indicates that the pattern of preferences produced by the male and female infants in the addition and subtraction conditions was unlikely to have occurred by chance. Although the data provided by the female participants alone did not generate the significant three-way interaction that Wynn's (1992) hypothesis predicts, the data provided by the male participants did not even trend in the predicted direction. Consequently, it appears that the replication of Wynn's findings indicated by the initial analyses reported above is due to the performance of the female participants.

### Discussion

An analysis identical to that used by Wynn (1992) revealed that infants in the current study, while exhibiting no reliable preference for one versus two checkerboards in baseline trials, nevertheless responded differently to the test displays after seeing the two different kinds of pretest events. Specifically, infants' preferences

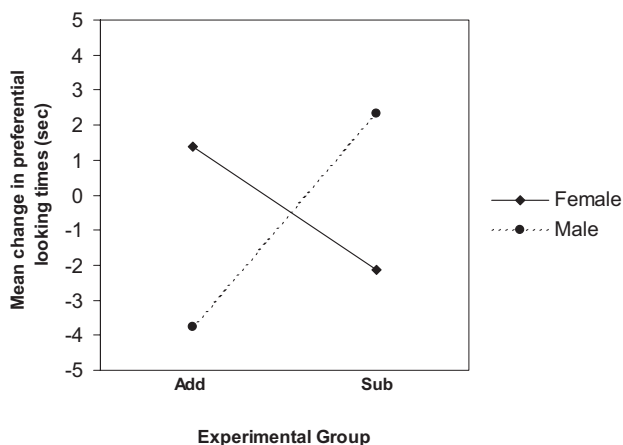


Figure 4. Experiment 2: Mean change from baseline to test in preference for one-object test displays as a function of experimental condition and participant sex. Add = addition; Sub = subtraction.

for one- over two-object displays were significantly greater in the addition than in the subtraction condition, and conversely, their preferences for two- over one-object displays were significantly greater in the subtraction than in the addition condition.<sup>6</sup> Following Wynn, one possible interpretation of these data is that these infants spent more time looking at test displays that contained the number of objects that would have been surprising if they had formed expectations on the basis of the results of computations performed while watching the pretest events. The role of sex in this effect, as revealed in the ANOVA reported above, is considered later in the General Discussion.

The data collected in this study, when analyzed with the same statistical techniques Wynn (1992) reported using in her original study, suggest that the phenomenon reported in that study is replicable even when the methods used to elicit it differ notably from those used by Wynn and other researchers who have replicated her work (e.g., Simon, Hespos, & Rochat, 1995; Uller et al., 1999). In particular, the current study used as stimuli two-dimensional pictorial representations of abstract shapes manipulated by a computer, whereas Wynn used three-dimensional, meaningful objects manipulated by people. Although it is known that infants in studies of numerosity perception can treat video displays of abstract shapes as they do displays of real objects (van Loosbroek & Smitsman, 1992), the current finding suggests that infants responded to the abstract shapes in this experiment as they responded to the real objects in Wynn's protocol, a precondition for the clear interpretation of the results of our Experiment 2, which used the same abstract stimuli.

The hypothesis that infants' behavior in this and other Wynn-style experiments can be explained as a function of infants' familiarity with the test displays is best explored with a methodology that changes nothing in the protocol other than infants' familiarity with the test displays. In Experiment 2, infants were allowed to view the display seen just prior to the appearance of the occluding screen—the same displays as seen in test trials—for as long as they needed to become disinterested in viewing it any longer. Although such a manipulation ensured that these infants were familiarized with these displays, it in no way altered the computational aspects of the pretest events. Consequently, if infants' behaviors in these sorts of protocols reflect familiarity with the test stimuli, infants who experience 1 (familiarization with 1) + 1 should look longer at the two-object than at the one-object test display, because they are bored with looking at one, and infants who experience 2 (familiarization with 2) – 1 should look longer at the one-object than at the two-object test display, because they are bored with looking at two. In contrast, if infants' behaviors in these sorts of

<sup>6</sup> Because our data would still have been theoretically interesting if they had revealed the opposite of what Wynn (1992) predicted, it is worth considering how they would have fared in a two-tailed test. In fact, the results of the current experiment would not have been significant if a two-tailed test had been used. Although the results of Wynn's Experiment 1 would have been significant even with the use of a two-tailed test, the results of her Experiment 2—her own attempt to replicate her Experiment 1—would not have been. Thus, the use of the more conservative two-tailed test would have rendered both Wynn's Experiment 2 and the current Experiment 1 failures to replicate Wynn's original experiment, even though the current experiment involved nearly twice the number of experimental participants as Wynn's first study, and so had more power to detect differences.



protocols reflect computation, infants who experience 1 (familiarization with 1) + 1 should look longer at the one-object than at the two-object test display, because they expect 1 + 1 to yield 2 and are surprised and intrigued by a result of 1; analogously, infants who experience 2 (familiarization with 2) – 1 should look longer at the two-object than at the one-object test display, because they expect 2 – 1 to yield 1, and are surprised and intrigued by a result of 2. If infants are really responding to the computational aspects of the pretest events, their responses to the test displays should not be influenced by their familiarity with those displays.

## Experiment 2

### Method

**Participants.** The final sample, recruited and screened as in Experiment 1, consisted of 89 full-term, healthy 5-month-old infants, including 47 girls (mean age = 156 days) and 42 boys (mean age = 155 days). Of these 89 infants, 41 were firstborn, 27 were second born, and 21 were third or later born, and 88.8% were living in two-parent, nuclear families at the time of testing. This sample was ethnically, socioeconomically, and educationally as diverse as the sample studied in Experiment 1, with 29.5% of the parents having obtained at least a college education, 38.6% having obtained no more than a high school education, and the remaining 31.9% having had educational experiences that fell between these extremes. An additional 64 participants were tested but were eliminated from statistical analyses; this group included 36 girls (out of 83 tested, or 43%) and 28 boys (out of 70 tested, or 40%), so male-generated and female-generated data were excluded from analysis in nearly identical proportions. Reasons for elimination from statistical analysis included not sufficiently fixating the pretest event on more than 50% of the trials (39 participants), fussing uncontrollably (24 participants), or having a mother who refused to remain unaware of the experimental stimuli during the test (1 participant). Thirty-five participants were excluded from the addition group, and 29 participants were excluded from the subtraction group.

**Design.** As in Experiment 1, each infant was randomly assigned to one of two groups. Between the same baseline and test events presented in Experiment 1, 21 boys and 25 girls were exposed to an addition pretest event, and 21 boys and 22 girls were exposed to a subtraction pretest event.

**Equipment and stimuli.** The stimuli and equipment were identical to those used in Experiment 1, with one exception. In both the addition and subtraction pretest events, a familiarization period preceded the lowering of the video curtain. The familiarization period lasted until the infant looked at the stimuli for at least 2 s (cumulatively) and subsequently looked away for 2 s consecutively (the familiarization period also ended either if 20 s elapsed without a cumulative 2 s of fixation having been elicited or if 30 s elapsed without the infant having looked away for 2 s consecutively).

Thus, as in Experiment 1, the addition pretest event began with one checkerboard emerging from the left side of the screen and moving to Final Position 1. At this point in Experiment 2, however, the checkerboard began to flash while remaining stationary, allowing the infant to become familiar with the stimulus. At the end of the familiarization period when the infant had been looking away from the display for 2 s, a digitized bell-like noise drew the infant's attention back to the screen, at which point the curtain descended, covering the right half of the screen, including the checkerboard. The trial then proceeded as in Experiment 1, with a second checkerboard emerging from the left and ultimately moving behind the curtain. Once again, as in Experiment 1, the curtain was raised at this point to reveal the test stimuli, flashing.

The subtraction pretest event, likewise, was similar to that seen in Experiment 1. After moving across the screen to Final Positions 1 and 2, respectively, two checkerboards began to flash while stationary, allowing the infant to become familiar with the display. At the conclusion of the familiarization period, the digitized bell-like noise drew the infant's attention back to the screen, at which point the curtain descended, covering the right half of the screen, including both checkerboards. As in Experiment 1,

one checkerboard then emerged from behind the curtain and moved off of the left edge of the monitor screen, whereupon the curtain was raised to reveal the flashing test stimuli.

**Apparatus and procedure.** The apparatus and procedure were identical to those used in Experiment 1. As in Experiment 1, each infant first saw both baseline events, and fixation duration was recorded once the checkerboard(s) stopped moving and began to flash (the computer randomly determined which baseline event was presented first). Likewise, as in Experiment 1, if a baseline trial ended without the infant having fixated the stimulus for at least 2 s cumulatively, it was repeated (20 out of 89 participants experienced at least one repeated baseline trial—or figured by trial, 30 out of 208 baseline trials were repeated—under this criterion).

Once both of the baseline events had been adequately fixated by an infant, he or she saw six additional trials, each constructed of either an addition or subtraction pretest event seamlessly followed by a test event (again, the computer randomly determined which test event was presented in the first postbaseline trial, and presentations of the one- or two-object test events were alternated thereafter). As in Experiment 1, fixation duration was recorded after the occluding curtain had been raised, and a trial was repeated if an infant fixated the stimulus for less than a cumulative 2 s (10 of 89 participants experienced 13 repeated trials—out of a total of 547 test trials—under this criterion).

### Results

Initial analyses were conducted as in Experiment 1, and identical criteria were used to exclude participants from the analyses. After calculating each participant's average duration of fixation of the two test displays, the average fixations of the two-object test display were subtracted from the average fixations of the one-object test display, and the resulting preference scores of participants in the two groups (addition or subtraction) were compared with the same test used by Wynn (1992), namely a one-tailed, between-subjects *t* test (see Table 2). This test failed to reveal a reliable difference between the two groups' fixation preferences,  $t(87) = -1.31, p > .09$ . Likewise, a two-tailed, between-subjects *t* test revealed no difference across groups in fixation during baseline trials,  $t(87) = -0.27, p > .79$ .

As in the analysis of the Experiment 1 data, the data from Experiment 2 were then subjected to a four-way, 2 (trial type: baseline vs. test)  $\times$  2 (test display: one object vs. two objects)  $\times$  2 (group: addition vs. subtraction)  $\times$  2 (sex: male vs. female) mixed-design ANOVA, in which trial type and test display were repeated measures variables and raw fixation time was the dependent measure.<sup>7</sup> As before, if participants' fixations of one- or

<sup>7</sup> As in Experiment 1, we investigated our baseline raw fixation time data by using a three-way, 2 (test display: one object vs. two objects)  $\times$  2 (group: addition vs. subtraction)  $\times$  2 (sex: male vs. female) mixed-design ANOVA, in which test display was the repeated measures variable. This analysis revealed a main effect of sex,  $F(1, 85) = 5.43, p < .03$ , which was qualified by a three-way Test Display  $\times$  Group  $\times$  Sex interaction,  $F(1, 85) = 6.39, p < .02$ . This interaction is uninterpretable, because infants were randomly assigned to the addition and subtraction groups, and infants in these groups had not yet had any differential experiences when they encountered the baseline trials. Consequently, this finding likely reflects random fluctuations in the looking behaviors of these groups of infants. Although the analysis of the baseline data collected in Experiment 1 revealed no analogous effects, the appearance of this anomalous effect in our Experiment 2 data suggests that our a priori decision to use trial type (baseline vs. test) as a variable in all of our primary analyses was a good one; this approach effectively allowed us to evaluate our test data in light of individual infants' preexisting (baseline) preferences for one- or two-object displays.

Table 2  
*Experiment 2: Mean Fixation Times and Mean Preferences for One Object Over Two Objects (in Seconds)*

Trial type and group	Mean fixation				Preference: one object	
	One object		Two objects		<i>M</i>	<i>SD</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
All participants						
Baseline						
Addition	14.92	9.75	15.27	10.03	-0.35	10.74
Subtraction	14.61	9.99	14.40	10.22	0.21	9.09
Test						
Addition	8.30	4.39	9.62	4.80	-1.32	5.84
Subtraction	8.85	6.20	8.59	4.52	0.26	5.47
Female participants						
Baseline						
Addition	12.23	9.42	14.98	10.64	-2.75	10.09
Subtraction	13.37	10.59	10.62	8.92	2.75	9.01
Test						
Addition	8.54	4.91	9.90	4.32	-1.36	5.22
Subtraction	8.55	6.68	7.92	4.55	0.63	6.49
Male participants						
Baseline						
Addition	18.13	9.35	15.62	9.51	2.51	11.01
Subtraction	15.91	9.40	18.36	10.18	-2.45	8.58
Test						
Addition	8.03	3.77	9.29	5.42	-1.26	6.63
Subtraction	9.17	5.81	9.29	4.49	-0.13	4.28

two-object test displays were affected by the pretest events, the expected result of this analysis would be a three-way interaction involving trial type, test display, and group. Such an interaction would indicate that the addition and subtraction groups' preferences for the one- or two-object test displays changed from baseline in differing ways after the infants in these groups experienced the addition and subtraction pretest events. However, whereas Wynn (1992) would have predicted that addition group infants should increase their preference for the one-object display and that subtraction group infants should increase their preference for the two-object display, the familiarity hypothesis leads to the opposite prediction, with addition group infants shifting to an increased preference for the two-object display and subtraction group infants shifting to an increased preference for the one-object display.

In fact, this analysis revealed a main effect of trial type,  $F(1, 85) = 43.09, p < .001$ , a Trial Type  $\times$  Sex interaction,  $F(1, 85) = 4.64, p < .04$ , and a Test Display  $\times$  Group  $\times$  Sex interaction,  $F(1, 85) = 4.77, p < .04$ . All of these findings were qualified by a four-way, Trial Type  $\times$  Test Display  $\times$  Group  $\times$  Sex interaction,  $F(1, 85) = 4.80, p < .04$ . Examination of the cell means contributing to this interaction suggested that while the girls' means fell into the three-way pattern predicted by Wynn (1992), the boys' means fell into the three-way pattern predicted by the familiarity hypothesis. Because this sex difference was not predicted, separate post hoc tests were conducted as in Experiment 1, first on the boys' and then on the girls' data. Once again, the dependent measure used in these analyses was the difference between the participants' preference scores during the baseline trials and their preference

scores during the test trials; averages of these difference scores, presented as a function of sex and experimental condition, are depicted in Figure 4. As in Experiment 1, the difference scores of the boys in the addition condition did not differ significantly from those of the boys in the subtraction condition,  $t(40) = -1.83, p > .07$ , two-tailed. Likewise, the difference scores of the girls in the addition condition did not differ significantly from those of the girls in the subtraction condition,  $t(45) = 1.22, p > .23$ , two-tailed. Consequently, when the boys' and girls' data were considered in isolation, neither Wynn's predictions nor the predictions of the familiarity hypothesis gain support. However, when the data from the boys and girls were considered relative to each other (as they were in the ANOVA just reported), the pattern of means obtained for the girls was found to differ from the pattern of means obtained for the boys.

Specifically, female participants in the addition condition shifted from a baseline preference for two-object displays to a weaker preference for such displays, reflecting a slight proportional increase in looking at one-object displays between the baseline and test trials; in contrast, male participants in this condition shifted from a baseline preference for one-object displays to a slight preference in test trials for two-object displays. Conversely, female participants in the subtraction condition shifted from a baseline preference for one-object displays to a much weaker preference for such displays in the test trials, reflecting a proportional increase from baseline to test in looking at two-object displays; male participants in this condition shifted from a baseline preference for two-object displays to an extremely weak preference for such displays in the test trials, reflecting a proportional increase from baseline to test in looking at one-object displays. As in the analyses of the data from Experiment 1, these preferences in individual cells are not statistically significant when considered in isolation, however the four-way interaction indicates that the differing patterns of preferences produced by the boys and girls in the two experimental conditions were unlikely to have occurred by chance. It seems likely that the absence of statistically significant findings in the initial  $t$  tests reported above reflects the fact that boys and girls responded in opposite ways to exposure to the different types of pretest events.

## General Discussion

Female infants in both of these studies exhibited no reliable baseline preference for one- or two-object displays. Nevertheless, in both studies, a significant interaction involving sex suggested that relative to male infants, female infants who saw a subtraction sequence alleged to represent the operation  $2 - 1$  subsequently increased their looking at two-object displays, whereas female infants who saw an addition sequence alleged to represent the operation  $1 + 1$  subsequently increased their looking at one-object displays.<sup>8</sup> Interpreting these data as Wynn (1992) interpreted hers

<sup>8</sup> Note that significant effects were revealed by the analyses of our Experiment 1 and Experiment 2 data sets only when our infants' fixations during test trials were evaluated in light of their baseline fixations (i.e., each of the significant interactions revealed by the ANOVAs reported in the main body of our article involved the trial type variable); significant effects were not revealed when infants' fixations during test trials were considered in isolation. This pattern of results is unusual inasmuch as experimenters who use violation-of-expectation paradigms typically analyze infants' performances on test trials alone and find significant effects when doing so.

would entail positing that in both cases, female infants increased their fixation of displays that violated expectations that were formed in the wake of computations elicited by the pretest events. Specifically, female infants would be understood to have behaved as they did in the  $1 + 1$  condition because a one-object outcome was unexpected (since  $1 + 1 = 2$ , not 1). Likewise, female infants in the  $2 - 1$  condition would be understood to have behaved as they did because a two-object outcome was unexpected (since  $2 - 1 = 1$ , not 2). Several theorists, Haith (1998) in particular, have argued cogently against rich interpretations of this sort, so it is not appropriate to conclude that the behavior of the female infants in these experiments actually reflects computation, particularly given the existence of nonnumerical accounts that could explain these data equally well (Simon, 1997, 1998) and the fact that the female infants' data alone (i.e., considered in isolation from the male infants' data) did not achieve statistical significance in either of the current experiments.

Nonetheless, the data from Experiment 2 do bear on the evaluation of the familiarity hypothesis, one of several hypotheses competing to explain Wynn's (1992) original data. Although the girls in the addition condition in this experiment were familiarized with one-object displays, their fixation of relatively unfamiliar two-object displays during test trials did not increase relative to their baseline fixation of such displays. Likewise, although the girls in the subtraction condition in this experiment were familiarized with two-object displays, their fixation of relatively unfamiliar one-object displays during test trials did not increase relative to their baseline fixation of such displays. This result is not consistent with the model proposed by Hunter et al. (1983) and suggests that 5-month-old girls' behavior in Wynn-style experiments cannot be explained with reference to their familiarity with the test stimuli.

The same pattern of results did not characterize the behavior of the boys in the current studies. Interpretation of the data generated by male infants in Experiment 2 is hindered by the fact that male infants in Experiment 1 behaved differently than either the female infants in Experiment 1 or the infants in Wynn's (1992) original studies. Nonetheless, in spite of this failure to replicate Wynn's findings, the data produced by the boys in Experiment 2 are in accord with the familiarity hypothesis outlined in the introduction. Infants in the addition condition in Experiment 2 were well familiarized with a one-object display, and in line with Hunter et al.'s (1983) model, male infants subsequently exhibited a marginally increased preference for two-object displays, even though a display containing two objects would have been expected—and so relatively uninteresting—if these infants were doing the computations they have been alleged to do in these sorts of studies. Likewise, infants in the subtraction condition in this experiment were well familiarized with a two-object display, and although our male infants continued to look longer at two-object test displays by a small fraction of a second, their time spent looking at one-object displays during the test trials increased marginally relative to baseline, even though the two-object displays were the mathematically unexpected result of the  $2 - 1$  pretest event. Thus, the trends in the data produced by the boys in Experiment 2 suggest that they were influenced by their familiarity with the test displays; these trends offer no support for the contention that the boys formed expectations on the basis of computations performed while watching the pretest events.

Taken together, the results of the current studies indicate that familiarization with the test displays did not lead 5-month-old girls

to change how they responded to addition and subtraction pretest events; even with increased perceptual experience with the test displays, these infants' behaviors continued to be more in line with Wynn's (1992) hypothesis than the familiarity hypothesis. Although the behavior of the boys in Experiment 2 is more difficult to interpret given that the boys in Experiment 1 did not behave as Wynn would have predicted, their responses were consistent with the hypothesis that familiarization with the test displays can increase looking at relatively novel displays, even if the alternative, familiar displays would have been more surprising to a perceiver capable of computation.

The finding that boys and girls reacted differently to the experimental displays in both of the current studies was unexpected but not inexplicable. While Wynn (1992) did not report such a sex difference, an interaction involving sex was reported in a study of 10- to 12-month-olds' discrimination of number of elements contained in visual displays (Strauss & Curtis, 1981); thus, a sex effect is not unprecedented in studies of this domain. In general, sex effects are rather common in the developmental literature (e.g., Antell & Keating, 1983; Caron & Caron, 1969; Cohen, Gelber, & Lazar, 1971; Creighton, 1984; Friedman, Bruno, & Vietze, 1974; Grattan, de-Vos, Levy, & McClintock, 1992; Humphrey & Humphrey, 1987; Tighe & Powlison, 1978); typically, these differences have been explained in terms of the faster rate of neurological maturation known to characterize female versus male brain development (Shucard, Shucard, & Cummins, 1981; Tanner, 1978). For example, Singer, Westphal, and Niswander (1968) found that for 8-month-old infants, "significant sex differences were found on all scales [and that] males had significantly poorer scores" on four of the five scales used in the study. In addition, female infants have been found to be more mature than male infants on particular measures of visual acuity (Held, Shimojo, & Gwiazda, 1984), in onset of binocular vision (Bauer, Shimojo, Gwiazda, & Held, 1986; Held, Bauer, & Gwiazda, 1988), and in the shortening of peak latencies of pattern-reversal visual evoked potentials (Malcolm, McCulloch, & Shepherd, 2002). These differences could have a variety of consequences, including differences in rates of processing or habituation, differences in motivation to attend to stimuli, and/or differences in fussing thresholds, among others.

Although any of the various differences between male and female infants could be responsible for the effects found in the current experiments, a particularly interesting possibility—however speculative—is that male infants, because of their relative immaturity early in life, are more likely than female infants to respond to the perceptual characteristics of a display; female infants, in contrast, might be quicker to apprehend a particular display with cognitive skills that are beginning to emerge at this point in the course of normal development. As early as 1966, Kagan, Henker, Hen-Tov, Levine, and Lewis concluded that a reasonable interpretation of their data was that "girls are maturationally ahead of the boys" (p. 528). Given the widely recognized delay in male infants' development relative to female infants' development, it is perhaps less surprising that boys and girls behaved as they did in the current experiments.

In light of these results, a reconsideration of some of the claims made in articles on this phenomenon seems warranted. First, although Wynn (1996) wrote, "infants' enumeration mechanism does not operate over low-level perceptual information" (p. 169), the current results suggest that at least for 5-month-old boys in this protocol, perceptual information might influence behavior in very

important ways. Thus, as Clearfield and Mix (1999) wrote, our results, like theirs, “highlight the importance of perception in infants’ numerical development” (p. 410). In support of concerns recently voiced by Haith (1998), Cohen and Marks (2002), Mix (2002), and others, the data produced by the boys in Experiment 2 highlight the importance of being cautious when “attributing sophisticated cognitive processes to young infants when simpler processes will suffice” (Cohen & Marks, 2002, p. 200); in fact, the current data suggest that in some cases, behavior that has looked to some investigators like genuine computational ability might actually be parsimoniously explained as a function of familiarity of the test displays. On the other hand, the current data weigh against the possibility that familiarity with the test displays has an important influence on the behaviors of 5-month-old girls. As data accumulate that rule out various alternative hypotheses to the one originally put forward by Wynn (1992), the possibility that some infants have developed relatively sophisticated numerical abilities by 5 months of age must be considered.

At the moment, it continues to seem premature to conclude that young infants actually have developed such abilities, the pronouncements of popular textbooks for relatively uncritical undergraduates notwithstanding (e.g., Feldman, 1998). The data produced by some studies of infants’ representations of number can be plausibly explained by hypotheses that do not require attributing complex numerical abilities to young babies. Although Carey (2002) wrote, “there is abundant convergent evidence from other paradigms . . . [for numerical processing in young infants] . . . including evidence not subject to a familiarity preference reinterpretation” (p. 203), it remains possible to understand the studies she cites without positing high-level cognitive processes. For example, she cites Spelke, Kestenbaum, Simons, and Wein (1995) as providing evidence for infant numerical representations that cannot be explained by the familiarity hypothesis, but these authors themselves noted that their Experiments 1 and 2 generated “weak and unstable differences between the experimental and control conditions” (p. 127), and their Experiments 3 and 4 generated null results (which never provide conclusive evidence of anything), so it is not at all clear that these results provide firm evidence for infants’ numerical abilities. Likewise, findings such as those reported by Koechlin, Dehaene, and Mehler (1997), in which an attempt to replicate Wynn’s (1992) findings by using a moving stage yielded significant results only in the 2 – 1 condition, can easily be explained with reference to infants’ known preference for visual displays with greater contour densities (Fantz & Fagan, 1975; Karmel, 1969); unless infants can *also* be shown to prefer one object after a 1 + 1 pretest event, infants’ preference for two objects after a 2 – 1 pretest event could reflect a simple preference for displays with more objects, not numerical processing.

That said, the finding that 5-month-old girls in the current experiments were not influenced in a statistically detectable way by pretest familiarization with the test displays now takes its place among other findings that have yet to be explained with reference only to low-level, perceptual processes. For example, the finding that 5-month-olds spend more time fixating mathematically unexpected visual displays after experiencing bimodal pretest events such as 1 object + 1 tone, or 1 object + 2 tones (Kobayashi, Hiraki, Mugitani, & Hasegawa, 2004), *if replicated*, would seem to offer some support for the contention that infants can perform simple computations, support that would not be undermined by extant proposals that explain apparent computation without refer-

ence to genuinely cognitive processes. Similarly, reports of infants’ enumeration of actions (Wynn, 1996) and their tracking and enumeration of collections of objects (Chiang & Wynn, 2000; Wynn, Bloom, & Chiang, 2002) lend support to the contention that 5- to 8-month-olds possess some numerical knowledge; these results, too, are not easy to explain with hypotheses that invoke only perceptual processes. Of course, as Haith (1998) pointed out, it remains the case that cognitive interpretations of these sorts of findings ought not be accepted until all possible perceptual interpretations have first been tested—including, possibly, some interpretations that have not yet been proposed—but the collection of additional data in studies designed to disconfirm the contention that infants are capable of numerical processing will, in time, bring some clarity to psychologists’ understanding of young infants’ abilities in this domain.

Under any circumstance, the finding that female infants in the current study behaved in line with Wynn’s (1992) predictions, but that male infants—whose normal development is just slightly delayed relative to female infants—did not, suggests that developmental events occurring around 5 months of age contribute to an emerging ability to resist the influence of perceptual familiarization in this protocol. Thus, Carey’s (2002) contention that “with older children (7 months and older), an avalanche of data support the existence of numerical representations” seems likely to gain support with additional empirical exploration of these phenomena. Nonetheless, as Cohen and Marks (2002) noted, the “important question is whether infants progress from a simple preference for familiarity to more sophisticated approaches . . . and proceed to true addition and subtraction” (p. 199). This question is inherently developmental, and as such, pursuing it will ultimately lead to an understanding of how component processes that contribute to mature cognition are assembled in development to allow for the emergence of numerical competence. The early appearance in development of some numerical competencies—and claims that they are evolved forms of cognition—ought not to distract researchers from the essential task of further exploring their development.

## References

- Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development, 54*, 695–701.
- Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in 5-month-old infants. *Cognition, 20*, 191–208.
- Bauer, J. A., Shimojo, S., Gwiazda, J., & Held, R. (1986). Sex differences in the development of binocularity in human infants. *Investigative Ophthalmology & Visual Sciences, 27*(Suppl.), 265.
- Bjorklund, D. F., & Bering, J. M. (2002). The evolved child: Applying evolutionary developmental psychology to modern schooling. *Learning and Individual Differences, 12*, 347–373.
- Bogartz, R. S., Shinsky, J. L., & Schilling, T. H. (2000). Object permanence in five-and-a-half-month-old infants? *Infancy, 1*, 403–428.
- Bogartz, R. S., Shinsky, J. L., & Speaker, C. J. (1997). Interpreting infant looking: The Event Set × Event Set design. *Developmental Psychology, 33*, 408–422.
- Capaldi, E. J., & Miller, D. J. (1988). Counting in rats: Its functional significance and the independent cognitive processes that constitute it. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 3–17.
- Carey, S. (2002). Evidence for numerical abilities in young infants: A fatal flaw? *Developmental Science, 5*, 202–205.
- Caron, R. F., & Caron, A. J. (1969). The effects of repeated exposure and

- stimulus complexity on visual fixation in infants. *Psychonomic Science*, 10, 207–208.
- Cashon, C. H., & Cohen, L. B. (2000). Eight-month-old infants' perception of possible and impossible events. *Infancy*, 1, 429–446.
- Chiang, W., & Wynn, K. (2000). Infants' tracking of objects and collections. *Cognition*, 77, 169–195.
- Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, 10, 408–411.
- Cohen, L. B., Gelber, E. R., & Lazar, M. A. (1971). Infant habituation and generalization to differing degrees of stimulus novelty. *Journal of Experimental Child Psychology*, 11, 379–389.
- Cohen, L. B., & Marks, K. S. (2002). How infants process addition and subtraction events. *Developmental Science*, 5, 186–201.
- Creighton, D. E. (1984). Sex differences in the visual habituation of 4-, 6-, and 8-month-old infants. *Infant Behavior and Development*, 7, 237–249.
- Fantz, R. L., & Fagan, J. F., III. (1975). Visual attention to size and number of pattern details by term and preterm infants during the first six months. *Child Development*, 16, 3–18.
- Feldman, R. S. (1998). *Child development*. Upper Saddle River, NJ: Prentice Hall.
- Ferraro, F. R. (1993). Potentially relevant alternative interpretation of infants' addition and subtraction. *Perceptual and Motor Skills*, 76, 110.
- Friedman, S., Bruno, L. A., & Vietze, P. (1974). Newborn habituation to visual stimuli: A sex difference in novelty detection. *Journal of Experimental Child Psychology*, 18, 242–251.
- Geary, D. C. (2000). From infancy to adulthood: The development of numerical abilities. *European Child and Adolescent Psychiatry*, 9, II/11–II/16.
- Grattan, M. P., de-Vos, E., Levy, J., & McClintock, M. K. (1992). Asymmetric action in the human newborn: Sex differences in patterns of organization. *Child Development*, 63, 273–289.
- Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development*, 21, 167–179.
- Hauser, M. D., Tsao, F., Garcia, P., & Spelke, E. S. (2003). Evolutionary foundations of number: Spontaneous representation of numerical magnitudes by cotton-top tamarins. *Proceedings of the Royal Society of London (Series B)*, 270, 1441–1446.
- Held, R., Bauer, J., & Gwiazda, J. (1988). Age of onset of binocularity correlates with level of plasma testosterone in male infants. *Investigative Ophthalmology & Visual Sciences*, 29(Suppl.), 60.
- Held, R., Shimojo, S., & Gwiazda, J. (1984). Gender differences in the early development of human visual resolution. *Investigative Ophthalmology & Visual Sciences*, 25(Suppl.), 220.
- Humphrey, D. E., & Humphrey, G. K. (1987). Sex differences in infant reaching. *Neuropsychologia*, 25, 971–975.
- Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, 19, 338–352.
- Kagan, J., Henker, B. A., Hen-Tov, A., Levine, J., & Lewis, M. (1966). Infants' differential reactions to familiar and distorted faces. *Child Development*, 37, 519–532.
- Karmel, B. Z. (1969). The effect of age, complexity, and amount of contour density on pattern preferences in human infants. *Journal of Experimental Child Psychology*, 7, 339–354.
- Kobayashi, T., Hiraki, K., Mugitani, R., & Hasegawa, T. (2004). Baby arithmetic: One object plus one tone. *Cognition*, 91, B23–B34.
- Koechlin, E., Dehaene, S., & Mehler, J. (1997). Numerical transformation in five-month-old human infants. *Mathematical Cognition*, 3, 89–104.
- Lickliter, R., & Berry, T. D. (1990). The phylogeny fallacy: Developmental psychology's misapplication of evolutionary theory. *Developmental Review*, 10, 322–338.
- Malcolm, C. A., McCulloch, D. L., & Shepherd, A. J. (2002). Pattern-reversal visual evoked potentials in infants: Gender differences during early visual maturation. *Developmental Medicine & Child Neurology*, 44, 345–351.
- Mix, K. (2002). Trying to build on shifting sand: Commentary on Cohen and Marks. *Developmental Science*, 5, 205–206.
- Moore, D. S. (2002). *The dependent gene: The fallacy of "nature vs. nurture."* New York: Times Books/Henry Holt.
- Pepperberg, I. M. (1991). A communicative approach to animal cognition: A study of conceptual abilities of an African grey parrot. In C. A. Ristau (Ed.), *Cognitive ethology, the minds of other animals* (pp. 153–186). Hillsdale, NJ: Erlbaum.
- Schilling, T. H. (2000). Infants' looking at possible and impossible screen rotations: The role of familiarization. *Infancy*, 1, 389–402.
- Shucard, J. L., Shucard, D. W., & Cummins, K. R. (1981). Auditory evoked potentials and sex-related differences in brain development. *Brain and Language*, 13, 91–102.
- Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A "non-numerical" account. *Cognitive Development*, 12, 349–372.
- Simon, T. J. (1998). Computational evidence for the foundations of numerical competence. *Developmental Science*, 1, 71–78.
- Simon, T. J., Hespos, S. J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development*, 10, 253–269.
- Singer, J. E., Westphal, M., & Niswander, K. R. (1968). Sex differences in the incidence of neonatal abnormalities and abnormal performance in early childhood. *Child Development*, 39, 103–112.
- Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatio-temporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, 13, 113–142.
- Strauss, M. S., & Curtis, L. W. (1981). Infant perception of numerosity. *Child Development*, 52, 1146–1152.
- Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.
- Tanner, J. (1978). *Fetus into man: Physical growth from conception to maturity*. Cambridge, MA: Harvard University Press.
- Tighe, T. J., & Powlison, L. B. (1978). Sex differences in infant habituation research: A survey and some hypotheses. *Bulletin of the Psychonomic Society*, 12, 337–340.
- Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development*, 14, 1–36.
- van Loosbroek, E., & Smitsman, A. W. (1992). Visual perception of numerosity in infancy. *Developmental Psychology*, 26, 916–922.
- Vilette, B. (2002). Do young children grasp the inverse relationship between addition and subtraction? Evidence against early arithmetic. *Cognitive Development*, 17, 1365–1383.
- Wakely, A., Rivera, S., & Langer, J. (2000). Can young infants add and subtract? *Child Development*, 71, 1525–1534.
- West, R. E., & Young, R. J. (2002). Do domestic dogs show any evidence of being able to count? *Animal Cognition*, 5, 183–186.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358, 749–750.
- Wynn, K. (1995). Origins of numerical knowledge. *Mathematical Cognition*, 1, 35–60.
- Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science*, 7, 164–169.
- Wynn, K. (2002). Do infants have numerical expectations or just perceptual preferences? *Developmental Science*, 5, 207–209.
- Wynn, K., Bloom, P., & Chiang, W. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition*, 83, B55–B62.

Received November 8, 2004

Revision received September 19, 2005

Accepted September 22, 2005 ■