Cognitive Development, 14, 1–36 (1999)
ISSN 0885-2014

WHAT REPRESENTATIONS MIGHT UNDERLIE INFANT NUMERICAL KNOWLEDGE?

Claudia Uller
Rutgers University

Susan Carey
New York University

Gavin Huntley-Fenner
University of California at Irvine

Laura Klatt
University of Pennsylvania

The sensitivity of young infants to number raises the question of the nature of their numerical representations. We contrast two broad classes of models that have been proposed in the literature: Integer-symbol models and Object-file models. Four infant-addition experiments ($1 + 1 = 2$ or 1) were conducted. Experiments 1 and 2 showed that the timing of the placement of the screen relative to the objects determined 8-month-olds' success (object-first condition) and failure (screen-first condition) in such addition tasks. Experiment 3 established that 10-month-old infants succeed in the single-screen, screen-first, paradigm. Experiment 4 examined whether the failure in the screen-first

This research was supported by CNPq/Brazil fellowship 202623/91.2 to Claudia Uller, NSF project grant BNS9012075 to Susan Carey, and NIH/NIGMS fellowship to Gavin Huntley-Fenner. We are grateful to Renée Baillargeon, Helene Intraub, Kelly Olguin Jaakkola, Susan Johnson, Ilona Kovačs, Alan Leslie, Sandeep Prasada, Zenon Pylyshyn, Liz Spelke, and Fei Xu for helpful discussions of the issues in this paper, and to two anonymous reviewers for comments on an earlier draft. We are grateful to Thomas M. Donnelly for the statistical assistance. We thank Jenny Welch for running the Infant Cognition Laboratory at MIT, and B. Azad-Tatari, S. Charles, J. Chang, A. Kakouros, C. Petrescu, and J. Tyan for recruiting infants, and coding infants' looking times. Finally, we express gratitude to the infants and their parents for their participation.

Direct all correspondence to: C. Uller, Center for Cognitive Science, Rutgers University, Busch Campus, Piscataway, NJ 08854 <uller@ruccs.rutgers.edu>.

Manuscript received September 29, 1997; revision accepted August 6, 1998

condition was due to an imprecise representation of object location. Separate screens enabled 8-month-olds to succeed in a screen-first paradigm. We argue that this pattern of success and failure favors the Object-file model over Integer-symbol models.

Over the past 20 years, simple habituation experiments have provided ample evidence that young infants, even neonates, are sensitive to numerical distinctions among sets of one, two, and three entities (e.g., dots: Antell & Keating, 1983; Starkey & Cooper, 1980; Tan & Bryant, 1996; Treiber & Wilcox, 1984; colored squares: Cooper, 1984; familiar objects: Starkey, Spelke, & Gelman, 1990; Strauss & Curtis, 1981; continuously moving figures: van Loosbroek & Smitsman, 1990; jumps of a puppet: Wynn, 1996). For example, infants who are habituated to pictures of sets of two familiar objects varying in type, size, and position, dishabituate when shown arrays of one object or three objects (Strauss & Curtis, 1981). Evidence for distinguishing four from other numbers of entities is weaker, but sometimes obtained (Tan & Bryant, 1996; van Loosbroek & Smitsman, 1990).

Of course, the ability to discriminate one from two from three entities does not entail that infants understand anything about the numerical relations between one and two and three, such as that two is less than three, or that three is exactly one more than two. However, using the methodology of violation of expectancy, Wynn (1992) showed that 5-month-old infants represent some of the numerical relations between one, two, and three objects, such as that $1 + 1$ is precisely two and that $2 - 1$ is precisely one. Take, for example, the $1 + 1 = 2$ or 1 task. One object was placed on a stage, covered with a screen, and then another object was introduced behind the screen. When the screen was removed, infants looked longer at impossible outcomes of either one object or three objects than at the possible outcomes of two objects, suggesting that infants expected precisely two objects (see also Baillargeon, Miller, & Constantino, 1994; Koechlin, Dehaene, & Mehler, in press; Simon, Hespos, & Rochat, 1995; Wynn, 1995).

These data leave open the question of the nature of the representations underlying infants' performance in the addition/subtraction tasks. Most fundamentally, the question remains whether the infants' representation of number contains explicit symbols for the integers. Some (e.g., Gallistel & Gelman, 1992; Wynn, 1995) have suggested that infants encode each of these events by representations that include a single symbol standing for the number of objects in the array, this symbol being arrived at through a process that embodies a counting algorithm. Such models are designated here as *Integer-symbol models.* Others (Carey, 1996; Huttenlocher, Jordan, & Levine, 1994; Simon, 1997; Uller, 1995; Uller, Carey, & Huntley-Fenner, 1994) have suggested that infants encode these events by representations of the objects in the arrays alone, such that each object is represented by a distinct symbol for that object, with no symbol for any integer involved at all, and thus no counting algorithm engaged in the encoding of the arrays. Models of this sort are designated here as *Object-file models.*

It is not easy to bring decisive empirical data to bear on deciding between Integer-symbol and Object-file models of infant number representation, and indeed, nobody has yet succeeded.[1] Simon's (1997) defense of the Object-file model was on parsimony alone; he argued that we know from other evidence that infants have the representational capacities called for on the Object-file model, and that this model is sufficient to account for infant success in number tasks. Whereas this statement may be true, it is also possible that infants are representing number symbolically in these studies. It is, of course, an empirical question. Here we review the currently available data that address this issue and suggest a new source of relevant empirical observation. It is important to our understanding of the course of cognitive development to characterize the infant's initial state correctly (as argued also by Simon, 1997). Both overattribution and underattribution of cognitive resources to infants have serious theoretical consequences.

One possible empirical avenue to follow in attempting to decide between the two classes of models derives empirically from the observation that symbolic models place different demands on short-term memory than do Object-file models. In what follows, we first characterize the two classes of models and show why the two engage short-term memory differently. We then review briefly the already existing infant-addition/subtraction literature as it bears on differential predictions from the two classes of models with regard to the effects of set size on infant performance. Finally, we develop a new approach to bringing data to bear on deciding between the two classes of models.

Integer-Symbol Models

In their ground-breaking book, Gelman and Gallistel (1978; see also Gallistel, 1990; Gallistel & Gelman, 1992) provided an abstract characterization of any symbolic representation of integers. Any such system includes a mentally represented list of symbols, a process that puts individuals to be enumerated in one-to-

[1] Simon et al. (1995) attempted to bring data to bear on the question of whether infant numerical knowledge is arithmetically based or object-based. They reasoned that if infants are encoding these events in terms of representations of the objects in the array, their looking times would be elevated to impossible outcomes that consist of object-identity mismatches and to outcomes that consist of object number mismatches. For example, if an Ernie doll is introduced onto the stage, a screen raised, and an Elmo doll added, infants looking times should be elevated to outcomes that consist of just one doll, but also to outcomes that consist of two Ernies or two Elmos. This prediction of the Object-file model was not supported. Looking times were affected only by numerical matches and mismatches, as would be predicted by Integer-symbol models. However, as Simon et al. noted, these data are not conclusive, for their interpretation depends on what criteria the infants use for individuation and numerical identity. For instance, the infants' models may be of two objects behind the screen, the property differences between Ernie and Elmo not entering into the infant's object file representations.

one correspondence with items in this list, always proceeding in the same order through the list. The number of items in the set of enumerated items is represented by the last item on the list reached, its numerical value determined by the ordinal position of that item in the list. Gelman and Gallistel originally envisioned that the mentally represented list of symbols, called by them "numerons," would be discrete, arbitrary symbols (e.g., "!, @, #, $, &, . . ."), their numeric value carried only by their place in the list. For example, in the above list, "@" represents 2, because it is the second item in the list. In their later work, Gelman and Gallistel acknowledged that a variety of analog representational systems, such as that proposed by Meck and Church (1983), could equally well serve as symbolic representation of integers (Gallistel, 1990; Gallistel & Gelman, 1992). Meck and Church proposed that animals represent integer values with a magnitude that is an analog of number. The idea is simple—suppose that the nervous system has the equivalent of a pulse generator that outputs neural activity at a constant rate, and a gate that can open to allow energy through to an accumulator that registers how much has been let through. When the animal is in a counting mode, the gate is opened for a fixed amount of time (e.g., 200 ms) for each item to be counted. The total energy accumulated will then be an analog representation of number. This system works as if length were used to represent number, with "–" being a representation of 1, "—" a representation of 2, "——" a representation of 3, and so on (see Gallistel, 1990, for a summary of evidence for the accumulator model of animal representation of number.) Although there are real differences between a non-analog numeron list system and an analog system such as the accumulator model, as Gallistel and Gelman point out, the accumulator system is a realization of their abstract characterization of symbolic counting algorithms. Items to be counted are paced in one-to-one correspondence with items on a list, and the numerical value of any given representation is given by its ordinal position in a series of states of the accumulator: ("–, —, ——, . . ."). In both systems, a counting process results in an abstract symbol for an integer; 2 is represented by "@" or "—" in the present examples. For the purposes of this paper, models of both sorts will be considered Integer-symbol models, but see Wynn (1990, 1992) for an argument that an analog system such as the accumulator model is more likely than a discrete numeron list system to underlie infant representation of number.

Object-File Models

Success in the infant-habituation studies, and in the addition/subtraction studies, could be supported by representations of the actual objects in the array. Simon (1997) argued that success may reflect nothing more than already well-documented physical reasoning abilities. Many studies show that infants of 5 months and even younger establish representations of an object placed on a stage and can reason about the existence of and physical interactions between objects that are hidden behind screens (e.g., Baillargeon et al., 1994, 1995; Spelke, Kestenbaum, Simons, & Wein, 1995; Spelke & Van de Walle, 1995). Infants may build a model of the objects behind the screen, updating this model when new objects are added or when objects are taken away. On the view we advocate here, the arrays are encoded in terms of a separate object-file (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992) for each object. Thus, an array of two objects would be encoded by a representation of the form: "O_i O_j," or in the form of any other representation that is equivalent to "There is an entity, there is a numerically distinct entity, each entity is an object, and there is no other object."[2] Note that in such a representation there is no single symbol for 2 at all, not "2" or "@" or "—" nor any other. For the purposes of this paper, we assume that object files are imagistic representations of the objects in the array, although we make no commitment to how abstract they might be, that is, how many of the perceptible properties of the objects may be instantiated in an object file. That is, two Xs may be represented "X_i X_j" or "O_i O_j," whereas two Ys may be represented "Y_i Y_j" or "O_i O_j."

Symbolic-file and Object-file models differ in two interdependent ways: (1) in the nature of the representations of the events and (2) in the algorithm underlying discrepancy detection between the representation of the set-up event (e.g., 1 + 1) and the representation of the final display revealed upon removal of the screen (e.g., 2 or 1). According to symbolic models, infants represent the number of items in the set-up event with a symbol (e.g., "@" or "—") that is stored in short-term memory. The number of objects in the outcome array is also represented symbolically ("@" or "—" in possible outcomes, "!" or "–" in impossible outcomes). The comparison process involves establishing whether the two symbols match or not, and whether they are tokens of a single type or of different types. In contrast, the Object-file model assumes that a representation of two object files, "O_i O_j," is constructed during the test event and stored in short-term memory, and is compared with a representation of two object files, "O_k O_l," constructed during the outcome (possible outcome) or to a representation of one object file, "O_k," (impossible outcome) by a process that detects one-to-one correspondence between object files in the two representations.

As is clear from a consideration of the differences in the processes underlying

---

[2] The term *numerically distinct* is required in this formulation because of the ambiguity of each of the words "same" and "different" in English. "Different" can refer to difference in kind or property or to difference in the sense of numerical identity, "different one." Similarly, "same" can be used in the sense of sameness of property or kind or in the sense of numerical identity, sameness in the sense of "same one." It is numerical identity that is the issue in the Object-file models. The Object-file models represent number in virtue of being an instantiation of the following: $(\exists x)(\exists y)\{(\text{object}[x] \;\&\; \text{object}[y]) \;\&\; x \neq y \;\&\; \forall z(\text{object}[z] \rightarrow [z = x] \vee [z = y])\}$. In English, this equation states that there is an entity, and there is another entity numerically distinct from it, and that each entity is an object, and there is not any other object. This sentence is logically equivalent to "There are two objects" and indeed, the above formula is the representation of that sentence in first-order logic.

discrepancy detection, the two classes of models differ in the memory demands they place on the infant. This difference in turn opens the possibility of bringing empirical data to bear on deciding between the two classes of models. Holding one particular symbol in memory (e.g., "%") does not require more memory resources than holding another (e.g., "#"), nor does the process of match/mismatch detection vary with the content of the symbols ("#" and "#" match, just as "%" and "%" match). If the symbolic representations are analogs, discrepancy detection will vary slightly with the content of the symbol: as the values of the numbers represented rise, if the differences between two numbers remain just one, the comparison process will become harder, for reasons of Weber's law ("——" vs. "——" is a harder discrimination than "—" vs. "–"). Memory loads for lists of object files, in contrast, vary dramatically with list length. Short-term memory representations of list of items, whether images or words, are among the most studied data structures in cognitive psychology (e.g., Baddeley, 1986, for a review). Memory load increases with the length of the list and the number of items in the list. Furthermore, the process of one-to-one correspondence detection will become much more complex as the numbers of items in the two representations increase.

The available data from existing addition/subtraction studies suggest dramatic effects of list length, as predicted by the Object-file model. Infants succeed on $1 + 1$ and $2 - 1$ tasks by 5 months of age (Koechlin et al., 1997; Simon et al., 1995; Wynn, 1992). The earliest reported successes with $2 + 1 = 3$ or 2 and $3 - 1 = 2$ or 3 tasks is 10 months of age (Baillargeon et al., 1994; Wynn, 1995). However, these data do not provide conclusive support for the Object-file model, as symbolic models can accommodate such results on the assumption that mistakes in the counting process are more likely the higher the number of objects to be counted, and in analog instantiations such as the accumulator model, discriminations between adjacent numbers become harder as the numbers become larger.

A systematic examination of the effects of list length arises in the course of comparing addition and subtraction versions of each study. Compare $1 + 1 = 2$ or 1 with $2 - 1 = 1$ or 2. The list length in an Object-file model of the test event is larger in the addition version (2) of the task than in the subtraction version of the task (1). That is, in the addition version the infant must hold a longer list in memory, and must evaluate one-to-one correspondences between larger sets during the outcome phase of the study. The Object-file model predicts, therefore, that addition will be more difficult than subtraction, and indeed infants succeed on subtraction tasks more robustly than on addition tasks. In Wynn's (1992) study, 5-month-old infants differentiated the outcomes in the $1 + 1 = 2$ or 1 version more weakly than in the $2 - 1 = 1$ or 2 version, as did the infants in the Koechlin et al. (in press) study. Indeed, only in Simon et al.'s (1995) replication did the 5-month-old infants succeed outright on the addition version of the task; in Wynn's (1992) and Koechlin et al.'s (in press) studies, success consisted merely of different looking-time patterns in the subtraction and addition versions. Further, Wynn (1995) also showed that infants had different looking-time patterns in $3 - 1$ versus $2 + 1$ tasks, but outright preference for the impossible outcomes was found only in the subtraction version of the study. Finally, in Starkey's (1992) addition and subtraction studies with young toddlers as well, subtraction was systematically easier than addition.

It is difficult to see how symbolic counting models would account for subtraction being easier than addition. In subtraction events, the counter must be incremented to the maximum value of the set before an object is removed, and then adjusted downward. During addition, the counter is simply incremented to the maximum value of the final set. That is, a representation of a $2 - 1$ subtraction test event includes all the steps of a $1 + 1$ addition event, plus an additional step. Symbolic models would seem to require that subtraction, if anything, would be more difficult than addition, as indeed it is in early arithmetic, when we know children are using a symbolic list-type of representation (Fuson, 1988; Siegler & Robinson, 1982).

An even more dramatic effect of list length is seen in the apparent upper limit on infant number representations. None of the infant habituation studies is successful in showing infant discrimination of numbers that exceed four. Most studies probing numbers higher than three have failed (Antell & Keating, 1983; Cooper, 1984; Starkey & Cooper, 1980; Strauss & Curtis, 1981), but see Tan and Bryant (1996), Treiber and Wilcox (1984), and van Loosbroek and Smitsman (1990) for limited success for higher numbers. Integer-symbol models have no intrinsic upper limit at three or four—rats and pigeons have been trained to enumerate larger numerosities, varying from 24 (rats: Capaldi & Miller, 1988; Meck & Church, 1983) to 50 (pigeons: Honig & Stewart, 1989; Rilling & McDiarmid, 1965; see also Gallistel, 1990, for a review)—nor is there any theoretically motivated reason to suspect that an innate numeron list would have such a limit. In an analog system such as the accumulator model, a 4 versus 6 discrimination places the same Weber fraction demands as a 2 versus 3 discrimination. Thus, the Integer-symbol models do not predict the almost categorical cutoff in performance as numbers of entities exceed three or four. However, such a limit is predicted by the Object-file account, for we know there is a limit on parallel individuation, the number of distinct objects that can be simultaneously tracked in a visual model of an array (Trick & Pylyshyn, 1994). Although the apparent upper limit on set sizes infants can discriminate is predicted by the Object-file model, it does not provide conclusive evidence for it, for there has yet been no systematic attempt to explore infant discriminations of larger numbers, taking into account Weber fraction considerations. In sum, we read the available evidence on the effects of set size on number representation as providing consistent but inconclusive support for the Object-file models over the Integer-symbol models of infant representations in these addition/subtraction tasks.

The studies in this paper adopt a different approach to finding data that may bear on comparisons between the two classes of models. Here we compare performance on different 1 + 1 addition tasks. Because these events always involve adding 1 and 1, symbolic models posit identical representations in each case. In each addition event, the hypothesized counter would be incremented twice, yielding the symbol for 2 ("@" or "—"). That symbol would then be compared with the symbol that represents the number of objects revealed in the outcomes: the symbol for 2 ("@" or "—") in possible outcomes, and the symbol for 1 ("!" or "–") in impossible outcomes). That is, according to the symbolic models, the representations constructed and the processes of comparison do not differ from one task to another. However, these tasks differ in terms of experimental manipulations that should affect the robustness of the object files constructed and held in short-term memory. These manipulations would be expected to affect performance if the Object-file model underlies infant performance on these tasks.

The manipulations in Experiments 1 and 2 are motivated by the following assumptions (hereafter referred to as *Model Robustness Assumptions*): (a) Representations of object files constructed in imagery are less robust representations than those constructed from perception of objects. That is, a representation of an object actually seen on the stage floor will be more robust than a representation of an object introduced behind a screen; (b) Each update on the represented object decreases the robustness of the representation. That is, a representation (of an object) that undergoes a process of change will be less robust than a representation (of an object) that remains as originally set up.

Baillargeon et al. (1994) reported a result consistent with these model robustness assumptions. They found that 10-month-old infants succeed in a 2 + 1 = 3 or 2 addition experiment, although success was weak, on the first pair of test trials only, but that infants of this age failed totally in a 1 + 1 + 1 = 3 or 2 addition experiment. These two experiments do not differ in the demands they make on a symbolic integer representation. Both require the counter to be incremented to the symbol for 3, and that symbol to be held in memory and to be compared with the result of the count of the outcome array. Now consider what is required to construct a mental model of three objects in each of these two situations. Both experiments began with an empty stage, upon which a screen is placed. In the 2 + 1 test event, two objects are placed, simultaneously, behind the screen. The infant mentally constructs a model of two objects behind the screen; this construction is done in imagery because the infant has not seen the two objects on the stage (one update in imagery). Then another object is introduced. The infant must operate on his or her mental model, updating it by adding another object to it, producing a model of three objects behind the screen (a second update in imagery). Ten-month-old infants succeed in this condition, looking longer at the impossible outcome of two objects when the screen is removed than at the possible outcome of three objects in the first pair of test trials.

Now consider the 1 + 1 + 1 = 3 or 2 condition, at which the infants failed. The infant is shown the empty stage and the screen is placed on it. An object is place behind the screen, and the infant must create in imagery a model of one object behind the screen (first update). Then a second object is placed behind the screen and the infant must operate on the model, updating it a second time to yield a model of two objects on the stage (second update). Then a third object is placed behind the screen and the infant must operate on the model again. The model is updated a third time, yielding a representation of three objects on the stage. On the assumption that each updating of a mental model in imagery yields a noisier representation, we can understand why a model of 1 + 1+ 1 objects (three updates) is harder for the infant to construct and maintain than a model of 2 + 1 objects (two updates). Note that the symbolic models have no natural explanation for why 1 + 1 +1 is harder to compute than 2 + 1, because both counts require a counter to be incremented for each item in the count (i.e., each require stepping through precisely "!, @, *" or "–, —, ——").

In Wynn's (1992) 1 + 1 addition experiments, and in the replications by Simon et al. (1995) and Koechlin et al. (1997), the infants saw the first object placed on the stage, the screen was raised and then the second object was introduced. This scenario is what we will call an object-first design. That is, the representation of the first object was constructed from a perceptual experience of it resting on the stage floor, and the construction of the final representation of the event required one update in imagery, as the original representation of one object file is modified with the addition of the second object, yielding a representation of one object file and another object file. The Baillargeon et al. (1994) design is what we will call a screen-first design. The screen is placed on the stage before the first object is introduced behind it. The two robustness assumptions yield the prediction that an object-first 1 + 1 test event will be easier than a screen-first 1 + 1 test event because in the object-first event the first object is actually seen on the stage floor and there is only one update in imagery. Experiments 1 and 2 test this prediction.

EXPERIMENT 1

Experiment 1 provides a test of the model robustness assumptions at the same time as providing data that will bear on a choice between the Object-file model and symbolic models of the representations underlying success on infant-addition/subtraction tasks. Wynn's (1992) object-first procedure can be changed from one that requires one update of a model constructed on the basis of perception to one that requires two updates, one of which is an update of a model constructed in imagery, simply by changing it to a screen-first procedure. If the Object-file model and the model robustness assumptions are correct, there should be an age at which infants will succeed in the object-first condition but will fail in the screen-first condition. Given 10-month-olds' success at a screen-first 1 + 1 = 2 or 3 procedure (Baillargeon et al., 1994), 8-month-olds were chosen to test this prediction.

The Object-file model is deliberately underspecified. As mentioned above, we make no commitment as to how many or which features of an object are represented in its object file. Also, we make no commitment as to how the spatial relations among objects are represented in the underlying object-file representations. Infants may represent the exact location of the seen object on the stage (but see Koechlin et al., in press, for evidence that such location information is not necessary for success on infant-addition/subtraction tasks.) In Wynn's (1992) study, the impossible outcome of one object was created by removing the second (previously unseen) object, leaving the object that the infant had actually seen in the location in which the infant had actually seen it. If infants are mentally representing the exact locations of the objects on the stage, then they may perform better on trials in which the impossible outcome is due to the first object being removed, especially in the object-first condition, in which the first object is actually seen on the stage. Experiment 1 also tests this hypothesis.

Method

Participants. Sixty-four full-term normal 8-month-old infants (32 boys, 32 girls) ranging in age from 7 months, 11 days to 8 months, 25 days (M_{age} = 8 months, 8 days) were tested. The participants were volunteers from a population in the middle-class, suburban Boston area. They were primarily white, but also reflected the Asian and the African-American segments of that population. Thirty-six additional participants were excluded because of fussiness (29), experimenter error (4), equipment failure (2), or extraordinarily long looking during the familiarization trials—more than 2 standard deviations above the mean familiarization looking time (1). The participants' names were retrieved from the birth records in the Greater Boston area and their parents were contacted by letter and by phone. Parents were compensated with token gifts (T-shirts, bibs, and plastic cups).

Materials and Apparatus. The stimuli were roughly the shape of a flattened cone.[3] They measured 15 cm in diameter at the base and 4.5 cm high at the center. The surfaces of these objects were completely coated in glued-on sand. A piece of string was attached to the top center of the object, allowing it to be raised and lowered without being held directly.

This experiment took place on a stage whose opening measured 38 cm × 88 cm × 34 cm and which was raised 100 cm from the ground level. A black felt backdrop for the stage hid the movements of the experimenter. Attached to the

[3] The objects were part of a series of experiments designed to compare infants' knowledge of sets of solid objects and non-solid substances. For that study, objects were made to look similar to a portion of sand. In the present studies, infants were given full evidence that the stimuli were objects, not piles of sand: they handled the objects and saw the objects moved, lifted, and shaken. Thus, that they looked like a pile of sand is irrelevant to the present study.

front of the stage was a black screen that could be raised by the experimenter to partially obscure the display. The stage area itself was surrounded by black curtains from floor to ceiling. These curtains hid an observer, who recorded looking-time data, and a video camera, which recorded the participant. During the experiment, the laboratory was darkened and the stage was lit directly from above. The participant was lit indirectly for videotaping by lamps in front of and on either side of the stage.

The participant sat facing the stage with his or her head about 150 cm away and his or her eyes slightly above the floor level of the stage. The participant's parent/guardian sat on the participant's left side facing away from the stage such that he or she could not see what was being presented to the participant. Parents were instructed not to turn around and not to look at the display. They were also instructed to interact with the participant as little as possible. An observer could see the participant through an "invisible" hole in the black curtains. The observer could not see the stage, and thus was blind to the details of the experimental manipulation. A white noise generator masked any sound from the movements of the observer and the experimenter. Looking time data were collected when the observer pressed a button connected to an IBM-486 computer every time the participant looked at the display. The experimenter calibrated the looking range of each participant at the beginning of each session, by shaking her laboratory keys on the left, center, and right sides of the stage. Trials began when the participant had looked at the stage for at least 0.5 s and ended when the participant looked away for 2 continuous seconds.

Typically, for the experiments in this paper, there was one highly trained and experienced live observer.[4] Each participant was also videotaped for offline observation by coders who were blind to the experimental conditions in which the participant was placed. Interobserver reliability between online and offline observations is reported for each experiment in the Results sections. For each trial, reliabilities were calculated as the ratio of the looking-time scores of the two observers (e.g., if one observer indicated 3.21 s looking time, and the other 3.15 s looking time, interscorer reliability for that trial would be 3.15/3.21 = .98).

Procedure

Each participant was assigned randomly to one of two conditions: the object-first condition (n = 32; M_{age} = 8 months, 2 days) or the screen-first condition (n = 32, M_{age}= 8 months, 7 days). The object-first condition was modeled on Wynn (1992) with one main difference: objects were lowered from above rather

[4] Observers were trained in online sessions as a second online observer. An observer was considered trained when his or her measurements corresponded to the measurements of the primary trained observer, to the level of accuracy of milliseconds, with 91% (or higher) agreement in three consecutive sessions.

than introduced from the side, making the prediction of the location of the objects behind the screen straightforward. Each condition consisted of three phases: an introductory phase, familiarization trials, and test trials. The introductory phase served to acquaint participants with the novel object stimuli. The familiarization trials served to introduce participants to the stage on which the objects were placed and to aspects of the events which were seen during the test trials—namely, the lowering of objects, the raising and lowering of the screen, and the fact that objects lowered into a hidden location are later found in that location. These trials also provided a measure of any baseline preference the participants might have between arrays containing two objects and arrays containing one single object. Test trials immediately followed the familiarization trials. The intertrial delay (approximately 4 s) between familiarization and test trials was the same as that between each familiarization or test trial. Test trials were $1 + 1 = 2$ or 1 events. Looking times when the screen was removed were used to infer the infants' expectations of one and two objects to be on the stage. Condition (screen-first, object-first) and object removed on one-object trials (first-seen, second-seen) were between participants variables, whereas trial-type (familiarization, test) and outcome (one object, two objects) varied within participants.

Introductory Phase (All Infants). Every participant was given a 60-s exposure to the object before the experiment began. The experimenter held the object by its string and drew the participant's attention to the object. The experimenter then handed the object to the participant, allowing him or her to play with it. If the participant was initially reluctant to grasp the object, the experimenter asked the parent to encourage the participant to handle the object.

Familiarization Trials (All Infants). There were two pairs of familiarization trials, one pair in full view and one pair involving a screen. The first pair took place entirely within the participant's view. The second pair was partially hidden by a screen. When objects were being introduced, the experimenter spoke to the participant: "Look, baby! Watch the object!" to keep his or her attention to the stage area. Each pair of familiarization trials consisted of a single-object trial and a double-object trial. The pair with no screen proceeded as follows. In the single-object trials the experimenter held an object by its string and introduced it into the display. The object was slowly lowered toward the stage floor and then stopped when it was about halfway down. It was then jiggled on its string for about 5 s and then lowered onto the stage floor. When the object reached the stage floor, it was tapped five times on the floor and then released. The hand that held the object was retracted from the display. As soon as the hand exited the stage area, the experimenter signaled the observer to record the participant's looking at the stage area by saying "now." In the double-object trials, the participant saw two objects held by their strings lowered simultaneously toward the stage floor. The objects were stopped halfway down and then jiggled for 5 s before simultaneously continuing toward the stage floor. When the objects reached the stage floor, they were simultaneously tapped on the floor five times and then released. The hands were then retracted and the participant's looking time was measured. After the participant looked away from the display for a period of 2 s the experimenter removed the stimuli by means of a movable flap in the bottom of the backdrop.

The second pair of familiarization trials was the same except that the objects were lowered behind a screen. Trials began with the stage empty. The experimenter first raised the screen to hide the stage floor. Then either a single object or two objects were lowered simultaneously into the area above the screen until their lower part was hidden behind the top of the screen but their tops were partially visible. The objects were jiggled in that position for about 5 s and then lowered behind the screen onto the stage floor where it/they were tapped five times on the floor and released. The experimenter's hand was retracted and the screen was lowered to reveal the objects directly below where they had been lowered. The experimenter then signaled the observer to begin measuring looking time by saying "now."

For 75% of the participants, there were four familiarization trials—one of each type. The pair of trials without a screen always preceded the trials with a screen. Familiarization trials alternated between one and two object outcomes in two orders (1-2-2-1 or 2-1-1-2). Order and side of the single-object trials (left, right) were counterbalanced across participants. The remaining 25% of the participants (the ones in the screen-first condition in which the first object lowered was the object removed in impossible outcome trials)[5] were given only the first pair of familiarization trials. Order (1-2 or 2-1) and side of the single-object trials (left, right) were counterbalanced across participants.

Test Trials (Object-First). Six test trials immediately followed the familiarization trials. During the test trials, the experimenter took care to draw the participant's attention to the objects as they were lowered by speaking to the participant until the second object had been released. Figure 1 shows the test trials of the object-first condition. First, participants were shown the empty stage. Next, a single object was lowered into one end of the display; it stopped halfway down, was jiggled for 5 s and then lowered, tapped five times on the stage floor and released. The hand that held the object was retracted and then the screen was raised to hide the object on the stage floor. Next a second object was lowered into the other end of the display behind the screen. This object was also stopped so that it

[5] This was actually the first experiment in this series of studies. It was the failure to replicate Wynn (1992) in this condition that led us to realize the significance of the object-first/screen-first manipulation. However, the infants in this condition received only two familiarization trials. The decision to include two other familiarization trials with screen (to familiarize infants with the screen, one reason why they might be failing) was taken after this condition had been completed.

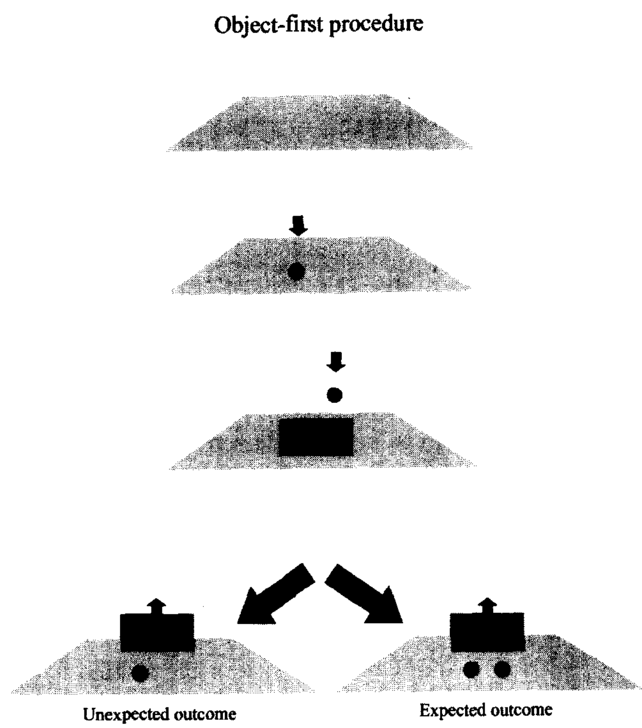

Figure 1. Schematic drawing of the test trials of the object-first condition.

was partially visible just above the top of the screen. The object was then jiggled for 5 s and then lowered to the stage floor, tapped five times on the floor, and released. After the second object was released, the screen was then lowered to reveal either one object (impossible outcome) or two (possible outcome). As soon as the experimenter lowered the screen, she signaled the observer to begin measuring the participant's looking at the display by saying "now."

Test Trials (Screen-First). All steps in the test trials of this condition were exactly the same as in the object-first condition except for the following. Figure 2 shows the test trials of the screen-first condition. After the participant was shown the empty stage, the screen was raised before any objects were introduced. Then the participant saw a single object lowered by its string toward the stage floor behind one end of the screen, followed by another identical object lowered in the same manner behind the screen. The trials then unfolded as in the object-first condition.

In the impossible (single-object) outcomes for half the participants in each condition, the object that was revealed after the screen was lowered was the

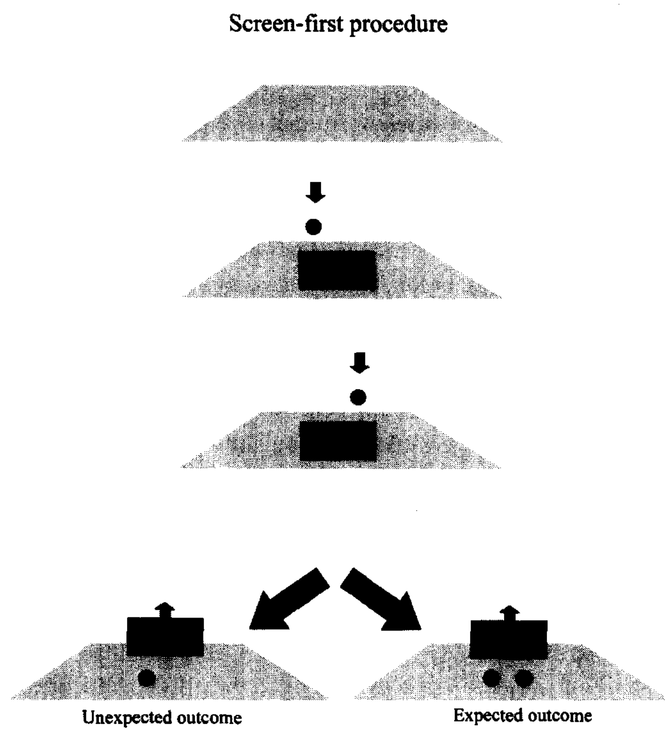

Figure 2. Schematic drawing of the test trials of the screen-first condition.

object that had been seen to be lowered onto the stage first (in full view in the object-first condition) and for the other half of the participants it was the object that had been lowered second (behind the screen in the object-first condition). That is, for half of the participants in each condition, the object that disappeared had been seen sitting on the stage floor and for the other half, the object that disappeared was last seen going behind the screen. In both conditions, there were three pairs of test trials. They alternated between one and two object outcomes in two orders (1-2-1-2-1-2 or 2-1-2-1-2-1). Order and side of the single-object trials (left, right) were counterbalanced across participants. Figure 1 schematizes the test trials in the two conditions with the single-object outcome (impossible) due to the removal of the second object lowered.

Results

Overall, 59 of the 64 participants were videotaped. The videotapes were coded by offline observers who were completely blind to the experimental conditions. The online–offline interscorer reliability was 93%. To ensure that offline observers were indeed blind to the experimental trials, we had an experienced observer guess the outcomes (1,2,1,2,1,2 or 2,1,2,1,2,1) in the test trials by viewing 15 randomly selected participants. He guessed the order of outcomes correctly in 8 of the 15 cases, which was no different from chance performance.

Both familiarization trials and test trials were presented in pairs consisting of a one-object outcome and a two-object outcome. For each pair of trials, we calculated a difference score: looking time on one-object outcome minus looking time on two-object outcome. Positive values represent longer looking at one-object trials; negative values represent longer looking at two-object trials. For all experiments in this paper, statistical tests were performed on the difference scores as calculated above. An alpha level of .05 was used for all statistical tests reported in the Results sections of each experiment.

Preliminary analyses examined the effects of order of outcome (one object first, two objects first), side of single-object outcomes (left, right), and sex (male, female) on the difference scores. As there were no effects of these variables, all subsequent analyses collapsed over them.

Test Trial Analysis. A first analysis examined the test trials only, for all those infants who completed all three sets of test trials (n = 58/64).[6] A 2 × 3 analysis of variance (ANOVA) examined the effects of trial pair (first, second, third) and condition (object-first, screen-first) on the difference scores. There was a main effect of condition: $F(1, 56) = 4.06$, $p = .049$: The difference score was greater in the object-first condition ($M = 1.56$, $SD = 2.02$) than in the screen-first condition ($M = .27$, $SD = 2.79$), reflecting a greater preference for the impossible outcomes in the object-first condition. This pattern of preference was sustained over all three pairs of trials (see Figure 3).[7] There were no effects of trial pair, $F(2,112) = .017$, $p = .98$, nor any interactions involving this variable.

The difference scores in the test trials on each condition were compared to 0. Overall, the difference scores in the object-first condition were greater than 0, $t(31) = 3.73$, $p = .001$, two-tailed, whereas those in the screen-first condition were not, $t(31) = .90$, $p = .37$, two-tailed. The analysis of the test trials revealed that only the infants in the object-first condition showed a significant preference for the impossible one-object outcome.

[6] Unless otherwise specified, as in this first test-trial-alone analysis, and the first test-trial-alone analysis of Experiment 2, all other statistical analyses were performed on the original data (difference scores) from all participants in each experiment. It is sometimes the case that values were missing in the last pair of test trials. For this test-trial analysis in Experiment 1, data from 6 participants, those who finished only two rather than three sets of test trials, were excluded (n = 58). For Experiment 2, data from 10 participants, those who finished only two rather than three sets of test trials, were estimated (n = 32).

[7] For all experiments, the means reported in Figure 3 include data from all participants in each experiment, not only from the participants that finished all three sets of test trials.

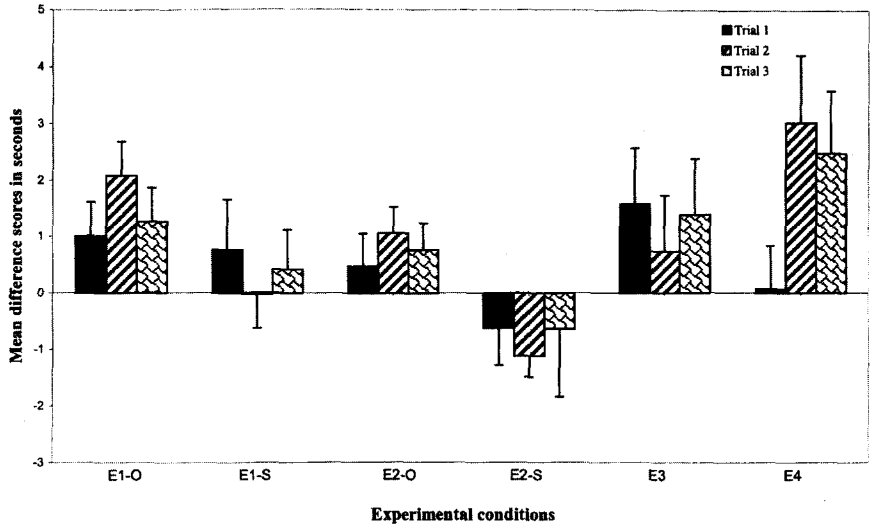

Figure 3. Mean difference scores in seconds for Trials 1, 2, and 3, for each experimental condition: Experiment 1, object-first (*n* = 32); Experiment 1, screen-first (*n* = 32); Experiment 2, object-first (*n* = 16); Experiment 2, screen-first (*n* = 16); Experiment 3 (*n* = 16); and Experiment 4 (*n* = 16). E1-O = Experiment 1, object-first condition; E1-S = Experiment 1, screen-first condition; E2-O = Experiment 2, object-first condition; E2-S = Experiment 2, screen-first condition; E3 = Experiment 3; E4 = Experiment 4.

Familiarization/Test Comparison. The preferences for the one-object outcome in the test trials must be interpreted in light of the preferences in the familiarization trials. Paired t-tests analyzed the difference scores of familiarization and test trials for each condition (see Table 1). In the object-first condition, there was a significant difference between the difference scores in the familiarization trials ($M = -.51$, $SD = 2.61$) and test trials ($M = 1.40$, $SD = 2.12$), $t(31) = -3.24$, $p = .003$, two-tailed. In contrast, in the screen-first condition, there was no significant difference between the difference scores in the familiarization ($M = -.87$, $SD = 3.20$) and test trials ($M = .44$, $SD = 2.73$), $t(31) = -1.78$, $p = .084$, two-tailed.[8]

In sum, infants in both the object-first condition and in the screen-first condition looked longer at two-object displays in the familiarization trials, but only the

[8] The means reported here, and on Table 1, differ from those reported in the test-trial-alone analysis above, because they reflect data from all 64 participants. The test-trial-alone analysis includes data from only 58 of the 64 participants, those who completed all three pairs of test trials.

Table 1. Mean Difference Scores (LT1 − LT2) for Experiments 1, 2, 3, and 4 as a Function of Trial Type

Experiment: Condition Type	*n*	Trial Type: Fam	Trial Type: Test
Experiment 1: Object-first, 8 mos.	32	−51 (2.61)	1.40 (2.12)***
Experiment 1: Screen-first, 8 mos.	32	−87 (3.20)	.44 (2.73)
Experiment 2: Object-first, 8 mos.	16	−1.43 (3.03)	.77 (.82)**
Experiment 2: Screen-first, 8 mos.	16	−1.12 (1.85)	−.86 (1.27)
Experiment 3: Screen-first, 10 mos.	16	−1.68 (3.07)	1.20 (2.69)*
Experiment 4: Split-screen-first, 8 mos.	16	−.89 (2.22)	1.90 (2.48)***

Note. The values represent the mean difference scores from original data for all infants (whether they finished three or two sets of test trials) in familiarization and test trials for each Experiment. Values enclosed in parentheses represent standard deviations. Statistical significance: familiarization vs. test comparison.
$^*p < .05$. $^{**}p < .02$. $^{***}p < .005$.

infants in the object-first condition differentiated the familiarization trials from the test trials.

Analysis of Object Removed. Recall that, for half of the participants, the object removed on impossible trials was the first object lowered, and for the other half, it was the second object lowered. This manipulation has different theoretical significance in the two conditions.

Object-First Condition: Here, the infants saw the first object in full view before it was occluded by the screen. Infants' sensitivity to whether the impossible outcome consisted of the seen-object alone or the unseen-object alone bears on whether infants are representing the objects in their precise locations.

A 2 × 2 ANOVA examined the effects of object removed (first lowered, second lowered) and trial type (familiarization, test) on the difference scores. As reported above, there was a main effect of trial type, $F(1, 30) = 10.29$, $p = .003$. The new result was that there was no effect of object removed, $F(1, 30) = 2.32$, $p = .138$, and, more important, there was no interaction between the two variables, $F(1, 30) = .37$, $p = .548$. Infants did not differentiate the familiarization and test trials better when the object removed was the one they had actually seen resting on the stage floor.

Screen-First Condition: Given that infants overall failed in the screen-first condition, and given that they never saw an object on the stage floor during the 1 + 1 event, we would not expect to see an effect of which object was removed. However, this analysis is of theoretical importance for another reason. The infants in the condition with the first object removed had only one pair of familiarization trials whereas those in the condition with the second object removed had two familiarization trials. Thus this analysis examines whether the number of familiarization trials affects the looking-time preferences in the test trials.

A 2 × 2 ANOVA was carried out to examine the effects of object removed (first lowered [1 familiarization pair], second lowered [2 familiarization pairs]) and trial type (familiarization, test) on the difference scores. There was a main effect of object removed, $F(1, 30) = 6.73$, $p = .015$. Overall, there was a significant difference between the difference scores in the group for which the object removed was the first one lowered ($M = -.23$, $SD = 2.28$) and the group for which the object removed was the second one lowered ($M = 1.10$, $SD = 3.05$). This effect is due to the fact that the group that had one pair of familiarization trials showed a preference for the two-object outcome throughout the experiment, whereas the group that had both sets of familiarization trials had a preference for the one-object outcome, both in the familiarization and in the test trials. There was no effect of trial type, $F(1, 30) = 3.12$, $p = .088$. Moreover, the important result is that there is no interaction between the variables trial type and object removed, $F(1, 30) = .42$, $p = .523$. As reported above, there is no evidence that the infants differentiated the outcomes of the test trials from the outcomes of the familiarization trials, nor did it matter whether they had seen two pairs of familiarization trials or one pair.

Discussion

As predicted by the Object-file model, infants in the object-first condition performed better than those in the screen-first condition. The infants' success in the object-first condition is an important result in itself, because outright success in a $1 + 1 = 2$ or 1 task has only once been found in the studies with 5-month-olds (Simon et al., 1995). The infants in Wynn (1992) and in Koechlin et al. (1997) were sensitive to the test events because they differentiated their patterns of looking to outcomes of one object or two objects according to whether the test events were $1 + 1$ or $2 - 1$, but they did not succeed outright on the $1 + 1$ trials. Apparently, success at object-first $1 + 1 = 2$ or 1 tasks is fragile at 5 months of age; Experiment 2 will address the robustness of success at 8 months of age.

Still, sensitivity to number in these experiments is well established. It is obtained in object-first designs with infants as young as 5 months and under conditions of varied amounts of familiarization (Experiment 1, this paper; Koechlin et al., in press; Simon et al., 1995; Wynn, 1992). In Experiment 1, different objects were used than in the previous studies and objects were introduced from the top instead of the side. Success across such different experimental conditions indicates that young infants certainly distinguish arrays with one and two objects in them, and are capable of representing the outcome of insertions and removals of objects into displays already containing hidden ones.

The important new result of Experiment 1 is the 8-month-olds' failure in the screen-first condition. If infants were simply counting the objects (as on any symbolic model, including the accumulator model), it is difficult to see why the screen being introduced should make such a dramatic difference. The Object-file model, in contrast, predicts this result, on the assumption that a model of an object actually seen on the stage is more robust than a model constructed from memory, and thus can better accommodate a subsequent update. In addition, the difference in success between the object-first and the screen-first conditions of Experiment 1 provides evidence for the model robustness assumptions, as does the difference in success between Baillargeon et al.'s (1994) $2 + 1 = 3$ or 2 condition and their $1 + 1 + 1 = 3$ or 2 condition.

Infants did not perform better when the impossible outcome was created by removing the first lowered object, even in the condition in which this object had actually been seen on the stage before being covered by the screen (object-first condition). This result may be because infants could well predict the location on the stage for both objects in this study, whether seen on the stage or not, given that the objects were lowered from above. It also may be because the match/mismatch between representations of test arrays and the representations of the outcome arrays do not take into account the exact location of the objects on the stage, being concerned solely with one-to-one correspondence. We favor the latter interpretation because of Koechlin et al.'s (in press) finding that 5-month-old infants succeeded equally well at addition/subtraction tasks when the outcomes were stationary and in predictable locations and when the outcomes were in unpredictable locations on a rotating plate behind the screen. Experiment 4 returns to the issue of the role of location representations in infants' mental models of the test events.

The failure in the screen-first condition, in the light of success in the object-first condition, adds to the support from set size considerations summarized in the introduction for the Object-file model. However, the contrast between the conditions in Experiment 1 is not ideal because one of the screen-first conditions had fewer familiarization trials than the other three conditions. It is unlikely, however, that the failure of this group of infants was due to this procedural difference from the object-first condition because those screen-first participants who received two sets of familiarizations also failed.

Experiment 2 seeks to replicate the pattern of failure on the screen-first condition in the face of success in the object-first condition under conditions that remove any differences in familiarizations between the two groups. Experiment 2 also explores the reliability of the infant-addition/subtraction results by exploring a markedly different methodology. In the infant looking-time studies, the experimenter is usually invisible to the infant, and objects are placed onto stages by disembodied hands. This experience is decidedly weird for the infants, most of whom must never have experienced anything remotely like it (most infants under 12 months have never seen a puppet show). Hauser, McNeilage, and Ware (1996) found that adult semi-free-ranging rhesus macaques, tested under totally different conditions, succeed even more reliably than do infants in a series of *screen-first* $1 + 1 = 2$ or 1 tasks. In this paradigm, the experimenter is fully visible, like a magician, putting objects into a fully visible stage. The magic trick is done by having a pocket in the back of the removable screen. Uller (1996), and Uller, Carey, and Hauser (1998) found that adult laboratory cotton-top tamarins also succeed on screen-first $1 + 1 = 2$ or 1 task under these conditions. These are adult animals and maturational factors may limit infant performance in these tasks. But it is also possible that the fully visible conditions of the monkey studies may promote success in both the screen-first and object-first conditions. Experiment 2 compares screen-first and an object-first versions of a $1 + 1 = 2$ or 1 task, using the procedures of the monkey experiments. Finally, the objects in Experiment 2 are small dolls, not the unfamiliar sand-pile objects of Experiment 1.

EXPERIMENT 2

Method

Participants. Thirty-two full-term 8-month-old infants (15 boys, 17 girls) ranging in age from 7 months, 9 days to 8 months, 16 days (M_{age} = 8 months, 3 days) were tested. Twenty-two additional participants were excluded because of fussiness (4), experimenter error (1), equipment failure (2), or distraction due to experimenter presence (15). The criteria for participant exclusion because of distraction due to experimenter presence were as follows: Participants stared at experimenter or in the direction of the experimenter instead of looking at the display area (a) from the very beginning of familiarization trials, as experimenter was presenting the objects and placing them onto the stage area, not looking at these events; (b) during familiarization outcomes, as observers were signaled to start timing, not looking at the display for 10 consecutive seconds; or (c) during test trial outcomes, as observers were signaled to start timing, not looking at the display for 10 consecutive seconds. Participants were contacted and compensated as in Experiment 1.

Materials and Apparatus. The experimental display was an opaque grey foam core box measuring 25 cm × 50 cm × 26 cm. The box had no top. The front part functioned as a screen which slid up and down concealing whatever was placed inside the box. The stimuli were two brightly colored green dinosaurs with blue pants and light blue striped shirts. They measured 17.7 cm high, 10.2 cm wide, and 8.9 cm from abdomen to tail.

This experiment took place on the same stage described in Experiment 1. Unlike standard preferential looking-time experiments, however, the experimenter stood on the side of the box in full view in front of the participant. The experimenter dressed in a laboratory coat with two pockets in which the dinosaur dolls were placed. The participant could see the experimenter draw the objects from the pockets and place them into the box, and could also see her draw such objects from the box into the pockets. The screen that covered the box was always placed behind the box whenever it was not needed, namely, in the familiarization trials without the screen and when the screen was removed to reveal the outcomes.

During the experiment, the laboratory was darkened and the stage was lit directly from above. The participant was lit indirectly for videotaping by lamps in front of and on either side of the stage. The participant sat facing the box with his or her head about 200 cm away and his or her eyes slightly above the floor level of the box. The participant's parent/guardian sat on the participant's left facing away from the box. Parents were instructed to interact with the participant as little as possible as described in Experiment 1. Each participant was observed by one live observer and was also videotaped for offline observation. Live observation, looking-time data collection, and offline coding proceeded as described in Experiment 1.

Procedure

As in Experiment 1, each participant was randomly assigned to one of two conditions: the object-first condition (n = 16; M_{age} = 8 months, 0 days) or the screen-first condition (n = 16, M_{age} = 8 months, 6 days). Both the object-first and the screen-first conditions were modeled on Experiment 1 with two major differences. First, objects were lowered into a box by the experimenter in full view. The experimenter drew the objects from pockets of her laboratory coat. When the objects were in the pockets, their tips were visible. This allowed participants to have more spatiotemporal information about the coexistence of the objects than is the case in all other previous addition/subtraction studies, where objects come from behind the stage.

Each condition consisted of three phases as in Experiment 1: an introductory phase, familiarization trials, and test trials. Condition (screen-first, object-first) was between participants variable, whereas trial-type (familiarization, test) and outcome (one object, two objects) varied within participants.

Introductory Phase (All Infants). Every participant was given a 60-s exposure to one of the objects before the experiment began. The experimenter held the object and drew the participant's attention to it. The experimenter then handed the object to the participant, allowing him or her to play with it. No participant was reluctant to grasp the object. At the end of this phase, the object was placed back into the pocket of the experimenter's laboratory coat.

Familiarization Trials (All Infants). As in Experiment 1, there were two pairs of familiarization trials. The first pair took place entirely within the participant's view. The second pair was partially hidden by the front side (screen) of the box. The experimenter called to the participant: "Look, baby! Look at the dinosaur!" to keep his or her attention at the objects and at what was happening inside the box. The first pair of familiarization trials consisted of a single-object trial and a double-object trial. In the single-object trial, the experimenter removed one dinosaur from one of the pockets and introduced it into the display. The object was slowly lowered toward the box floor and then stopped when it was about halfway down. It was then jiggled for about 5 s and then lowered onto the box floor. When the object reached the box floor it was tapped five times and then released. The hand that held the object was slowly retracted from behind the screen. During the familiarization and the test trials, for both conditions, after the experimenter withdrew her hands from the box area, she lowered her arms along side and she fixed her look to the ground in front of her to provide no cues to the participant. The experimenter then signaled the observer to record the participant's looking at the stage area by saying "now." After the participant looked away from the display, as signaled by a beep from the computer, the experimenter removed the stimulus by grabbing the object and placed it back into the pocket from which it had been taken. In the double-object trials, the experimenter removed the two dinosaurs from the pocket simultaneously, showed them to the participant, and then introduced them into the stage at the same time. The objects were lowered and tapped simultaneously on the box floor, as looking times were recorded. The objects were then removed from the stage and replaced in the pocket.

The second pair of familiarization trials happened behind a screen. As in the first pair of trials, there were two types of trials, one with one object and one with two objects. Both trials began with the empty open box. Then the experimenter placed the screen to hide the box interior. Both trials, one with a single object, and one with two objects, proceeded the same way as in the two trials with no screen.

There were four familiarization trials—one of each type. The pair of trials without a screen always preceded the trials with a screen. Familiarization trials alternated between one and two objects in two orders (1-2- 2-1 or 2-1-1-2). Order and side of the single-object trials (left, right) were counterbalanced across participants.

Test Trials (Object-First). As in Experiment 1, there were six test trials that immediately followed the familiarization trials. During the test trials, the experimenter took care to draw the participant's attention to the objects as they were drawn from the pockets of the laboratory coat by speaking to the participant. First, participants were shown the empty stage. Next, a single object was placed into one side of the box; it stopped halfway down, was jiggled for 5 s and then lowered, tapped five times on the box floor, and released. The hand that held the object retracted and then the screen was placed to hide the object. Next, a second object was drawn from the other pocket of the laboratory coat and placed into the other side of the box behind the screen. This object was also stopped so that it was partially visible just above the top of the screen. The object was then jiggled for 5 s and then lowered to the floor, tapped five times, and released. After the second object was released, the screen was then removed to reveal either one object (impossible outcome) or two (possible outcome). The "magic trick" of one-object outcome was performed by placing the second object presented into a hidden pocket attached to the back side of the screen. The experimenter then signaled the observer to begin measuring the participant's looking at the display by saying "now."

Test Trials (Screen-First). As in Experiment 1, there were six test trials that immediately followed the familiarization trials. Here, the participants saw the empty stage. Then the screen was placed before any objects were introduced. After the screen was in place, the objects were lowered behind the screen much the same way as in the object-first condition.

In both conditions, test trials alternated between one and two object outcomes in two orders (1-2-1-2-1-2 or 2-1-2-1-2-1). Order of test outcomes (one object first, two objects first) and side of the single-object outcome (left, right) were counterbalanced across participants. The object removed in the one-object outcomes was always the second one lowered into the box. The side of the one-object outcome was always the same as the one-object trials in the familiarization.

Results

In the present experiment, all participants were videotaped. Interscorer reliabilities between the primary online observer and the offline observer of the videotapes were 91%.

Preliminary analyses examined the effects of order of outcome (one object first, two objects first), side of single-object outcomes (left, right), and sex (male, female) on the difference scores. There were no effects of any of these variables. All subsequent analyses collapsed over them.

Test Trial Analysis. As in Experiment 1, a first analysis examined the test trials only. Data from all 32 participants were included in this analysis. However, 10 participants failed to complete all three pairs of test trials. These participants come from the screen-first group, which reflects the greater difficulty of this condition.

To estimate the missing data for the last pair of test trials of these 10 participants, a series of steps were taken. First, a multiple regression was carried out to determine whether the data from these 10 participants were missing at random. A variable was coded for missing data. This variable was regressed on the difference scores. The amount of variance accounted for by missing data was not significant (Shrunken $R^2 = .081$). This finding implies that the data were missing at random.

Second, to further test the possibility of differences between the group of participants that had missing data and the group of participants that did not have any missing data, independent-samples t-tests were performed on the difference scores of test Trial sets 1 and 2. There was no significant difference between the difference scores on either trial sets of those participants who finished all three sets of test trials (Trial set 1: $M = -.18$, $SD = 3.57$ Trial set 2: $M = -1.13$, $SD = 1.29$) and those who finished only two sets (Trial set 1: $M = -.87$, $SD = 2.11$; Trial set 2: $M = -1.11$, $SD = 1.68$): Test set 1, $t(14) = .63$, $p = .687$, two-tailed; Test set 2, $t(14) = -.03$, $p = .977$, two-tailed.

Third, to replace the missing data values, a multiple regression analysis was performed on the difference scores. Condition (object-first, screen-first), missing (0 = not missing, 1 = missing), order (one object first, two objects first), side (left, right), sex (male, female), and trial number (two variables were created using dummy-coding to represent trial number) were regressed on the difference scores. The regression equation used to predict the missing values was the following: $\hat{Y} = 1.13 - 1.43$ (condition) $- .34$ (missing) $- .009$ (order) $+ .22$ (sex) $- .45$ (side) $+ .04$ (trial 1) $+ .03$ (trial 2). The predicted difference scores were used to replace the missing values of the third set of test trials. The following analysis was conducted both ways, namely, with and without the replaced missing values.

A 2×3 ANOVA examined the effects of test trial pair (first, second, third) and condition (object-first, screen-first) on the difference scores. With the missing values replaced, there was a main effect of condition, $F(1, 30) = 27.39$, $p = .000$: The difference score was greater, $M = .77$, $SD = 2.03$, in the object-first condition than in the screen-first condition, $M = -.86$, $SD = 2.06$, reflecting a greater preference to look longer at one-object outcomes than at two-object outcomes in the object-first condition. There was no effect of trial pair, $F(2, 60) = .00$, $p = .997$, nor was the interaction involving these variables significant, $F(2, 60) = .46$, $p = .635$. When the missing data were not replaced, again, there was a main effect of condition, $F(1, 20) = 26.43$, $p = .002$: The difference score was greater, $M = .77$, $SD = 2.03$, in the object-first condition than in the screen-first condition, $M = -.83$, $SD = 2.31$, reflecting a greater preference to look longer at one-object outcomes than at two-object outcomes in the object-first condition. There was no effect of trial pair, $F(2, 40) = .02$, $p = .980$, nor was the interaction involving these variables significant, $F(2, 40) = .38$, $p = .690$. Thus, the results are the same whether the missing values were replaced or not.

The difference scores in the test trials of each condition were compared to 0. The analyses of the test trials revealed that the infants in the object-first condition showed a preference for the impossible one-object outcome in the test trials, $M = .77$, $SD = .82$, $t(15) = 3.78$, $p = .002$, two-tailed. The infants in the screen-first condition, however, looked longer at the two-object outcome in the test trials, $M = -.86$, $SD = 1.27$, $t(15) = -2.73$, $p = .016$, two-tailed.

Familiarization/Test Comparison. Paired *t*-tests analyzed the difference scores of familiarization and test trials for each condition (see Table 1). In the object-first condition, there was a significant difference between the difference scores in the familiarization trials, $M = -1.43$, $SD = 3.03$, and the test trials, $M = .77$, $SD = .82$, $t(15) = -2.85$, $p = .012$, two-tailed. In the screen-first condition, however, there was no significant difference between the difference scores in the familiarization, $M = -1.12$, $SD = 1.85$, and the test trials, $M = -.86$, $SD = 1.27$, $t(15) = -.39$, $p = .702$, two-tailed.

In sum, participants in both conditions looked longer at the two-object outcomes in the familiarization trials, but only the infants in the object-first condition looked longer at the one-object outcomes in the test trials.

Discussion

Experiment 2 exactly replicated Experiment 1. Infants looked longer at the impossible outcome of one object if they had seen the first object placed in the box before the screen was raised and the second object introduced, whereas they failed to do so if they saw the objects placed one at a time onto an already screened box floor. Given that the two groups of infants in Experiment 2 began with exactly the same familiarization trials, this replication relieves the lingering worry that the failure of the screen-first infants in Experiment 1 may have been due to the fact that half of them had fewer familiarization trials than did the object-first infants.

Taken together, the results of Experiment 1 and Experiment 2 relieve other worries about $1 + 1 = 2$ or 1 experiments with designs similar to those of these studies. When infants have a preference for two objects on the familiarization trials, they will have looked longer at, and thus become more familiar with, arrays of two objects than arrays of one object by the end of these trials. Then, when the test trials come along, the looking-time preference for one object could be a novelty preference rather than a looking-time preference for the impossible outcome. From the point of view of this alternative explanation of success on $1 + 1 = 2$ or 1 studies, however, the screen-first and the object-first conditions are identical. The failure in the screen-first condition in both Experiments 1 and 2 shows that there is no overall tendency to switch preferences from two objects to one object with repeated viewing of pairs of one object and two object trials. In the screen-first conditions, infants maintain their looking-time preference for two objects. At 8 months of age, at least, the looking-time preference for one object in an object-first $1 + 1 = 2$ or 1 event seems genuinely a preference for the impossible outcome.

The failure of 8-month-old infants in the screen-first condition differs from the findings obtained from different species of monkeys (rhesus macaques: Hauser et al., 1996; cotton- top-tamarins: Uller, 1996; Uller et al., 1998) who succeed on screen-first $1 + 1 = 2$ or 1 tasks exactly like that of Experiment 2. Thus, it is not the case that the fully visible conditions of the monkey studies promote success; the pattern of performance was identical in the standard hidden experimenter procedure of Experiment 2 and the visible experimenter paradigm borrowed from the primate studies. Most probably, the difference between the monkey and the human infant results derives from the fact that the monkeys were adults. Immaturity of information processing capabilities most probably contributes to the infant's inability to create robust object-file representations under the conditions of the screen-first experiments. If so, then slightly older infants would be expected to succeed in a screen-first $1 + 1 = 2$ or 1 study, a prediction tested in Experiment 3.

The success in the object-first condition of Experiment 2 expands dramatically the range of circumstances in which infants succeed on such tasks. No looking-time experiment with infants that we are aware of has been carried out under these circumstances (with the experimenter present and visible.) There are many reasons not to run infant looking-time studies as in Experiment 2. Fully 28% of the participants could not be enticed to watch the events; they kept looking at the experimenter's face instead of at the objects and the box. On the other hand, only

7% became fussy, versus 29% of the participants in Experiment 1. It seems possible that fussiness in standard experiments with hidden experimenters partially reflects lack of normal social interaction.

Several arguments favor the standard hidden-observer procedures. First, there is less opportunity for experimenter bias. We took great care to guard against experimenter bias in Experiment 2; the experimenter looked at the floor during recording of the infant looking at the outcome events and could not tell whether the infant was looking at the object on the stage or not. Still, in the standard procedure, the infant cannot see the experimenter and the experimenter cannot see the infant. Second, in most experiments, the magic tricks needed are too intricate to be carried out by a fully visible magician. The similarity in results between Experiments 1 and 2 lend generality to the results obtained in standard hidden-experimenter procedures.

The failure in the screen-first version of Experiment 2 is especially striking, because the objects that were placed into the box were visible in the experimenter's pockets when they were not in the box. The infants had ample opportunity to encode that there were exactly two objects involved in all of these events, two objects that were moved from pockets to box and back again. The failure to use this information in the screen-first condition provides evidence relevant to any model of performance. If the Object-file model is correct, and the infants are creating a mental model of the actual objects on the stage floor, then the main information they use is what they see placed there. If the Integer-symbol model is correct, then infants are counting only what they see placed onto the stage.

The difference in performance in the object-first and screen-first conditions supports the Object-file model by the same arguments offered for the results of Experiment 1. The two conditions place identical demands on a counting algorithm; a counter/accumulator must be incremented to a representation of two during the test events, and to a representation of one or two in the outcomes, and the results of the two counts compared. Thus, the Integer-symbol model has no natural way to explain the consistent failure in the screen-first condition in the face of success in the object-first condition (Experiments 1 and 2). Again, the model robustness assumptions receive support; two updates in imagery exceed the processing capabilities of infants under these conditions whereas one update in imagery does not.

EXPERIMENT 3

If the 8-month-olds' failure in the screen-first conditions reflects information processing limitations in creating models of the objects in the array, older infants should succeed. Baillargeon et al. (1994) reported a success on the part of 10-month-olds at a screen-first version of a $1 + 1 = 2$ or 3 task. In this case, the impossible outcome (3) is also more numerous than the possible outcome (2), which may contribute to the infant's success on this task. However, Baillargeon et al. found no baseline preference for arrays of three over arrays of two. Thus, it seems reasonable to expect that 10-month-olds will succeed at the screen-first task of Experiment 1. Experiment 3 establishes whether this expectation is warranted.

Method

Participants. Sixteen full-term, 10-month-old infants (9 boys, 7 girls), ranging in age from 9 months, 14 days to 10 months, 17 days (M_{age} = 10 months, 1 day) were tested. Six additional participants were excluded because of fussiness (4) and extraordinarily long looking during the familiarization trials (2). Participants were contacted and compensated as in Experiment 1.

Materials. The materials were exactly as in Experiment 1.

Procedure

The procedure was exactly as the screen-first condition in Experiment 1 with two sets of familiarization trials. The object that was displayed in the single-object trials was always the object that had been lowered first behind the screen.

Results

Seven of the 16 participants were observed by both a live observer and a second offline observer from the videotaped record of the experimental session. Interobserver reliability scores averaged 95%.

Preliminary analyses examined the effects of order of outcome (one object first, two objects first), side of single-object outcomes (left, right), and sex (male, female) on the difference scores. As there were no effects of any of these variables, all subsequent analyses collapsed over them.

Test Trial Analysis. The difference scores in the test trials were compared to 0. As can be seen in Figure 3, the difference scores on each of the three test trials was greater than 0, reflecting longer looking at the impossible outcome of one object. However, the average difference score over the three test trials, $M = 1.20$, $SD = 2.69$, was not significantly greater than 0, $t(15) = 1.79$, $p = .09$, two-tailed. What is important, of course, is whether the pattern of preference on the test trials differed from that of the familiarization trials.

Familiarization/Test Comparison. The preferences for the one-object outcome in the test trials were interpreted in light of the preferences in the familiarization trials. A paired t-test analyzed the difference scores of familiarization and test trials (see Table 1). There was a significant difference between the difference scores in the familiarization trials, $M = -1.68$, $SD = 3.07$, and test trials, $M = 1.20$, $SD = 2.69$, $t(15) = -2.51$, $p = .024$, two-tailed. Thus, a preference for two-object displays on the familiarization trials was reversed in the test trials on which the infants had a looking-time preference for the impossible one-object outcomes.

In sum, the 10-month-olds in Experiment 3, unlike the 8-month-olds of Experiments 1 and 2, succeeded at differentiating the familiarization trials from the test trials of a screen-first 1 + 1 = 2 or 1 procedure.

Discussion

These data confirm those of Baillargeon et al. (1994), demonstrating that 10-month-old infants expect two objects as the outcome of a 1 + 1 addition performed in a screen-first procedure. It is clear that the capacity to construct models of multiple unseen objects is acquired gradually. During the period between 5 and 8 months of age, success at object-first 1 + 1 procedures becomes more robust as 8-month-olds succeed outright whereas younger infants often merely differentiate the outcomes of 1 + 1 from 2-1 events (Koechlin et al., in press; Wynn, 1992). During the period between 8 and 10 months, the infant overcomes the problems posed by a screen-first 1 + 1 task, but a screen-first 1 + 1 + 1 task still exceeds the capacity of 10-month-olds (see also Baillargeon et al., 1994).

We interpret the 8-month-old's failure in the screen-first condition in terms of the model robustness assumptions laid out in the introduction. Infants of this age cannot perform two updates of models of objects constructed in imagery under the conditions of these experiments. However, in other circumstances, young infants have been shown to be able to create models of two objects in imagery. Spelke et al. (1995) showed 5-month-old infants objects emerging alternately from the left side of the left of two screens and from the right side of the right of two screens; the objects were never seen together and no object ever appeared in the middle. Under these circumstances, the infants established representations of precisely two objects behind the screens, one behind each; they looked longer at outcomes of one object than at outcomes of two objects when the screens were removed.

Spelke et al.'s (1995) procedure differed from the standard infant-addition/subtraction procedures in many respects. The infant was given much more and repeated spatiotemporal evidence that there were two objects in the event; the objects were shown repeatedly emerging from and reentering behind each screen while not appearing in the middle. In the standard 1 + 1 test event, the spatiotemporal evidence that there are two objects is much more subtle; the infant must establish a representation of the first object behind the screen and infer that the second object is numerically distinct from it because it is coming from a different location. Thus, the additional spatiotemporal information in the Spelke et al. procedure may contribute to the earlier success on this screen-first, two-update event.

It is also possible, however, that the mere fact that there are two screens instead of one helps the infant. Each screen provides a perceptually available marker of the location of a hidden object, which could serve the role of helping to individuate it and thus to maintain it in memory. If this is so, then infants of the same ages as those in Experiments 1 and 2 might succeed in the screen-first 1 + 1 procedures of those studies if there were two separate screens involved in the test events. Experiment 4 tests this prediction.

EXPERIMENT 4

Method

Participants. Participants were 16 full-term 8-month-old infants (7 boys, 9 girls), ranging in age from 7 months, 15 days to 8 months, 14 days (M_{age} = 8 months, 0 day). Eight additional participants were excluded because of fussiness (7) and extraordinarily long familiarization looking times (1). Participants were contacted and compensated as in Experiment 1.

Materials. The materials were the same as in Experiment 1, except that the single screen was replaced by two, brightly colored screens, each measuring 35 cm × 35 cm. These screens were lavender and contrasted clearly with the blue stage and the black background of the rest of the stage display. When placed on the stage, the two screens were separated by a distance of 16 cm. The screens were introduced and withdrawn by the experimenter through an opening in the ceiling of the stage.

Procedure

The procedure was identical to that of the screen-first condition in Experiment 1. There was an introductory section, which served to familiarize the participant with the experimental stimuli. Next there were two pairs of familiarization trials in which participants were familiarized with the apparatus including the two screens, and during which familiarization preferences were measured. The familiarization trials served to alert the participants to the fact that one object or two objects would be involved in these events, and that objects lowered on the stage would be seen where expected. Of course, they provided no information about what to expect in the crucial 1 + 1 = 2 or 1 test trials.

Unlike the participants in Experiment 1, participants in Experiment 4 were introduced to the screens after the introductory section and before the familiarization trials. The participant was first shown an empty stage, then the two screens were lowered into place, side by side on the stage floor. The experimenter drew the participant's attention by calling out to the participant as the screens were lowered. Once the screens were on the stage floor the experimenter tapped her hand on the stage floor, first to the left of the two screens, then in the space separating the screens, and finally she tapped the other end of the stage to the right of the screens. The partially hidden trajectory of the hand served as an additional source of information that there were two separate screens on the stage. After this event, the screens were removed and the familiarization trials were started.

Introductory Section. As in Experiment 1 each participant was given a chance to manipulate the object.

Familiarization Trials. There were two pairs of familiarization trials. The events of the first pair happened in full view while the events of the second pair involved two screens. Each pair consisted of a single-object trial and a double-object trial. In the double-object trials, the objects were lowered simultaneously. Each lowering event followed the time course of the events of Experiment 1; objects lowered to the level of the top of the screens, jiggled for about 5 s, lowered, and tapped on the stage floor. Before the second pair of trials, the screens were introduced simultaneously. For all familiarization trials involving screens, after the object(s) were lowered behind the screens, both screens were removed simultaneously before looking time to the display was measured. There were two orders of familiarization trials (1-2-2-1 or 2-1-1-2). Order (one object first, two objects first) and side of the single-object trials (left, right) were counterbalanced across participants.

Test Trials. Six test trials immediately followed the familiarization trials. Participants saw an empty stage and then both screens were lowered onto the stage simultaneously. After the two screens were in place, the experimenter lowered one object toward one of the screens. At the same time she drew the participant's attention to the object by calling out to the participant, while jiggling the object for about 5 s partially behind the top of the screen, before lowering it and tapping it on the stage floor. When the object was partially behind the top of the screen it was stopped and jiggled for about 5 s and then completely lowered behind the screen, tapped on the stage floor, and released. The experimenter's hand was then retracted from the display. The second object was then lowered behind the remaining screen the same way as the first. After the second object was in place and the experimenter's hand was retracted, both screens were removed simultaneously to reveal either one object (impossible outcome) or two (possible outcome). Looking times at the outcomes were measured by hidden observers. Each participant saw an alternating series of three pairs of test trials in one of two orders (1-2-1-2-1-2 or 2-1-2-1-2-1). The object that remained in the impossible outcome was always that object that had been placed first on the stage. Order (one object first, two objects first) and side of the single-object trials (left, right) were counterbalanced across participants.

Results

All participants' looking times were measured by one live observer and a second offline observer from the videotaped record of the experimental session. Interobserver reliability scores averaged 93%.

Preliminary analyses examined the effects of order of outcome (one object first, two objects first), side of single-object outcomes (left, right), and sex (male, female) on the difference scores. As there were no effects of these variables, all subsequent analyses collapsed over them.

Test Trial Analysis. The difference scores in the test trials were compared to 0. As can be seen in Figure 3, the difference scores on each of the three test trials was greater than 0, reflecting longer looking at the impossible outcome of one object, and the average difference score over the three test trials, $M = 1.90$, $SD = 2.48$, was significantly greater than 0, $t(15) = 3.07$, $p = .008$, two-tailed.

Familiarization/Test Comparison. The preferences for the one-object outcome in the test trials were interpreted in light of the preferences in the familiarization trials. A paired t-test analyzed the difference scores of familiarization and test trials (see Table 1). A significant difference was found between the difference scores in the familiarization trials, $M = -.89$, $SD = 2.22$, and the test trials, $M = 1.90$, $SD = 2.48$, $t(15) = -3.34$, $p = .004$, two-tailed. Thus, a preference for two-object displays on the familiarization trials was reversed in the test trials, on which the infants had a looking-time preference for the impossible one-object outcomes.

Discussion

Infants succeeded in Experiment 4. Its screen-first paradigm required two successive updates of the infants' initial mental model of the empty stage, just as did the screen-first paradigm of Experiments 1 and 2. Thus, the number of updates of a model constructed from perceptual evidence cannot be the sole variable explaining the difference in success between the object-first and the screen-first conditions of Experiments 1 and 2.

We hypothesize that infants succeeded in Experiment 4 because the two screens provided continuously perceptually available markers of location, which helped the infants individuate and maintain two object files in their model of the array. That is, they could encode one object behind screen A and the other object behind screen B.

Success in Experiment 4 in the face of failure in the screen-first conditions of Experiments 1 and 2 suggests that the model of the array that infants constructed in the earlier studies was "object behind the screen, another object behind the screen." With such an imprecise specification of location, the infants had some difficulty constructing or maintaining a representation of two numerically distinct objects behind the screen. Notice that the object-first condition of Experiments 1 and 2 required them to do this also, as do the experiments in Koechlin et al. (in press), Simon et al. (1995), and Wynn (1992)—success requires a model of two numerically distinct objects behind the single screen. Apparently, it is the conjunction of the number of updates required in the screen-first condition and the lack of perceptually available markers of separate locations that accounts for the failure in the screen-first condition of Experiments 1 and 2.

General Discussion

Experiments 1, 2, and 4 provide support for the model robustness assumptions. Mental models of the objects on the stage in experiments such as these are more robust if an object is actually seen on the stage floor rather than merely imagined there, as suggested by the contrast between the object-first and the screen-first conditions in Experiments 1 and 2. Also, each update of a mental model in imagery decreases robustness, as also suggested by the contrast between the object-first and the screen-first conditions and by the contrast between the 2 + 1 and the 1 + 1 + 1 conditions of Baillargeon et al. (1994). Finally, visible markers of distinct individuals increase the robustness of a model of two hidden objects, as suggested by the contrast between the two-screen, screen-first procedure of Experiment 4 and the single-screen, screen-first procedures of Experiments 1 and 2.

At the same time as providing evidence for the model robustness assumptions, these experiments also provide support for the Object-file model over the Integer-symbol models of the representations underlying infant success in the addition and subtraction studies. The order of placement of the screen and objects (Experiments 1 and 2), or the number of screens (Experiment 4), or the grouping of the objects (Baillargeon et al., 1994) would not be expected to make any difference to the process of incrementing an accumulator (or stepping through a mental list of numerons) as each new object is introduced into the array, nor to the processes by which the result of one count is compared with the result of another. In other words, the Integer-symbol models have no natural way of accounting for the effects of the manipulation of these studies, whereas they are predicted by the Object-file model.

An adherent of the Integer-symbol position might make the following response: the infant cannot count individuals she does not represent. The infant must make a model of the individuals on the stage floor before she can count them. So, yes, these manipulations affect the robustness of such models, but these models are merely the input to the counting process.

We fully agree that one cannot count what one has not individuated. But it is not necessary to make a model of the objects on the stage floor as input to a counting process. If one is counting marbles put into a cup, one need not make a model that contains a unique symbol for each marble one counts. One need only count. In the version of the Integer-symbol model we are contrasting with the Object-file model, objects are counted as they are introduced into the array. Of course, the infant must individuate the objects to succeed at such a count. The infant must establish the representations *an object, another object* in these events. The spatiotemporal evidence on the basis of which this is achieved would be that the second object comes from a different place from which the already counted object is represented to be.

It is possible that the mental models of the objects on the stage are input to a counting process. However, there is no evidence that it is the case. Rather, the evidence from the present studies, as well as from the set size considerations reviewed in the introduction, suggests that object-file models of the individuals in the array are being created and evaluated for consistency with object-file models of the outcome arrays.

These experiments lend empirical support to the hypothesis that object-file models underlie success in the infant-addition/subtraction studies. Even if this hypothesis is true, we would not want to conclude that human infants lack the analog, number line/accumulator representations shared so widely in the animal kingdom (see Gallistel, 1990, for a review), including in that peculiar animal, the human adult (see

Dehaene, 1997, for a review). Recently, Xu and Spelke (1997) showed that 7-month-old infants discriminate 8 from 16 dots, when total dot area and density of dots was controlled for. These numbers exceed the limitation of parallel individuation of object files. The claim of this paper is merely that the infant-addition/subtraction studies are unlikely to exploit a symbolic integer model of number. If so, they cannot provide evidence relevant to the nature of infant representation of integer.

Suppose the Object-file model of infant representations in the addition/subtraction events is correct. Does this mean that these experiments are irrelevant to the conceptual roots of human number representation? Not at all, for these representations are numerical in a variety of respects. First, they require that the infant have criteria for individuation and numerical identity (the ability to distinguish one entity seen on different occasions from two numerically distinct entities). Second, comparisons of the representations of the test event and the outcome arrays are based on one-to-one correspondence among individuals, an operation that establishes numerical equality or inequality among sets. Third, any representation that is equivalent to "There is one object; there is another object; and there are no other objects" is logically equivalent to "There are two objects." Fourth, the representations that underlie success on infant-addition/subtraction studies are spontaneous; they require no habituation and they require no training. Thus, they may well be the first that are available as a basis for making sense of the linguistic expression of number. However, they fall short of fully symbolic representations of number, as there is no unique symbol for each integer, and because there is no counting process involved in building the representations of the arrays.

The developmental changes in infants' model building capacities between ages 5 and 10 months may have several sources. General information processing capacities may increase (e.g., see Diamond, 1991, for a review). Alternatively, infants' knowledge of objects per se may become more elaborated, allowing them to take into account more information about objects and their locations in constructing mental models of the arrays. Most probably, both types of changes contribute to the developmental progression witnessed in these studies.

References

Antell, S., & Keating, D. (1983). Perception of numerical invariance in neonates. *Child Development, 54*, 695–701.

Baddeley, A. (1986). *Working memory.* Oxford, UK: Oxford University Press.

Baillargeon, R. (1995). Physical reasoning in infancy. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 181–204). Cambridge, MA: MIT Press.

Baillargeon, R., Miller, K., & Constantino, J. (1994). *Ten-month-old infants' intuitions about addition.* Unpublished manuscript, University of Illinois at Urbana Champaign.

Capaldi, E. J., & Miller, D. J. (1988). Counting in rats: Its functional significance and the independent cognitive processes which comprise it. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 3–17.

Carey, S. (1996). Continuity and discontinuity in cognitive development. In D. Osherson

& E. E. Smith (Eds.), *An invitation to cognitive science: Vol. 3. Thinking* (2nd ed., pp. 147–172). Cambridge, MA: MIT Press.

Cooper, R. G. (1984). Early number development: Discovering number space with addition and subtraction. In C. Sophian (Ed.), *The origins of cognitive skill* (pp. 157–192). Hillsdale, NJ: Erlbaum.

Dehaene, S. (1997). *La Bosse des Maths.* Paris: Editions Odile Jacob.

Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 67–110). Hillsdale, NJ: Erlbaum.

Fuson, K. C. (1988). *Children's counting and concepts of number.* New York: Springer Verlag.

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

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

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

Hauser, M., McNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences, 93,* 1514–1516.

Honig, W. K. & Stewart, K. E. (1989). Discrimination of relative numerosity by pigeons. *Animal Learning and Behavior, 17*(2), 134–146.

Huttenlocher, J., Jordan, N., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General, 123,* 284–296.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

Koechlin, E., Dehaene, S., & Mehler, J. (1997). Numerical transformations in five-month-old infants. *Mathematical Cognition, 3*(2), 89–104.

Meck, W., & Church, R. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 320–334.

Rilling, M., & McDiarmid, C. (1965). Signal detection in fixed ratio schedules. *Science, 148,* 526–527.

Siegler, R., & Robinson, M. (1982). The development of numerical understanding. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 6, pp. 227–287). New York: Academic Press.

Simon, A. (1997). Reconceptualizing the origins number knowledge: A non-numerical account. *Cognitive Development, 12,* 349–372.

Simon, A., Hespos, S., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development, 10,* 253–269.

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.

Spelke, E., & Van de Walle, G. (1995). Perceiving and reasoning about objects: insights from infants. In N. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial representation* (pp. 297–330). Cambridge, MA: Blackwell.

Starkey, P. (1992). The early development of numerical reasoning. *Cognition, 43,* 93–126.

Starkey, P., & Cooper, R. (1980). Perception of numbers by human infants. *Science, 210,* 1033–1035.

Starkey, P., Spelke, E., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36,* 97–127.

Strauss, M., & Curtis, L. (1981). Infant perception of numerosity. *Child Development, 52,* 1146–1152.

Tan, L., & Bryant, P. (1996). *Discrimination of simultaneously presented large-number arrays by 6-month-old infants.* Poster session presented at the 10th biennial meeting of the International Society of Infant Studies, Providence, RI, April.

Treiber, F., & Wilcox, S. (1984). Discrimination of number by infants. *Infant Behavior and Development, 7,* 93–100.

Trick, L., & Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101,* 80–102.

Uller, C. (1995). *On infant number representations.* Poster session presented at the biennial meeting of the Society for the Research on Child Development, Indianapolis, IN, April.

Uller, C. (1996). *Origins of numerical concepts: A comparative study of human infants and nonhuman primates.* Unpublished doctoral dissertation, MIT, Cambridge, MA.

Uller, C., Carey, S., & Hauser, M. (1998). *Evolutionary roots of human numerical capacities.* Manuscript in preparation.

Uller, C., Carey, S., & Huntley-Fenner, G. (1994). *The representations supporting infant addition.* Poster session presented at the 9th biennial meeting of the International Society of Infant Studies, Paris, France, March.

van Loosbroek, E., & Smitsman, A. (1990). Visual perception of numerosity in infancy. *Developmental Psychology, 26,* 916–922.

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

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

Wynn, K. (1995). Origins of numerical knowledge. *Mathematical Cognition, 1,* 36–60.

Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science, 7,* 164–169.

Xu, F., & Spelke, E. (1997). *Large number discrimination in 6-month-old infants.* Poster session presented at the annual meeting of the Cognitive Science Society, August.