Developmental Psychology
2003, Vol. 39, No. 1, 61–70

0012-1649/03/$12.00 DOI: 10.1037/0012-1649.39.1.61

Looking and Search Measures of Object Knowledge in Preschool Children

Bruce Hood, Victoria Cole-Davies, and Melanie Dias
University of Bristol

The same preschoolers were tested on an observation task and a search task involving the invisible displacement of an object. In the observation task, children watched an object roll behind a screen from which protruded the top of a solid wall. Analyses revealed significantly longer looking to impossible than to possible outcomes in all children. In search, the child was allowed to retrieve the rolled object. Most 3-year-olds but significantly fewer 2.5-year-olds completed the search successfully. An unexpected sex difference was found, with boys outperforming girls. Search performance was not associated with observation measures. The findings indicate that children visually discriminate violations of solidity but that this sensitivity is not associated with successful search performance.

According to Piaget (1954), mental representation is constructed in infancy through the process of interacting with objects in the physical world. Initially, the newborn has no representation of the physical world other than fleeting, uncoordinated perceptions. However, as the sensory and motor systems mature and combine with new experiences to enable more complex interactions with the world, the infant, in the view of Piaget, gradually builds representations of the physical world. For Piaget, this representation emerged as a consequence of acting on objects and experiencing the consequences of those actions, and it was not until approximately the end of the 1st year of life that infants began to construct representations of unseen objects as independent entities. This account was illustrated and supported by Piaget's famous search experiments, in which he demonstrated that 6-month-olds behaved initially as if hidden objects ceased to exist, reflecting the absence of representation. When the infants did search for hidden objects a few months later, they returned to locations where they had previously found these objects because they did not represent them independently at the new locations.

This account was challenged in the 1980s with the introduction of a new paradigm for studying infant representation. Renée Baillargeon and Elizabeth Spelke argued that Piagetian search was simply too difficult for infants (Baillargeon, Spelke, & Wasserman, 1985). Rather than reflecting limited *competence* to form representations of objects, the difficulty of Piaget's tasks illustrated *performance* limitations because of the need to coordinate actions and plan behavior. In other words, infants either did not search or did not search accurately because the task demands had overwhelmed their ability to execute the retrieval successfully.

The performance limitation account was supported by neurophysiological research which revealed that planning and coordination of actions required brain structures that were not fully mature in the young infant. In particular, working memory and inhibition were identified as limiting factors in executing successful search, and these functions had been linked to the dorsolateral prefrontal cortex, a region that was not fully operational in the infants who failed search tasks. This was supported by comparative primate studies which showed that both cortically immature young monkeys and mature adult monkeys with lesions to the dorsolateral prefrontal cortex exhibited search behaviors similar to those of human infants who did not successfully complete searches (Diamond & Goldman-Rakic, 1989). Likewise, electroencephalogram studies showed that successful performance on Piagetian search tasks was related to increased activity in the frontal cortex of human infants (Bell & Fox, 1992). Anatomical (Huttenlocher & Dabholkar, 1997) and imaging (Chugani & Phelps, 1986) studies also supported the idea that there were significant changes in frontal structures occurring at the same time as major transitions in search performance.

Because of these performance limitations, Spelke and Baillargeon argued that a more accurate evaluation of the infant's representational capacity could be obtained by using measures that did not require the same levels of coordinated behavior necessary for search (Baillargeon, Spelke, & Wasserman, 1985). Rather than requiring infants to act, they simply observed infants' responses to events that were either consistent with or not consistent with physical laws. The "violation of expectancy" paradigm, as it became known, was based on the principle of the conjurer's trick; namely that a trick triggered an increase in the observer's attention because it contravened an expectancy or belief about the physical world. For example, in a direct adaptation of one of Piaget's search tasks, Baillargeon and Graber (1988) demonstrated that infants who observed an object placed behind an occluder at Location A looked significantly longer at the display if the object was retrieved from behind a different occluder at Location B. The interpretation was that the infants' expectation that the object continued to exist at Location A had been violated.

On the basis of this simple logic, a large number of studies involving the violation-of-expectancy paradigm were conducted, and these studies revealed infant discriminations on an extensive range of physical attributes and properties. Not only did infants

Bruce Hood, Victoria Cole-Davies, and Melanie Dias, Department of Experimental Psychology, University of Bristol, Bristol, England.

This research was supported by Economic and Social Research Council of Great Britain Grant R000223293 to Bruce Hood. We would like to thank the Avon Health Authority and the families who made this research possible.

Correspondence concerning this article should be addressed to Bruce Hood, Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, England. E-mail: bruce.hood@bristol.ac.uk

understand that objects continued to exist, but they made judgments about the nature of the objects' constitution and interactions with other objects. It was argued that in the same way that adults reasoned about the physical world, infants' looking time behavior revealed that they appreciated that objects continued to exist, retained their spatial and physical properties, and were subject to physical laws when out of sight.

Violation of expectancy became the single most influential technique introduced in recent years in developmental cognitive research. Backed by this paradigm, new accounts of infants' understanding of the physical world appeared. Spelke (1991) and Baillargeon (1995) proposed similar accounts detailing an early appreciation of the physical world that was built in at birth and did not need to be learned from experience. Spelke (1991) took a nativist stance and argued that certain core principles remained constant throughout development, whereas Baillargeon (1995) argued that infants possessed an early appreciation of physical knowledge in an all-or-none manner and needed experience to fine-tune aspects of this knowledge. For example, Baillargeon (1995) showed that whereas 3-month-olds looked longer as a box was pushed off a cliff but remained suspended, 4.5-month-olds recognized that the box should still fall if most of it was balanced over the edge.

The violation-of-expectancy paradigm also generated considerable controversy and criticism. Criticisms fell into a variety of categories, including methodological, statistical, procedural, and interpretational (Haith, 1998). In the current study, we set out to address interpretations of the infant findings. We were motivated by recent studies showing that older children who should have been well beyond the period when there are performance constraints still exhibited limited appreciation of physical knowledge in tasks that were direct adaptations of the infant looking time studies (Berthier, DeBlois, Poirer, Novak, & Clifton, 2000; Hood, 1994; Hood, Carey, & Prasada, 2000). For example, Spelke (1991) claimed that there were immutable core principles of knowledge that were built in at birth and remained relatively unchanged throughout the course of development. She identified the three core principles of contact, cohesion, and continuity. In particular, the principle of continuity specified that a moving object traces exactly one connected path over space and time. Objects do not jump over space and time, ceasing to exist in the interim, nor can two solid objects occupy the same point in space and time.

As shown by looking time studies, the continuity principle is appreciated by infants as young as 4 months (Spelke, Breinlinger, Macomber, & Jacobson, 1992). Spelke et al. (1992) showed that infants who observed a ball roll behind a screen above which a solid wall protruded and then saw the same ball emerge from beyond the other side of the wall looked significantly longer than infants who observed the same sequence but without a wall in place. Similarly, 4-month-olds familiarized to a ball landing on a stage after an occluder had been removed showed a significant increase in looking time when the same outcome was revealed in spite of the insertion of a solid shelf (Spelke et al., 1992). Control studies with no falling event revealed that the position of the ball under the shelf was not intrinsically more interesting than the outcome of the ball resting on top of the shelf. Both the wall and the shelf studies provided converging evidence that 4-month-olds appreciate that a solid ball should not pass through a solid barrier, thereby violating the core principle of continuity.

Hood and colleagues (Hood, 1994; Hood et al., 2000) adapted Spelke's infant looking time studies to search tasks for preschoolers. The familiarization phase was identical to the infant looking time experiments. Children watched as a ball was rolled or dropped behind an occluder that was then removed to reveal the final resting location. However, on the test trial after a wall or shelf was inserted, children were asked to search for the ball instead of removing the occluder. Search was random at 2 years of age, indicating that the children did not know the location of the ball. When there was no insertion of a barrier, children could find the ball accurately, indicating that the limitation was not due to an inability to search. However, because these studies involved two hiding places and a period of familiarization to one location, children could be biased to the incorrect location during familiarization. In a much more thorough paradigm with four hiding locations, which enabled an analysis of search patterns, Berthier and colleagues (2000) replicated the finding of search errors among preschoolers and demonstrated that although the children took into account the position of a solid wall, they were still unable to search accurately on the first attempt. Taken together, these findings strongly suggest that core principles demonstrated in infancy do not constrain the search behavior of older children.

These studies appeared to indicate that there was a difference between the outcomes of experiments based on looking time in infants and search behavior in preschool children. Indeed, some commentators described this as a dissociation (Munakata, 2001). However, without a direct comparison of the two methodologies within the same age group, it is difficult to determine whether there was, in fact, a dissociation. Either the difference in methodologies or the difference in age could have accounted for the pattern. The current study set out to address whether the two measures of physical knowledge dissociated within the same individual preschool child. Using Berthier and colleagues' (2000) apparatus, we sought to replicate the search behavior they found in an age group in the transition period between failing and passing the four-door search task.

We also tested our participants on a looking time or observation task using the same apparatus to demonstrate possible and impossible outcomes based on the relative position of the wall and the resting location of the rolled object. Analysis could then determine whether the two measures of looking and search dissociated. Unlike the case with standard infant looking time paradigms, we assessed five dependent measures: duration of first look, total duration of looking over 10 s, "double takes," vocalizations, and social referencing. First look and total duration have commonly been reported in infant looking time studies. The double takes analysis was based on the number of times the children refixated between the two open doors. This measure was used to determine whether children anticipated the appearance of the toy at a particular door as well as the discrepancy between where they saw the object and the alternative location. Vocalizations and social referencing were included as potential measures of discrimination. Vocalizations may result in clear, explicit appreciation of the violation. Social referencing, defined as looking toward the caregiver or commentator, is found in ambiguous situations in which children seek out information from adults to interpret events (Klinnert, 1985).

Method

Participants

One hundred forty-five children 2.5 years ($n = 83$) and 3 years ($n = 62$) of age were recruited by means of local health authority birth records. Failure to complete the observation task through noncompliance or difficulty in obtaining clean data was the major reason for exclusion from the analysis. The final analysis included 41 children in the 2.5-year age group (19 boys and 22 girls) and 44 children in the 3-year age group (22 boys and 22 girls). Mean ages were 30.8 months (range: 29.9 to 31.7 months) in the 2.5-year group and 36.8 months (range: 35.5 to 38.6 months) in the 3-year group.

Design

Each child was assessed on an observation task and a search task. Both tasks were preceded by the same familiarization phase. For the observation task, a test trial included a possible and an impossible outcome, with the order counterbalanced across children. The search task involved a minimum of four trials and a maximum of eight trials. The order of the two types of tasks was also counterbalanced across participants to avoid the possibility of order effects. This accounts for the high participant dropout rate; children who had been actively encouraged to approach the apparatus on the search task were often unable to sit still on the observation task.

Apparatus

The apparatus is shown in Figure 1. The apparatus was presented on a metal frame with a wooden base, 58 cm from the floor, and displayed as a "stage." Black felt curtains hung above, below, behind, and on either side of the apparatus to hide the experimenter who carried out the tasks. The apparatus consisted of a wooden box with four doors and a ramp down which a soft red "Mr. Strong" toy inside a small tin was rolled, behind the closed doors. A wall could be placed in one of four locations (A, B, C, or D in Figure 1) between the doors to stop the tin from rolling to the end of the track. The wooden box was 80 cm wide and 32 cm high and had a depth of 21 cm. Both ends and the top of the box were left open, and a large hole was cut out of the back of the box and covered with a black felt curtain, enabling the experimenter to push his or her hands surreptitiously through the back of the box to reposition the tin during the looking tasks.

The four doors in the front of the box were positioned 9 cm up from the base on which the apparatus lay. The doors were 17 cm high, 14 cm wide, and 4 cm apart. They were constructed out of thin plywood, hinged on the bottom and with a wooden knob on the top, enabling the child to open the door with an effortless downward pull. The ramp was 100 cm long, with two white dowels glued 10 cm apart to the top surface to create a track that constrained the path of the tin. The tin could be rolled down the ramp from the child's right to the child's left. A solid wooden wall was placed in one of four positions between the doors so that the tin came to rest to the right of the wall and was easily retrievable through the open door. The 3-cm-thick wall (height: 30 cm; width: 21 cm), once inserted into position, stood 8 cm higher than the top of the wooden box.

Procedure

The experimental session started with the child seated on a small infant chair positioned in the middle of the room in front of the apparatus, which was set approximately 2 m back. The child's caregiver was seated to the left-hand side behind the infant chair, and the experimenter–commentator was seated to the right behind the infant chair. A running commentary was used, because pilot studies showed that many caregivers spontaneously offered commentary to their children while the event sequences were taking place and that they often directed the children. We found that if one of the experimenters provided the same commentary to every participant, caregivers did not speak during testing. Before both the observation task and the search task, the child was shown a familiarization sequence. All behaviors were recorded on two digital video cameras for later analysis. One camera was situated in front and to the side of the child and focused on his or her face; the other recorded search behavior from behind and to the side of the child.

Familiarization. The familiarization phase enabled the child to become accustomed to the apparatus and the methodology before the test phase. The experimenter demonstrated that all four doors could be opened, left them lying open, and then held up the "Mr. Strong" toy by the apparatus while it was introduced by the commentator. The experimenter then placed the toy into the tin, positioned it on the peak of the ramp, and released it while the commentator described what was happening. The tin rolled past all of the open doors and out the other side of the apparatus. Next, the experimenter showed the solid wooden wall to the child, and again the commentator described the actions as they happened. The experimenter placed the wall in Position B (see Figure 1), tapped it three times to draw attention to it, and then rolled the tin down as before. When the tin stopped at Door 2, the commentator animatedly drew attention to the fact. The experimenter then moved the wall to Position D, and the previous dialogue and actions were repeated so that the tin stopped at Door 4. The experimenter then removed the wall and closed all of the doors while the commentator talked the children through what was happening.

Observation task. In the first test phase, the experimenter placed the wall in Position C and tapped it three times to draw attention to it. The

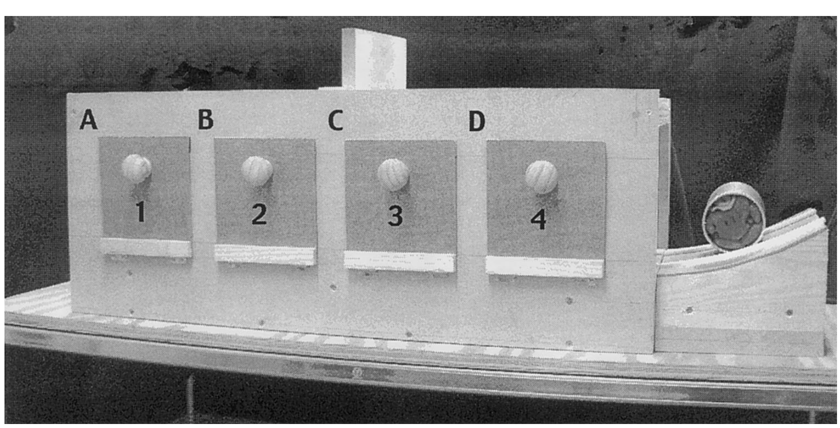

Figure 1. Picture of the apparatus, with wall positions and door codes superimposed on the image (see text).

commentator drew the child's attention to the apparatus and wall and made it clear that the child only had to watch this part of the task. The experimenter then rolled the tin, and, once it had stopped, the two doors on either side of the wall were opened simultaneously so that the child could observe. There were only two test conditions in this task. In the *possible condition*, the tin was located in front of the wall. After a period of 10 s, the tin was lifted out, the toy was shown to the child, and the doors were closed once more. In the *impossible condition*, the experimenter tapped the wall three times and, unbeknownst to the child, placed an identical tin with a second Mr. Strong toy on the other side of the wall through the back of the apparatus (Door 2). The visible tin was then rolled down the ramp. When the tin stopped, the experimenter simultaneously took that tin out from behind Door 3 and went to open Doors 2 and 3 to show the toy (now on the other side of the wall). The order of the event presentation was counterbalanced across participants.

Search task. As in the observation task, the familiarization phase was run through once more to show how the tin rolled out the other side of the apparatus in the absence of the wall or was stopped by the wall when in place. All of the doors were then closed once more, and the wall was placed in one of four door locations. The experimenter tapped on the wall three times, and the commentator directed the child's attention to the wall. The experimenter then released the tin with the Mr. Strong character inside, from the top of the ramp. Next, the commentator encouraged the child to approach the apparatus and retrieve the toy by pulling down the handle of the door of her or his choice. During each trial, the child was allowed unlimited time and attempts to retrieve the toy. If the child retrieved the toy within the first two attempts, the wall was removed and the remaining open doors were closed. If the child had difficulty finding the toy, the experimenter opened the correct door (if applicable), and the commentator explained that the wall stopped the tin. The wall was then removed and the doors closed. Testing consisted of two blocks of four trials in which the locations of the wall varied in random order. The same location was never repeated over two consecutive trials.

Differences between observation and search procedures. There were significant differences between the observation and search task procedures that require further consideration. The major difference was that there were only two trials in the observation task and the wall did not move, whereas there were multiple trials in the search task and the wall moved from trial to trial. The reason for this discrepancy was that a comparison between possible and impossible outcomes on the observation task required all aspects of the sequence to be identical other than the order of outcomes. Otherwise, any difference in looking time could reflect a change in the sequence. The wall was moved from trial to trial in the search task to replicate the original Berthier et al. (2000) procedure as well as to avoid the likelihood that children would find it too easy to open the same door on repeated trials without reasoning about the wall. These differences are not trivial and have implications for the analyses and interpretations (as described later).

Results

Half of the children began with the observation task followed by the search task, and the other half completed the tasks in the reverse order. There was a major effect of task order on completion in that children who were tested with the search first were less likely to produce data for the observation task. For example, of the children who failed to produce any observation data on a single trial ($n = 31$), 71% had experienced the search task first.

The criterion for inclusion in the study was at least four clean attempts on the search task and clean data on the observation trials. Clean search attempts were defined as child-initiated approaches to the apparatus without the aid or prompting of the supporting adult(s). Clean observation data were defined as trials in which the child was attending to the apparatus when the doors were opened. Each tape was scored by two observers; interrater reliability was 94% for the duration measures and 96% for the direction of gaze.

Observation Task

In the looking task, five dependent measures were evaluated during the 10-s period following the child's exposure to either the possible or impossible outcome: (a) duration of first look, (b) total duration of looking, (c) number of times the child looked back and forth between the possible and impossible locations, (d) number of times the child commented during the outcome, and (e) number of times the child engaged in social referencing with the caregiver. The data were subjected to a repeated measures analysis of variance in which the measures taken on the possible versus impossible trials were the dependent variables. This measure of difference between conditions was deemed the "magic" effect to reflect the extent of the child's discrimination between possible and impossible outcomes. The first analysis was an assessment of order variables related to the design of the experiment that might influence the magic effect for each measure: (a) order of task presentation and (b) order of possible and impossible trials. Analysis of variance revealed no significant effect of either variable on any of the five dependent measures.

The main independent variables chosen were age group (2.5 years vs. 3 years) and sex. Age was chosen because previous search studies involving this apparatus (Berthier et al., 2000), as well as the current search, indicated that age was a significant factor in performance. Likewise, the current analysis of search also revealed a significant sex effect (described subsequently), and so this variable was included in the analysis of observation data. Of the five dependent measures, the only two that differentiated between the possible and impossible trials were duration of first look and total duration. Because these two measures were only moderately correlated ($r = .62$, $p < .001$), separate analyses are reported. Although the other dependent measures did not show discrimination between the possible and impossible outcomes, there were significant age effects, with significantly fewer vocalizations, $F(1, 81) = 9.51$, $p < .01$, and double takes, $F(1, 81) = 13.46$, $p < .001$, among the 3-year-olds than among the 2.5-year-olds.

Duration of First Look

This measure was defined as the duration of the child's first fixation on the apparatus after the doors had been opened by the experimenter; it was terminated by the first look away to either the other door or elsewhere. The data are shown in Figure 2. As a group, children exhibited significantly longer first looks on the impossible than on the possible trials, $F(1, 81) = 5.32$, $p < .05$. The effect size was .27. There was also a marginal effect for age, with the younger group tending to look longer than 3-year-olds on first looks, $F(1, 81) = 3.03$, $p = .08$. Sex was not a significant factor, and there were no significant interactions.

Total Duration

This measure was defined as the sum duration of fixations to each door opened by the experimenter within the 10-s exposure

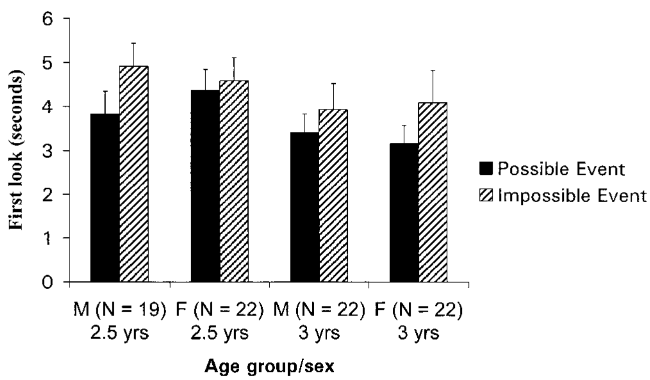

Figure 2. Mean durations of first looks to possible and impossible outcomes, by age and sex. M = male; F = female.

period. Again, a repeated measures analysis of variance was conducted; the dependent variable was the difference between possible and impossible outcomes, and the independent variables were age and sex. The data are shown in Figure 3. In terms of total duration of looking, there was a significant "magic" effect, with all children looking significantly longer at the impossible outcome, $F(1, 81) = 4.85$, $p < .05$. The effect size was .26. There were no other significant main effects or interactions. Twelve children (14%) looked continuously for the 10-s period during an observation trial. Of these 12 children, all but 1 did so during an impossible trial. Three of these children were 2.5-year-olds, and the remainder were 3-year-olds.

Search Task

A minimum of four and a maximum of eight search trials were required for children to be included in analyses. Of the 85 children included (41 in the 2.5-year group and 44 in the 3-year group), 66 completed eight trials, 4 completed seven trials, 3 completed six trials, and 12 completed four trials. Two types of analyses were conducted. The first was a parametric analysis based on percentage of correct searches. The second was a nonparametric analysis of participants characterized as passing or failing the search task according to a binomial statistic.

Percentage of Correct Searches

The number of correct searches on the first attempt divided by the number of trials was subjected to an analysis of variance in which order of task, age, and sex were the independent variables. Again, there was no effect of order of task presentation, but both age, $F(1, 81) = 25.80$, $p < .001$, and sex, $F(1, 81) = 11.10$, $p < .001$, were significant; there were no interactions. The results are shown in Figure 4.

Successful Versus Nonsuccessful Search

With a four-choice decision, the number of correct searches necessary for significance at the .05 level was as follows: three or four correct for four trials, four or more correct for six trials, five or more correct for seven trials, and five or more correct for eight trials. According to this criterion, 20% of 2.5-year-olds ($n = 8$) and 59% of 3-year-olds ($n = 26$) passed the search task. A chi-square analysis showed that this distribution was significant, $\chi^2(1, N = 85) = 14.35$, $p < .001$. A breakdown by sex revealed further effects, with only boys passing ($n = 8$; 42%) among the 2.5-year-olds and significantly more boys ($n = 17$; 77%) than girls ($n = 9$; 41%) passing at 3 years of age (see Figure 5). Again, this distribution was significant: $\chi^2(1, N = 41) = 11.51$, $p < .001$ and $\chi^2(1, N = 44) = 6.02$, $p < .05$ for 2.5-year-olds and 3-year-olds, respectively. There was no significant effect of task order.

Considering the high attrition rate, one possible explanation for the sex difference is that the children differed in terms of their selection for inclusion in the study. However, an analysis of children dropped from the study revealed that there was no significant difference between the percentage of boys ($n = 35$, 58%) and the percentage of girls ($n = 25$, 42%) excluded. When the analysis was broken down by age, male-to-female percentages were equal in the 2.5-year group (male: $n = 24$, 59%; female: $n = 17$, 41%) and the 3-year group (male: $n = 11$, 58%; female: $n = 8$, 42%). Therefore, there was no evidence that sex differences were attributable to different attrition rates among boys and girls at the different ages.

Analysis of Errors

An examination of the errors addressed by Berthier et al. (2000) was undertaken to investigate whether certain strategies were exercised by the children during the search task. Possible strategies

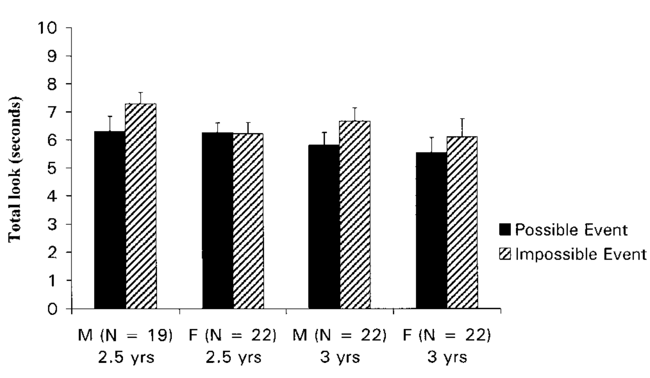

Figure 3. Mean durations of total looking over 10 s to possible and impossible outcomes, by age and sex. M = male; F = female.

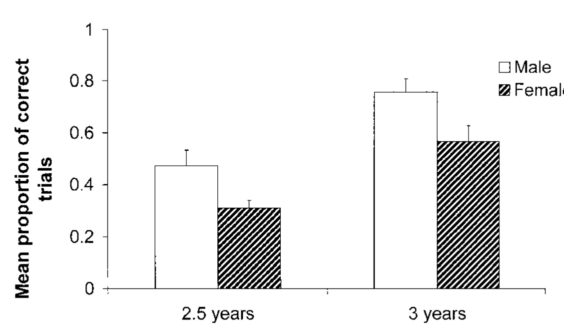

Figure 4. Mean number of correct searches on first attempt, divided by number of search trials, by age and sex.

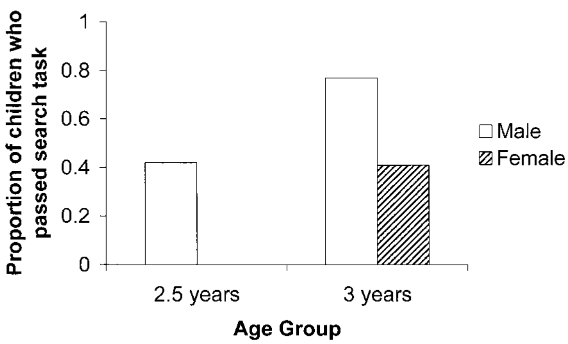

Figure 5. Proportion of children in each age group who passed the search task according to the binomial criterion (see text for details). Note that no girls passed this criterion at 2.5 years of age.

included (a) a preference for a particular position or door; (b) perseverative errors, defined as opening the door that was correct on the immediately preceding trial; and (c) choosing a door adjacent to the wall, either to the left (which would be correct) or to the right (incorrect). Only data from the 2.5-year-old group were analyzed in this way, because most 3-year-olds passed the search task.

The proportion of trials in which each door was opened was computed for each child. The favorite door was defined as the door with the highest proportion. Door 3 (third from the left end of the apparatus) was identified as the favorite door for 2.5-year-olds (mean proportion: .36); it was chosen 98 times out of a total of 274 trials. Twenty children chose this door on their first reach in at least 50% of trials completed. Door 1 was the least favored position; only 4 children selected it repeatedly over trials. Preferences to search at Door 3 presented a problem in terms of interpreting the magic effect reported in the observation task. If there was an inherent bias to expect an object at this location on the search task, then significant looking to Door 2 on an observation task could simply reflect a novelty preference.

The analysis of the magic effect was repeated with the 20 children who showed a preference for Door 3 removed. Again, there was a significant discrimination between possible and impossible outcomes in terms of both duration of first look, $F(1, 61) = 3.99$, $p < .05$, and duration of total looking, $F(1, 61) = 3.97$, $p < .05$. The corresponding effect sizes were .26 and .28. We also checked to see whether the Door 3 preference tended to occur among children who witnessed the magic trick first relative to children who underwent the search task first. There was no difference between these groups, indicating that this preference was not due to children trying to re-create the magic trick in the search task after witnessing the observation task.

The second possible strategy investigated was perseveration in choosing the door behind which the ball was found on the preceding trial (not including familiarization). In the 2.5-year-old group, perseverative errors occurred, on average, on 24.5% of trials (67 of 274), no more than would be expected by chance (25%). However, 4 of the 41 children did perseverate on at least 50% of trials.

Finally, the strategy of selecting a door adjacent to the wall (left or right) was investigated. If the children were not attending to the wall position, they would have chosen an adjacent door 43.75% of the time. This is because there was a 50% chance of choosing an adjacent door for the three wall positions (B, C, and D) and a 25% chance of choosing an adjacent door when the wall was in Position A. The 2.5-year-olds, on average, chose an adjacent door 65.7% of the time, a value significantly different from 43.75%, $t(40) = 6.9$, $p < .01$. When the data from the 8 children in the 2.5-year group who successfully completed the search task were omitted, it was found that an adjacent wall was selected 60.5% of the time, an average value that was still significantly different from 43.75%, $t(33) = 5.5$, $p < .01$.

Through further examination of the children's second reaches, it can be shown whether the position of the wall did indeed influence children's door selection. If choosing an adjacent door had been a strategy used, it would be expected that if the toy were not found behind the door adjacent to the wall on the first reach, the door on the other side of the wall would be selected on the second reach. Among the 2.5-year-olds who did not solve the search task, 1 always chose an adjacent door on second reaches, and 2 others chose an adjacent door on 75% of trials if they had chosen an adjacent door on the first reach that did not reveal the toy. The majority of 2.5-year-olds (71%) chose an adjacent door on their second reach on 25% or less of trials. On average, children in this age group chose an adjacent door on the second reach 20% of the time when they had chosen an adjacent door on the first reach, which was significantly less than would be expected by chance (20% vs. 33%), $t(33) = 2.70$, $p < .05$. This result did not support the suggestion that children followed a strategy of attending to the wall and opening doors adjacent to it to retrieve the toy.

However, we found that, in general, children who did not pass the search task did choose a door adjacent to the one previously selected 95% of the time if they had not chosen the correct door on their first reach. This rate was significantly higher than the chance level of 33%, $t(33) = 29.9$, $p < .01$. Of the 34 children, 28 always chose a door one removed (in either direction) from their previous door selection on their second reach. In fact, all of the 2.5-year-olds who did not pass the search task chose a door adjacent to the previous choice at least 50% of the time when their first choice had been incorrect. This result strongly suggests that the search strategy of 2.5-year-olds was to methodically work through the doors in turn to find the toy, regardless of the wall position. As indicated by the door preference analysis, the 2.5-year-olds tended to begin their search by selecting a middle door (i.e., Door 3), after which they worked outward, opening an adjacent door in either direction rather than randomly selecting a door in any location.

Comparison of Observation and Search Task Performance

We undertook a correlation analysis to investigate whether there was any relationship between measures of looking time discrimination on the observation task and search performance. We entered the five dependent measures into nonparametric correlation with search performance. In addition, we calculated two new looking time variables to reflect the magnitude of the reported discrimination between possible and impossible outcomes. These variables represented the differences in values from impossible and possible trials, that is, (a) differences in first looks and (b) differences in total looking. There was no evidence that these or the other dependent measures were correlated with search performance, as defined by either the cutoff criterion for passing the search task or the proportion of correct searches on the first attempt.

As noted earlier, there were important differences between the procedures for the observation and search tasks that could have implications for a direct comparison between the two tasks. In particular, the wall moved between every trial in the search task but remained constant during the observation task. In this sense, the observation task may have been easier. This issue was addressed by repeating the preceding analysis but considering only the data for the first trials in the observation and search tasks. In the observation task, an analysis of variance was conducted on the magic effect by comparing the children who witnessed the possible trial first with the children who witnessed the impossible trial first. Unlike the within-subject analysis, this group comparison revealed no significant differences between the possible and impossible outcomes on any of the five dependent measures.

Performance on the initial search trial was also considered. On their initial attempt, 37% ($n = 15$) of 2.5-year-olds and 61% ($n = 27$) of 3-year-olds chose the correct door. Again, a chi-square analysis revealed a significant age effect, $\chi^2(1, N = 85) = 3.96$, $p < .05$. Whereas the performance of 3-year-olds remained the same on the first trial and all other trials (61% vs. 59%), 2.5-year-olds appeared to perform better on the first trial than over all trials (37% vs. 20%). However, almost all of these successful children (93%) directed their search to either Door 2 or Door 4, the locations that had just been demonstrated during the familiarization.

Note that the children who failed the search at Doors 1 and 3 must also have directed their search to these two doors. These children simply repeated the search demonstrated by the experimenter. The 3-year-olds who were successful exhibited equally good performances on all baited doors. This finding supports the general concern that perseveration may represent a particular problem for 2.5-year-old children engaged in search tasks (Hood et al., 2000).

The correlation analysis between the dependent measures on the observation task and performance on the search task was conducted independently for the two groups of children who had witnessed either the possible sequence first or the impossible sequence first. Again, no significant associations were found.

Discussion

Within-subject analyses revealed that there were significant differences between possible and impossible outcomes on two measures of looking time duration for all children. This effect was independent of which task or condition children experienced first. However, when only the first observation trial was considered, group comparisons revealed no significant differences between possible and impossible outcomes. This suggests that the two events were only discriminated in our study when both outcomes were observed over the first two trials. Because looking time studies typically involve within-subject designs, the remainder of the discussion is restricted to the evidence for positive discrimination.

The absence of order effects indicated that witnessing a magic trick first did not influence search, and conversely, interacting through search first did not affect looking behavior. On the first search trial, 3-year-olds produced the same performance as that on multiple trials and again were better than the younger children. The 2.5-year-olds showed apparently better performance when only the first trial was considered, but this was due to their repeating search at the same locations as those just witnessed during the familiarization. The most unexpected finding was the sex effect on search.

In the observation task, both age groups looked significantly longer at the impossible outcome of seeing the object at Door 2 behind the wall in comparison with seeing the object at Door 3 in front of the wall. Note that this occurred even though children had not witnessed the object at Door 3 but had observed the object at Door 2 during the familiarization phase. This discrimination could be attributable to at least two candidate explanations that need to be considered. First, the preference could reflect that the children recognized a violation of solidity in that the object should not have been located on the far side of the wall. This interpretation is consistent with the position of advocates of early physical knowledge of solidity (Spelke et al., 1992).

Alternatively, children may have responded to the novelty of the object at this position. Search analysis of all trials revealed that when errors were made, there was an overall bias for Door 3. This bias could have contributed to a novel preference for the object at Door 2 in the observation task. However, the preference for the impossible outcome remained when the children who had this door bias were removed and the analysis was repeated.

The preference could also have arisen because of a bias introduced during familiarization. In the familiarization phase, children always saw the object in front of the wall. Any preference for the impossible outcome could have reflected the novelty of the object at the position behind the wall without an appreciation of solidity constraints. This interpretation is consistent with the position of advocates of simple perceptual accounts (Bogartz, Shinskey, & Speaker, 1997). In planning the current study, we were aware of this potential problem. A stronger control would have included additional trials in which the object was rolled in both directions so that it could be observed in both relative positions. Alternatively, the familiarization phase could have involved trials in which the object was simply placed in front of and behind the wall. However, although such controls are necessary for an analysis of whether children are sensitive to the effect of walls on rolling objects, and have already been addressed in infant studies (Spelke et al., 1992), the inclusion of such controls in the current study would have been problematic for a number of reasons.

First, control trials could not be easily included with test trials in a within-subject design. Second, this would have meant increasing the number of trials or return visits in a paradigm that was already difficult for preschoolers. Finally, differences (or a lack of differences) would have been difficult to interpret when the children had to remember which event they had just witnessed (“Was it rolled or placed, and if so, from what direction?”). Rather, in the current study we set out to determine whether there was discrimination on a looking time measure that was related to search performance in the same individual. Conceding that the necessary controls were missing from our study, we deferred to the infant studies that had ruled out novelty of relative position as an explanation. Instead, we interpreted the increased looking time in our studies to reflect sensitivity of the constraint imposed by the wall.

Another methodological issue that became apparent in the review process was that the search task may have been more difficult than the observation measure because the wall was always moving.

Although the first trial analysis of search did not support this notion, future observation studies should consider including outcomes in which the wall is moved between trials.

With these issues in mind, the search performance observed here was still very similar to that reported in the original study involving this type of paradigm and apparatus (Berthier et al., 2000). Again, older children searched more successfully. However, unlike the original Berthier et al. study, there was also a strong sex effect, with boys exhibiting significantly better search performance at both ages. This finding was consistent with the results of other studies of invisible displacement search in preschoolers which found better search performance in boys (Hood, 1995). Although Berthier and colleagues did not report sex differences in their studies, their sample size was considerably smaller. It is not clear why boys and girls should differ on this particular search task, though we note that the present findings are consistent with those of other studies (Halpern, 1992).

Conclusion

In comparison with a possible outcome, children looked significantly longer at the impossible outcome of an event sequence that violated the laws of solidity, suggesting that they were sensitive to the apparent violation. However, on a search measure, 3-year-olds (but significantly fewer 2.5-year-olds) used the same sort of information to successfully retrieve that hidden toy, as found in previous studies (Berthier et al., 2000). Our analysis of first search trials and patterns of search thereafter suggested that one of the major impediments to successful search among the 2.5-year-olds was the tendency to repeat search at locations baited during familiarization on the first attempt. After that, search patterns were characterized by children selecting an initial door and then continuing with adjacent doors until successful. The older children were more flexible in their responses.

Unexpectedly and still unexplained, sex was also found to be an important determinant of search performance, with boys outperforming girls at both ages. Most important was the lack of evidence that search performance was systematically related to any of the measures obtained during the observation task. Furthermore, only duration measures of observation, and not the other measures (double takes, vocalizations, and social referencing), gave any indication of discrimination of possible from impossible events. The apparent dissociation of duration measures from search performance gravely undermines the position of proponents of a rich interpretation of the looking time paradigm (see similar argument by Haith, 1998). On the other hand, the findings provide support for continuity models by demonstrating that looking time reveals sensitivity to violations of core physical principles that may underpin commonsense conceptions in adults (Spelke, 1991).

However, this conclusion must be moderated by the absence of significant discrimination when a group comparison was made of looking on the first trial alone. In our study, the differences emerged only when the responses to the two outcomes were compared. Most infant looking time studies, including those of Spelke et al. (1992), have involved the use of within-subject comparisons as a "statistically desirable approach" for determining changes in behavior (Kirk, 1968, p. 131). However, repeated measures analysis is not without fault. In particular, systematic changes could be related to order effects (this criticism does not apply to our study because of the randomized order used). We believe that our design and analysis support positive discrimination. Nevertheless, it may be a discrimination that occurs in preschoolers only when both outcomes are available.

Other studies have reported similar results. Mash, Clifton, and Berthier (2002) also found that 2-year-olds looked longer at impossible outcomes on the Berthier apparatus, although the effect size was better (.46) for first look. In the Mash et al. study, children observed a rolling sequence and saw a puppet open a door. The puppet opened the correct door and found the ball (possible outcome), the puppet opened an incorrect door and did not find the ball (possible outcome), or the experimenter secretly moved the ball around the barrier and the puppet did not find the ball despite having opened the correct door (impossible outcome). As in our study, 2-year-olds looked significantly longer at impossible outcomes than at possible outcomes. Although they were not also tested on search as in the current study, previous research indicates that they would have not been able to pass the search version (Berthier et al., 2000; Hood et al., 2000). Therefore, if search had been attempted, a similar dissociation between measures could have been expected.

Although, to our knowledge, there are no studies focusing on the two measures in the same preschool child, there are studies of object permanence comparing search and looking in the same infant (Ahmed & Ruffman, 1998). In the standard Piagetian A not B search task, infants who witness an object hidden at Location A on initial trials typically direct their search back to Location A after observing the object being hidden at Location B. In their experiments, Ahmed and Ruffman (1998) defined passing the search task as not returning to search at Location A on the reversal trial after a delay. Infants were deemed to have passed the looking task if they looked longer at the impossible event than at the possible event. Having established that their sample of 8- to 11-month-olds failed the standard A not B search task, Ahmed and Ruffman presented these infants with a looking time version in which the toy was retrieved from either possible or impossible locations. They found that the significant proportion of infants who failed the search task reversed to a significant proportion who passed the looking task, indicating a dissociation between measures. Although their analysis was statistically weaker because individual variations were not taken into consideration, their findings are consistent with the conclusion that the two measures can dissociate.

Dissociations between looking and search are also found in nonhuman primates. In a study of physical knowledge in free-ranging rhesus monkeys, Hauser (2001) reported that monkeys did not pass the search tasks involving invisible displacement of falling pieces of food. Experimenters tested animals with a portable version of the shelf apparatus used in human shelf studies (Hood et al., 2000). However, in a subsequent looking time study (Santos & Hauser, 2002), the same population of monkeys was presented with a looking time version in which they again saw the food being dropped but simply observed the outcome of the food at possible and impossible locations. This measure revealed longer looking times to the impossible outcomes.

Taken together, these findings from diverse laboratories, paradigms, and populations support the assertion that looking time reveals sensitivity to violations of the continuity principle but that sensitivity is not sufficient to support search. A number of explanations have been offered to explain this difference. One class of explanations centers around the role of working memory in search and the limitations related to maturation of neural mechanisms (Diamond & Goldman-Rakic, 1989). Berthier and colleagues (2000) argued that search on their task is more dependent on working memory than looking time and, as such, may be relatively poor owing to maturational delay. Working memory may play a role in many standard search tasks such as Piaget's A not B paradigm, but it seems unlikely that impaired working memory was the major impediment in the current paradigm. Children could solve the task by simply noting the position of the wall and directing their search accordingly. Indeed, the analysis of errors suggested that they did note the wall position but did not exercise the logical search on the second attempt.

Also, in a follow-up study conducted by Berthier and colleagues (Butler, Berthier, & Clifton, 2002), children still made search errors when all of the relevant information other than the final position was made available. In this study, they used a similar apparatus that was constructed of transparent Plexiglas. The only opaque components were the doors. Therefore, the children could see the wall and the movement of the object as it rolled behind the doors. In this situation, much of the work for working memory was already done in that the visible wall defined the position and the visible movement enabled the child to track the object. Although there was a modest improvement in search when this apparatus was used, this improvement occurred only among children who accurately tracked the visible movement of the object. In general, performance was still very poor. Such results are in contrast to looking time event sequences that clearly require working memory. For example, one of the most remarkable reports is that 4-month-olds looked longer at an outcome in which a rolling toy car was not impeded by a solid object (a Mickey Mouse toy) situated in its path behind an occluder (Baillargeon & DeVos, 1991). The infants had to represent not only the continued existence of the impeding object but also its relative position behind the occluder in the path of the rolling car. Because much more information was provided in the current set-up in terms of both invisible (Berthier et al., 2000) and visible (Butler et al., 2002) displacements, a working memory explanation seems inadequate.

Others have argued for a graded representation account in which search requires a more substantiated representation to support accurate retrieval, whereas looking time can operate with a weaker representation (Munakata, McClelland, Johnson, & Siegler, 1997). Again, graded representation does not seem to deal adequately with the problem that the major determinant of the event outcome—namely, the wall—is visible. Children know that the object continues to exist and that it is somewhere near the wall. However, they do not appear to appreciate that it must be located in front of the wall.

Another type of explanation draws the distinction between the need to generate a prediction in the search task and the need to discriminate the outcomes in the looking task (Hood et al., 2000; Meltzoff & Moore, 1998). Kyeong and Spelke (1999) showed that children failed to anticipate the correct motion of a ball launched from a cliff by predicting a straight-down path, but they looked longer at this trajectory than at the correct parabolic trajectory when viewed in full. Therefore, the prediction measure produced an outcome different from that of the postdiction event. This prediction–postdiction account may provide a resolution to the discrepancies in findings between infant looking time studies and search studies.

In the looking paradigm, children observe a sequence of events in which a given component occurs out of sight. The object is then revealed at a location that may trigger a post hoc analysis. When the object and wall are both visible, the child determines whether the outcome is consistent. Note that there is no prediction. In search, children do not have an outcome from which to work backward, and there is no triggering event to generate the mental reconstruction. This does not mean that children who do not complete search successfully lack the physical knowledge that solidity impedes the motion of another body. Rather, they are unable to generate a model of what has just happened in the absence of explicit information of the current location of the object. This would suggest that children are not necessarily predicting the outcome in a looking time experiment.

If rich interpretations of looking time are unwarranted, where does that leave the paradigm? The data cannot address whether sensitivity to core principles is built in; if it is, however, additional factors need to be in place before core principles can be used in a predictive manner. Sex and age appear to be two factors that are relevant in this particular task. Whereas age seems to be an obvious candidate for improved performance on search tasks, the role of sex is unclear and requires further replication and analysis. Contrary to Spelke, we do not believe that core principles unambiguously operate to segment the world of object events throughout development. If this were the case, core principles should be capable of supporting predictions. Likewise, we accept the proposal that search failures can be largely attributable to performance limitations. However, rather than obscuring the state of knowledge representations, we argue that the performance constraints of search tasks reveal the fragility of core knowledge. Inevitably, natural selection works on the ability to act in a predictive way, and so knowledge derived from post hoc analyses seems very vulnerable.

However, the current findings do not demean the importance of using looking time as a means of investigating infants' sensitivity to object knowledge. There are examples from other domains in which looking behavior reveals the rudiments or precursors of the full, mature ability at a much earlier age than the onset of functional behavior. For example, in joint attention, infants do not typically turn to see where another's eyes are directed until well into the 2nd year of life (Corkum & Moore, 1995). Yet, sensitivity to the direction of eye gaze can be observed as early as 4 months of age if looking behavior is measured during presentation of faces with diverted gaze (Hood, Willen, & Driver, 1998). In this situation, the interesting developmental story is that additional attentional mechanisms may need to be online before the infant can engage in joint attention. Likewise, future studies of infant object knowledge may profit from approaches that attempt to identify the initial sensitivity and then the nature of the factors that constrain that sensitivity in other situations such as search.

However, this does not necessarily mean that looking time is a precocious measure of physical knowledge because it reflects a violation of expectancy. Consider again the experiment conducted by Kyeong and Spelke (1999). This study showed that preschoolers look longer at an impossible motion of an object launched from a cliff. Yet up to the age of 5 years, preschoolers predict that the object will take the very same trajectory. How can one interpret their longer looking at impossible object motions? Clearly, it cannot be due to a violation of their expectancy because their expectancy based on their beliefs was otherwise. Likewise, adults incorrectly predict that objects continue on a curvilinear path on exiting a curved tube (McCloskey, Caramazza, & Green, 1980) but correctly judge such trajectories to be anomalous when viewed in motion (Kaiser, Proffitt, & Anderson, 1985). If such adults were tested in a looking time study, it is conceivable that they would look longer at these impossible movements, not because they violated their explicit predictions but, rather, because they violated their implicit knowledge of object motion. These examples of naive physics may be exceptions to most commonplace physical knowledge, but they highlight the danger of assuming that the information derived from observing events reflects explicit predictions held by the observer about what should happen (Hood, 2001).

References

Ahmed, A., & Ruffman, T. (1998). Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a non-search task? *Developmental Psychology, 34,* 441–445.

Baillargeon, R. (1995). A model of physical reasoning in infancy. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 9, pp. 305–371). Norwood, NJ: Ablex.

Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. *Child Development, 62,* 1227–1246.

Baillargeon, R., & Graber, M. (1988). Evidence of location memory in 8-month-olds in a nonsearch AB task. *Developmental Psychology, 24,* 502–511.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition, 20,* 191–208.

Bell, M. A., & Fox, N. A. (1992). The relations between frontal brain electrical activity and cognitive development during infancy. *Child Development, 63,* 1142–1163.

Berthier, N., DeBlois, S., Poirier, C. R., Novak, M. A., & Clifton, R. K. (2000). What's behind the door? Two- and three-year-olds reason about unseen events. *Developmental Psychology, 36,* 394–401.

Bogartz, R. S., Shinskey, J. L., & Speaker, C. (1997). Interpreting infant looking. *Developmental Psychology, 33,* 408–422.

Butler, S. C, Berthier, N. E, & Clifton, R. K. (2002). Two-year-olds' search strategies and visual tracking in a hidden displacement task. *Developmental Psychology, 38,* 581–590.

Chugani, H. T., & Phelps, M. E. (1986). Maturational changes in cerebral function in infants determined by FDG positron emission tomography. *Science, 231,* 840–843.

Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 61–83). Hillsdale, NJ: Erlbaum.

Diamond, A., & Goldman-Rakic, P. (1989). Comparison of human infants and rhesus monkeys on Piaget's A-not-B task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research, 74,* 24–40.

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly*? Infant Behavior and Development, 21,* 167–179.

Halpern, D. (1992). *Sex differences in cognitive ability.* Hillsdale, NJ: Erlbaum.

Hauser, M. (2001). Searching for food in the wild: A non-human primate's expectation about invisible displacement. *Developmental Science, 4,* 84–93.

Hood, B. (1994, June). *Searching for falling objects in 2-year-olds is different from watching them fall in 4-month-olds.* Poster presented at the 9th biennial International Conference for Infant Studies, Paris, France.

Hood, B. M. (1995). Gravity rules for 2–4-year-olds? *Cognitive Development, 10,* 577–598.

Hood, B. (2001). When do infants know about objects? *Perception, 30,* 1281–1284.

Hood, B. M., Carey, S., & Prasada, S. (2000). Predicting the outcomes of physical events: Two-year-olds fail to reveal knowledge of solidity and support. *Child Development, 71,* 1540–1554.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adults' eyes trigger shifts of visual attention in human infants. *Psychological Science, 9,* 131–134.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology, 387,* 167–178.

Kaiser, M. K., Proffitt, D., & Anderson, K. (1985). Judgements of natural and anomalous trajectories in the presence and absence of motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 795–803.

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioural sciences.* Belmont, CA: Wadsworth.

Klinnert, M. D. (1985). The regulation of infant behavior by maternal facial expression. *Infant Behavior and Development, 7,* 447–465.

Kyeong, I. K., & Spelke, E. S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science, 2,* 339–362.

Mash, C., Clifton, R., & Berthier, N. E. (2002, April). *Two year-olds' understanding of event outcomes when looking and reaching.* Paper presented at the Conference on Infant Studies, Toronto, Ontario, Canada.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. *Science, 210,* 1139–1141.

Meltzoff, A. N., & Moore, M. K. (1998). Object representation, identity, and the paradox of early permanence: Steps toward a new framework. *Infant Behavior and Development, 21,* 201–235.

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences, 5,* 309–315.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104,* 686–713.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.

Santos, L., & Hauser, M. D. (2002). A non-human primate's understanding of solidity: Dissociations between seeing and acting. *Developmental Science 5,* F1–F7.

Spelke, E. S. (1991). Physical knowledge in infancy: Reflections on Piaget's theory. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 133–169). Hillsdale, NJ: Erlbaum.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99,* 605–632.

Received December 10, 2001
Revision received August 2, 2002
Accepted August 20, 2002 ■