Developmental Psychology
2006, Vol. 42, No. 6, 1103–1115

0012-1649/06/$12.00 DOI: 10.1037/0012-1649.42.6.1103

Blickets and Babies: The Development of Causal Reasoning in Toddlers and Infants

David M. Sobel
Brown University

Natasha Z. Kirkham
Stanford University

Previous research has suggested that preschoolers possess a cognitive system that allows them to construct an abstract, coherent representation of causal relations among events. Such a system lets children reason retrospectively when they observe ambiguous data in a rational manner (e.g., D. M. Sobel, J. B. Tenenbaum, & A. Gopnik, 2004). However, there is little evidence that demonstrates whether younger children possess similar inferential abilities. In Experiment 1, the authors extended previous findings with older children to examine 19- and 24-month-olds' causal inferences. Twenty-four-month-olds' inferences were similar to those of preschoolers, but younger children lacked the ability to make retrospective causal inferences, perhaps because of performance limitations. In Experiment 2, the authors designed an eye-tracking paradigm to test younger participants that eliminated various manual search demands. Eight-month-olds' anticipatory eye movements, in response to retrospective data, revealed inferences similar to those of 24-month-olds in Experiment 1 and preschoolers in previous research. These data are discussed in terms of associative reasoning and causal inference.

Keywords: causal inference, statistical learning, cognitive development

Causal knowledge enables children to interpret the current state of the world rationally as well as engage in predictive inference and explanation. Although Piaget (1929) described young children as precausal, it is clear that they have complex causal learning and inferential abilities. Young children recognize causal properties of objects and events in the physical (e.g., Leslie & Keeble, 1987; Spelke, Breinlinger, Macomber, & Jacobson, 1992), biological (e.g., Inagaki & Hatano, 1993; Kalish, 1998), and psychological domains (Perner, 1991; Wellman, 1990). More generally, children are sensitive to Hume's (1978/1739) principles of causality: temporal priority, spatial priority, and contingency (see, e.g., Bullock, Gelman, & Baillargeon, 1982).

What is missing from this research is a formal description of how children represent and acquire new causal knowledge from the relatively sparse data they are exposed to. Several psychologists have adopted a particular computational model—causal graphical models—as a way of describing how causal knowledge is represented by adult learners (Lagnado & Sloman, 2004; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; Waldmann & Hagmayer, 2001). Gopnik et al. (2004) argued that this computational model offers a description of how children represent causal knowledge and learn causal information given limited data.

Causal graphical models are representations of a joint probability distribution—a list of all possible combinations of events and the probability that each combination occurs. These representations embody conditional probability information among events. Events are represented as nodes, and causal relations are represented as edges between nodes. Critically, these models rely on a particular assumption, the *Markov assumption,* for translating between those causal relations and conditional probability information (Pearl, 2000). The Markov assumption states that the value of an event (i.e., a node in the graph) is independent of all other events except its children (i.e., its direct effects) conditional on its parents (i.e., its direct causes). For example, consider the causal model shown in Figure 1A: A → B → C. In this model, the values of Events A and C are dependent. The Markov assumption states that these values are independent conditional on the value of Event B. C has no children, and B is its only parent. If one wants to predict the value of C and know the value of B, additional knowledge about the value of A does not help: The only influence that A has on C is through B.

It is well known that adults can recognize conditional independence information among events (e.g., Shanks, 1985; Spellman, 1996). Can children reason in this manner? For example, two events (A and B) might co-occur because of a causal relation: A may cause B, or B may cause A. However, these two events might also co-occur because of a third event (C) that causes them both. Can children distinguish between these two situations? Gopnik, Sobel, Schulz, and Glymour (2001) demonstrated that young children can make this inference. They presented children with a "blicket detector," a device that lit up and played music when certain objects were placed upon it. After children learned to label as blickets objects that activated the machine, they were shown the test trial with two objects (A and B). One object (A) activated the

David M. Sobel, Department of Cognitive and Linguistic Sciences, Brown University; Natasha Z. Kirkham, Department of Psychology, Stanford University.

We thank Alisa Breetz, Kathryn Noe, Lisa Smythe, and Daniel Richardson for assistance in data collection and analysis; Jim Morgan, Melanie Soderstrom, Megan Blossom, Carl Dambowski, Marco DeLeon, Lisa Flores, Amelia Hausauer, Debbie Kim, Ayla Nero, Judith Pond, Natalie Ramirez, Karen Rathbun, and Lori Rolfe for assistance in participant recruitment; and Adele Diamond, Alison Gopnik, Laura Schulz, and Daniel Richardson for helpful discussions about this article.

Correspondence concerning this article should be addressed to David M. Sobel, Department of Cognitive and Linguistic Sciences, Box 1978, Brown University, Providence, RI 02912. E-mail: dave_sobel@brown.edu

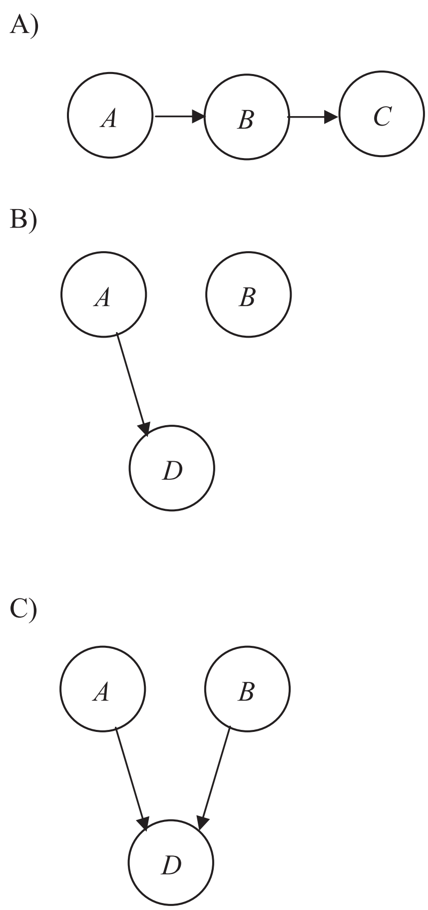

Figure 1. Three examples of causal graphical models representing the causal structure among events. In each, nodes represent the values of events, and edges represent causal relations among those events. Figure 1A indicates that Events A and C are dependent but independent conditional on the value of B. Figures 1B and 1C are the two potential models that could describe data in the screening-off and indirect screening-off trials in Experiment 1. In these figures, Events A and B correspond to the two objects being placed on the blicket detector, and Event D corresponds to the detector activating. In both Figures 1B and 1C, A and D are dependent. In Figure 1B, B and D are independent; in Figure 1C, they are dependent.

machine, and the other (B) did not. Then both objects were placed on the machine together twice, activating it both times.

Figures 1B and 1C show two causal models that are potentially consistent with these events. The nodes labeled *A* and *B* correspond to Objects A and B being placed on the detector, and the node labeled *D* represents the detector activating. The edges represent that the object has the causal efficacy to activate the detector—that is, it is a blicket. At issue in Gopnik et al.'s (2001) experiment was which model children inferred: Would children categorize Object B as a blicket? It did not activate the machine by itself, only dependent on the presence of Object A. Gopnik et al. (2001) found that children as young as 30 months old categorized only Object A as a blicket. Schulz and Gopnik (2004) demonstrated similar findings with 4-year-olds reasoning outside the domain of blicket detectors. These data suggest that children use conditional independence information to make causal inferences and reason according to the Markov assumption.

A difficulty with this conclusion is that recognizing conditional independence in this manner is a form of *blocking* (Kamin, 1969), a phenomenon from the associative reasoning literature. Various associative learning models were designed to explain this phenomenon (e.g., Mackintosh, 1975; Rescorla & Wagner, 1972). These models assume that candidate causes and effects have been identified; they output the strength of each cause–effect association. For example, the Rescorla–Wagner rule computes the change in associative strength of a candidate cause on the basis of the salience of the cause and effect, the potential associative strength the effect can provide, and the current associative strength of all candidate causes present when the cause–effect relation was observed. In the Gopnik et al. (2001) experiment, children made causal inferences by calculating how strongly each object was associated with activating the blicket detector: Object A was more associated with the detector's activation than Object B, so it should be chosen as a blicket.

However, there are three problems in positing associative models as an explanation of causal inference. The first is that many of these models fail to account for certain kinds of retrospective data involving multiple potential causes (e.g., Shanks, 1985). In these cases, learners are presented with initially ambiguous data and then evidence that potentially resolves that ambiguity. Again, it is well known that adults make retrospective inferences (e.g., Shanks, 1985; Shanks & Dickinson, 1987). Sobel, Tenenbaum, and Gopnik (2004) replicated some of these findings with preschoolers. In one condition (the *indirect screening-off* condition[1]), 3- and 4-year-olds observed that two objects (A and B) activated the blicket detector together and then that one of those objects (A) failed to activate the machine by itself. Children inferred that only Object B was a blicket. Further, when asked to activate the machine, they placed only Object B on the detector instead of imitating what they had previously seen activate it.

In a second condition (the *backward blocking* condition), the same children observed that two new objects (A and B) activated the detector together and then that one of those objects (A) activated the machine by itself. Preschoolers made opposite inferences as in the indirect screening-off condition—they labeled Object A as a blicket and Object B as not a blicket. The associative strength between Object B and the blicket detector was the same between the backward blocking and indirect screening-off trials. Children only observed Object B activating the detector in the presence of Object A. When children observe the efficacy of Object A, the associative model we have described posits no change to the associative strength of Object B, because it was not present when Object A was placed on the machine. Object B should be treated the same across these two conditions; children, however, treated it differently.

A variety of contemporary causal learning models do account for these kinds of retrospective causal inference. These models rely

[1] The philosopher Reichenbach (1956) described a screening-off inference as distinguishing between the case in which two events possess a causal relation and cases in which two events are independent from each other conditional on the presence of a third event—the exact basis of the Markov assumption. We adopt this vernacular in the present article. In previous work and in the present article, *indirect screening-off* refers to making this inference without being shown the direct effect of each event.

on either calculations of causal strength derived from measures of associative strength (e.g., Dickinson, 2001; Kruschke & Blair, 2000; Wasserman & Berglan, 1998; see also Rogers & McClelland, 2004) or calculations of causal strength based on the frequency with which events co-occur (e.g., Cheng, 1997; Shanks, 1995). However, a second problem in positing associative models as an explanation of children's causal inference abilities is that these models were designed for cases in which learners have to observe large amounts of data to make accurate estimations of frequency or to reach particular learning states. Children, however, make causal inferences on the basis of only one or two data points.

In addition, a third problem is that these models also fail to account for particular findings in the cognitive and developmental literature. For example, Sobel et al. (2004) introduced children to the blicket detector but first trained them that blickets were either rare or common: Children observed 12 identical objects placed on the detector 1 at a time, and either 2 or 10 of them activated the detector and were categorized as blickets. Then children were given the same backward blocking procedure as before. When blickets were common (i.e., 10 out of 12 objects were blickets), children categorized the B object as a blicket; when blickets were rare (i.e., 2 out of 12 objects were blickets), they did not. Children used the base rate of blickets to resolve the ambiguity presented in the backward blocking paradigm. No associative model predicts this pattern of inference.

On the basis of these and similar data (Kushnir & Gopnik, 2005; Tenenbaum, Sobel, Griffiths, & Gopnik, 2006), Gopnik et al. (2004) argued that causal graphical models describe how children represent causal knowledge not only during the preschool years but throughout development. However, there is little evidence in the present literature to support this hypothesis directly and almost none with very young children or infants. Gopnik et al. (2001) did not examine whether children younger than 30 months could detect conditional independence among events when they observed all the information directly. Sobel et al. (2004) did not test children younger than 3 years on their indirect screening-off or backward blocking measures. Experiment 1 replicates some of these procedures using the blicket detector on 19- and 24-month-olds.

What about even younger children? Infants seem to be able to perceive causal events (e.g., L. B. Cohen & Amsel, 1998; Leslie & Keeble, 1987) as well as engage in associative reasoning in both auditory and visual domains (e.g., Fiser & Aslin, 2003; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996). Infants also show sensitivity to visual spatial relations among repetitive events. For example, 3.5-month-olds can learn simple (two-location), predictable sequences (e.g., Haith, 1993). Wentworth, Haith, and Hood (2002) found that infants recognized spatial contingency in temporal sequences. They presented 3-month-olds with a simple spatiotemporal sequence that was either predictable or random; in some cases there was a contingent relation between the identity of one stimulus and the location of the next. Both the predictable sequence and the contingent trials resulted in more anticipatory eye movement. Furthermore, in displays of greater complexity, infants can also recognize statistical structure. Given exposure to a sequence of static multielement scenes, 9-month-olds recognized the underlying statistical structure of the scene layout (Fiser & Aslin, 2003). Many of these experiments used oculomotor anticipation as the index of learning; researchers considered where infants looked as a measure of their predictive inferences. Experiment 2 makes use of these skills and uses a similar methodology.

These previous studies provide compelling evidence that infants are sensitive to statistical patterns in event occurrence. The question of whether this ability bootstraps an eventual understanding of causality later in infancy and early childhood or whether it indicates a mechanism for causal reasoning that is already in place remains unanswered. In Experiment 1, we examined how toddlers reasoned about a set of novel causal relations concerning objects and the blicket detector. Some of these inferences involved retrospection, which potentially distinguishes between models of associative reasoning and more complex models of causal inference. In Experiment 2, we used an anticipatory eye-tracking measure to assess infants' ability to use retrospection in predicting temporal sequences of events. Critically, the inferences we used examined whether certain associative models (e.g., Rescorla & Wagner, 1972) are consistent with children's predictive inferences. To our knowledge, there are no studies investigating retrospective inferences in the infant statistical learning literature. Experiment 2 offers a first demonstration of such a phenomenon.

Experiment 1

Gopnik et al. (2001) and Sobel et al. (2004) demonstrated that preschoolers could reason according to the Markov assumption and that they did so in a way that certain models of associative reasoning (e.g., Rescorla & Wagner, 1972) did not predict. Experiment 1 attempts to replicate these data in toddlers. Because many of these children were preverbal, we did not rely on children's categorization abilities. Instead, in the present experiment we examined what intervention children would make to activate a blicket detector after being shown various patterns of data. Given data in which one object only activated the detector dependent on the presence of another object, would children recognize that the second object and not the first caused the machine to activate? Would they do so even when they had not observed that information directly? Because Gopnik et al. (2001) demonstrated positive results with 30-month-olds, our goal was to investigate slightly younger children.

Method

Participants. The final sample consisted of twenty-four 19-month-olds (11 girls; range = 18.00–21.00 months; M = 18.75 months) and twenty-five 24-month-olds (12 girls; range = 23.75 months to 26.75 months; M = 24.40 months) recruited from a list of hospital births. Eleven other children were recruited but not included in the analysis; 6 were excluded for failing control tasks (3 from each age group), 4 refused to participate (2 from each age group), and one 24-month-old generated uncodable data. Forty-five of the children were Caucasian, and 4 were Latino/Hispanic. No information about parental education or socioeconomic status was available.

Materials. The blicket detector used by Gopnik et al. (2001) was used. The detector was 5 in. × 7 in. × 3 in. (13 cm × 18 cm × 8 cm), made of wood (painted gray) with a red Lucite top. Two wires emerged from the detector's side; one was plugged into an electrical outlet, and the other was attached to a switchbox, hidden from the child's view, that controlled whether each object activated the detector. When the detector activated, it lit up and played music (*Für Elise*). Objects that activated the detector did so as soon as they came into spatial contact with it, and the detector deactivated as soon as the spatial contact was broken.

Four sets of blocks, 9 in total, all different in shape and color, were used in the experiment. An assortment of approximately 50 blocks was used in the warm-up. None of the blocks used in the experiment was the same shape and color as any of the blocks in the warm-up. The detector and each block set sat on a 12 in. × 20 in. (30 cm × 51 cm) white tray.

Procedure. The child, experimenter, and caregiver first interacted with each other by playing with a set of blocks. After this familiarization, children were brought to a child-size table. The experimenter sat on one side of the table. The caregiver sat on the other side of the table, and the child stood at the table between the caregiver's legs. The blicket detector was on the experimenter's side of the table, sitting on the white tray.

Children were first shown an *imitation* trial. Two blocks were taken out from under the table (hidden from the child) and placed at opposite ends of the tray. One block activated the detector by itself. The other block did not activate the detector by itself. Each block was demonstrated on the detector twice. For each of these events and each of the subsequent events in which an object was placed on the detector, the experimenter narrated the causal efficacy of the object by saying either, "Wow," or, "No."

The experimenter then slid the tray over to the child and said, "It's your turn. Make it go." Although the majority of children produced an immediate response, some did not, and it was necessary to encourage them to "make the machine go." If children did not place the causally efficacious object on the detector, they were excluded from the analysis. This failure demonstrated that children did not understand the basic nature of the task and that their data in further trials could be random. Six children were excluded for this reason, divided equally from the two age groups.

Children were then given three experimental trials, which paralleled causal inference procedures used in previous investigations. The spatial location of the causally efficacious block in each trial was counterbalanced. First, children were given a screening-off trial. Two new blocks were taken out (A and B). Children saw that Object A activated the machine by itself and Object B did not. Then both blocks were placed on the machine together, and it activated. This was demonstrated twice. The experimenter then slid the tray with the blocks and blicket detector over to the child and said, "It's your turn. Make it go." Children were allowed to make one response before the experimenter slid the tray back to his side of the table (and out of the child's reach). This trial was similar to the procedure used by Gopnik et al. (2001).

Second, children were given an indirect screening-off trial. Two new blocks (A and B) were taken out. Both were placed on the machine together twice, and it activated both times. Then one of the two blocks (B) was placed on the machine by itself, and it did not activate. The experimenter then slid the tray over to the child and said, "It's your turn. Make it go." Again, the child was only allowed to make one response before the experimenter slid the tray back to his side of the table. This procedure was similar to the one used by Sobel et al. (2004).

Finally, children were given a backward blocking trial. Three objects were taken out from under the table. Two (A and B) were placed on either end of the tray, and the third (C) was placed in the middle. Objects A and C activated the machine together twice. Then Object C was placed on the detector by itself, and it activated. Object C was then removed, and the experimenter slid the tray with only Objects A and B and the detector over to the child and said, "Now it's your turn. Make it go."

This measure of backward blocking is different from previous research with children (Sobel et al., 2004). It does, however, serve as a control for whether certain associative models underlie children's reasoning. In several associative models (including the Rescorla–Wagner model), the associative strength of Object A is the same between the indirect screening-off and backward blocking trials. In both cases, Object A activates the detector twice with another object, and then that other object either does or does not activate the detector. If children reason on the basis of recognizing associative strength between objects and the detector, then they should treat this object similarly across the two trials (i.e., place Object A on the detector).

Results

Preliminary analyses revealed that neither gender nor the spatial location of the causally efficacious block affected performance. Table 1 shows performance on the three test trials. Because children were given the opportunity to respond only once on each trial, we consider only a nonparametric analysis. We first compared responses between the two age groups. No significant differences on the screening-off or the backward blocking trials were found. On the indirect screening-off trials, 24-month-olds tended to place Object A (the object they had not seen on the machine alone) on the detector by itself more often than the younger children (76% vs. 50%), $\chi^2(1, N = 49) = 3.56$, $p = .059$.

Performance was next considered against chance responding. There were at least three types of responses children could make on each trial. They could place either object on the detector, or they could place both objects on the detector together. For each trial, children did generate all three responses. However, instead of comparing performance with 33%, we felt the most conservative measure of chance would be to compare correct performance on each trial with 50%, as there were only two objects present. According to this criterion, performance on the screening-off task differed from chance levels; 73% of the children placed Object A on the detector alone (binomial test based on a z approximation, $p < .001$). Looking at the two age groups individually, we found that responses from both the 19-month-olds (75% of the time) and the 24-month-olds (72% of the time) differed from chance levels (binomial tests, $p = .023$ and $p = .043$, respectively).

On the indirect screening-off trial, overall performance showed a nonsignificant trend to differ from chance performance; children placed Object A on the detector by itself 63% of the time (binomial test based on a z approximation, $p = .085$). Looking at the two age groups individually revealed a different pattern of performance. The 19-month-olds responded at chance levels: Exactly 50% placed Object A on the detector by itself (binomial test, *ns*). The 24-month-olds responded at a level greater than chance: 76% placed Object A on the detector by itself (binomial test, $p = .015$). Neither age group's performance differed from chance levels on the backward blocking trial.

We next compared whether children treated Object A differently between the indirect screening-off and the backward blocking

Table 1
Distribution of Responses in Experiment 1

Trial type and object	19-month-olds	24-month-olds
Screening-off trial (%)		
Object A (effective)	75	72
Object B (ineffective)	17	24
Both objects together	8	4
Indirect screening-off trial (%)		
Object A (effective)	50	76
Object B (ineffective)	42	16
Both objects together	8	8
Backward blocking trial (%)		
Object A (blocked object)	46	40
Object B (novel object)	42	40
Both objects together	13	16

Note. One 24-month-old generated no response on the backward blocking trial.

trials. Overall, children were more likely to place Object A on the blicket detector by itself in the indirect screening-off trial than in the backward blocking trial (63% vs. 43% of the time), McNemar $\chi^2(1, N = 49) = 4.50$, $p = .031$. Looking at the two age groups individually revealed that the 19-month-olds showed no difference between these two trials (50% vs. 46% of the time). The 24-month-olds, however, did respond differently on these two trials (76% vs. 40% of the time), McNemar $\chi^2(1, N = 25) = 5.82$, $p = .012$.

Finally, because the three trials of interest were presented in a fixed order, we were concerned that performance on one might affect performance on the others. We examined performance on the indirect screening-off and backward blocking trials as a function of performance on the screening-off trial; across both age groups, performance on the screening-off trial did not affect performance on the indirect screening-off trial, $\chi^2(1, N = 49) = 0.23$, *ns,* or backward blocking trial, $\chi^2(1, N = 49) = 0.74$, *ns.* Similarly, performance on the indirect screening-off trial did not affect performance on the backward blocking trial, $\chi^2(1, N = 49) = 2.55$, *ns.* This suggests that performance on the latter trials was not influenced by performance on the trials that came previously.

Discussion

In Experiment 1, 19- and 24-month-olds were presented with a device that produced a novel causal property when objects were placed on it. When allowed to interact with this machine, children placed objects on it that they had observed activate it. In the screening-off trial, when presented with one object that activated the machine alone and another that only activated the machine dependent on the presence of the first object, both 19- and 24-month-olds reliably placed only the first object on the machine. This behavior extends the results of Gopnik et al. (2001) from preschoolers and 30-month-olds to younger age groups and demonstrates that younger children can recognize conditional independence information and reason accordingly.

The indirect screening-off and backward blocking trials revealed a different pattern of performance between the 19- and 24-month-olds. The older children placed only Object A on the blicket detector in the indirect screening-off condition and treated Object A differently between the indirect screening-off and backward blocking trials. This behavior was consistent with the Markov assumption and also inconsistent with certain models of associative reasoning that do not account for retrospective inferences (e.g., Rescorla & Wagner, 1972). Although some more complex models of causal inference potentially explain these data (e.g., Cheng, 1997; Wasserman & Berglan, 1998), a possible conclusion is that 24-month-olds use a similar causal reasoning mechanism as that used by older children in previous research (Gopnik et al., 2001, 2004; Sobel et al., 2004).

It is important to note that the present procedure is a different and possibly imperfect measure of backward blocking. Previous methods presented only two objects that activated the machine together, followed by one of those objects activating the machine by itself. Those studies asked children specifically to categorize the other object. This could not be done with the present intervention procedure. Instead, we examined whether children could track the causal status of three objects, then removed one and examined whether children would differentiate between the blocked object and an object whose causal efficacy was unknown. One concern is that this trial, unlike the other two, required children to keep track of three objects instead of two. This might have been too much for children to remember, resulting in chance performance. However, Ross-Sheehy, Oakes, and Luck (2003) demonstrated that, by 13 months, infants could hold the properties of four different objects in working memory. This suggests that older children could keep track of the causal properties of three objects.

As such, there are several interpretations of both age groups' performance on the backward blocking trial. One possibility is that children recognized that the data they observed were inherently ambiguous. If this was the case, they might have placed either Object A or Object B on the machine at random to determine the causal status of either. It is also possible that children responded in one of two other ways with equal frequency. Children could have reasoned that the blocked object (A) did not activate the detector on its own, so they should choose an object they had not seen on the detector (B). Alternatively, children could have made an inference on the basis of the associative strength between the two objects. To our knowledge, all associative models predict that the associative strength of Object A is greater than that of Object B, because Object B is never placed on the detector (thus, it does not gain any associative strength) and Object A always activates the detector (i.e., its associative strength must be positive). As such, the 24-month-olds' data are difficult to interpret beyond recognizing that these children treated Object A differently between the backward blocking and indirect screening-off trials. As such, in Experiment 2, we wanted to use a different measure of backward blocking that more closely resembles procedures used in previous research.

The 19-month-olds, however, responded in a similar manner on the indirect screening-off trial and the backward blocking trial and did not respond differently from chance on either. They appeared unable to make inferences that involved retrospection about ambiguous data when presented with new evidence. A potential interpretation of these data is that children develop this ability between 19 and 24 months. Before the age of 19 months, children might be able to make causal inferences when all of the data are directly presented (possibly through calculations of associative strength) but might not be able to come to these causal conclusions from retrospection.

However, there are several concerns with this procedure that suggest we might have underestimated 19-month-olds' (and perhaps even younger children's) performance. First, some of the participants could have had superior verbal abilities, making them more likely to understand the "make it go" instruction. This possibility is unlikely: All children included in the analyses successfully imitated the experimenter activating the machine, and the children excluded for failing to do so were equally likely to come from both age groups. Also, children in both age groups had no trouble generating the appropriate inference when shown all the relevant data in the screening-off trial.

Second, there could have been fatigue effects, as children were always given the three experimental trials in the same order. It is possible that 24-month-olds' responses on the backward blocking trial and 19-month-olds' responses on the indirect screening-off and backward blocking trials would have differed from chance if those trials were presented first and the toddler were not fatigued or bored by the procedure. We do not think this is the case. It seems unlikely that toddlers would be fatigued by only one or two trials. Statistical analyses of order effects showed that performance on early trials did not influence performance on later trials. Anecdotally, we note that after the experiment was finished, while the experimenter was debriefing caregivers, most children went to the other side of the room, picked up blocks at random (which were used in the familiarization phase of the experiment), and placed them on the blicket detector. This also suggests that fatigue or boredom had not influenced performance.

A final concern stems from the motor demands placed on the children. Correct responding on the indirect screening-off and backward blocking trial required children to inhibit what they observed and generate a novel intervention on the detector. There is much evidence suggesting that 18-month-olds can imitate goal-directed actions (Uzgiris & Hunt, 1975) and generate novel actions on the basis of observing another's intended goal (Meltzoff, 1995). To succeed on the present tasks, children had to inhibit an imitative response in favor of a novel action. Younger children might correctly reason about the causal efficacy of the objects but fail to make the particular behavioral response required of them. This could have caused them to become confused and imitate previous (ineffective) actions, resulting in chance-level performance.

This dissociation between knowledge and action in toddlers appears throughout the cognitive development literature (e.g., Hood, Cole-Davies, & Dias, 2003; Kirkham, Cruess, & Diamond, 2003; Munakata, 2001; Zelazo, Frye, & Rapus, 1996). Many investigations have suggested that children's poor success in manual search or reaching paradigms reflects performance limitations rather than limited competence (e.g., Baillargeon, Spelke, & Wasserman, 1985). For example, infant researchers have provided robust evidence through both eye-tracking and habituation (looking time) paradigms that young infants can predict the outcome of a ball's partially hidden trajectory, perceive object unity under conditions of occlusion, display their understanding of the continuity principle, and reason about fully hidden objects (e.g., Johnson, Bremner, Slater, Mason, & Foster, 2002; Johnson et al., 2003; Spelke et al., 1992; Wang, Baillargeon, & Brueckner, 2004). All these competencies are evidenced even though older infants demonstrate extremely limited understanding about hidden objects when tested in the manual search A-not-B paradigm (e.g., Diamond, 1991; Diamond, Cruttenden, & Niederman, 1994; Thelen & Smith, 1994), and toddlers and preschoolers perform poorly on search tasks investigating the continuity principle (e.g., Hood, Carey, & Prasada, 2000; Keen, 2003).

Experiment 1 demonstrated that younger children could make some causal inferences similar to preschoolers in previous research, but the data from 19-month-olds suggests possible developmental differences in children's causal reasoning abilities. Because of concerns about the performance limitations of the 19-month-olds in Experiment 1, in Experiment 2 we modified the experimental method to one in which infants could predict the occurrence of a particular event without having to manipulate objects. To do this, we designed an anticipatory eye gaze paradigm. This eliminated the need for participants to select and generate a motor response. Although eye movements are motor responses, in an eye gaze paradigm, participants need only react to the presented events. Eight-month-olds were chosen for this study as they are old enough to attend to and learn from the bimodal stimuli needed for the anticipatory eye gaze experiment (Lewkowicz, 2000; Richardson & Kirkham, 2004) and young enough to sit calmly through an eye-tracking paradigm (i.e., pilot data suggested that 19-month-olds would move their head too much and too often for eye gaze calibration to be successful). Do these infants make retrospective inferences like the older children in Experiment 1?

Experiment 2

Eight-month-olds were shown a sequence of events: Two events together predicted a target event that occurred in a particular spatial location. Then one of the two predictive events occurred alone; this was either followed or not followed by the target event. This presented children with a form of indirect screening-off or backward blocking inference about the appearance of the target event. We then presented infants with the other predictive event and, using an eye tracker, measured their eye gaze in response to the next event's occurrence. If 8-month-olds reasoned about indirect screening-off inferences and distinguished them from backward blocking inferences, it would provide evidence that children's inferential abilities are consistent with a more complex causal reasoning system than calculations of associative strength.

A critical difference between this procedure and the procedures used in Experiment 1 and in previous investigations is the nature of the relation between events. For instance, in Experiment 1 the inference was whether an event would occur (i.e., whether the detector would activate). In the present experiment, the spatial location of the next event must be inferred. However, given the initial information presented to the infants, the experimental procedures do have certain parallels. Both show two events predicting a third, followed by one of those events alone either predicting the third event or predicting an alternative. Similar to the methods used on preschoolers, infants in the current experiment must infer the outcome given the other predictor event by itself.

Method

Participants. Twenty-five full-term 8-month-old infants (13 girls) composed the final sample (M = 8.23 months, SD = 0.30 months). Ten additional infants were observed but not included in the analyses because of fussiness or poor calibration of point of gaze (POG). Thirteen infants (7 girls) were tested in the indirect screening-off condition; 12 infants (6 girls) were tested in the backward blocking condition. The ethnic breakdown of the sample was as follows: Six children were Caucasian, 6 children were Asian American, 1 child was Hispanic, 8 children were of mixed race, and the parents of 4 refused to provide race information. The parental education background of the sample was as follows: Two parents had a high school education, 21 parents had some college education, and 27 parents had a graduate degree. No other information about socioeconomic status was available. Infants were recruited by letter and telephone from an established database of parents. Parents and infants received a small gift (a baby T-shirt or toy) for their participation.

Apparatus. The stimuli were presented by a Macintosh G5 computer and 152-cm rear projection screen. Infants sat on their caregiver's lap 180 cm away from the screen. An Applied Science Laboratories (Bedford, Massachusetts), Model 504 corneal reflection eye-tracking system was used to collect eye movement data as infants were shown the stimulus displays. A remote pupil camera with a pan/tilt base was placed on the table below the projection screen. The stimuli viewed by the infant were imported directly into the eye tracker from the Macintosh. The eye tracker also fed a signal into a mini digital video recorder in the form of crosshairs superimposed on the stimulus. Figure 2 shows the laboratory layout.

Procedure. Infants were randomly assigned to either the indirect screening-off condition or the backward blocking condition. Each session

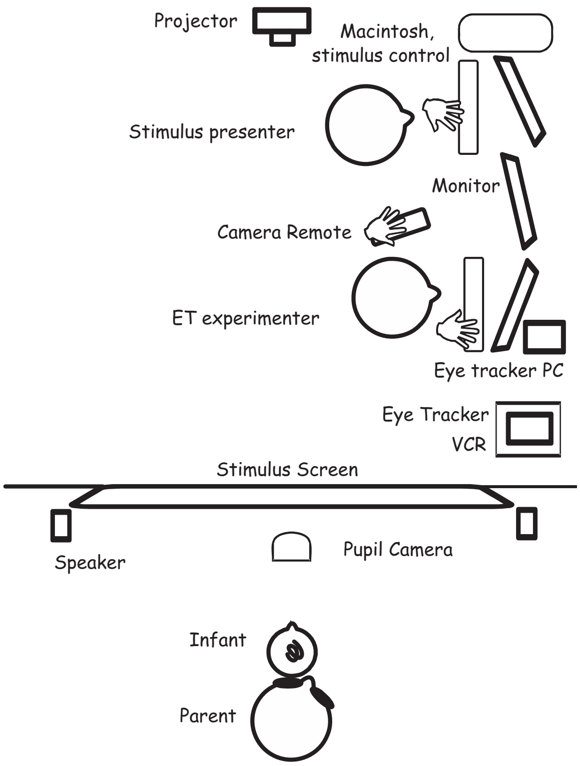

Figure 2. Layout of the laboratory in Experiment 2. ET = eye-tracking.

required two experimenters: an eye-tracking experimenter (ET experimenter) and a stimulus presenter, both of whom sat behind the rear projection screen out of sight of the infant (see Figure 2). The ET experimenter worked the eye-tracking system, watching an image of the infant's pupil, the infant's POG, and the stimulus on a split-screen monitor. If the infant's eye moved away from the presented stimuli, the ET experimenter used a remote control to redirect the pupil camera. The stimulus presenter controlled presentation of the calibration and test stimuli on the basis of when the infant was attending to the projection screen. When the baby was looking directly at the screen (determined by eye gaze and crosshair location), the stimulus presenter started the presentation.

Before the stimulus presentation, the room lights were turned off, and the infant was shown a *Sesame Street* cartoon clip to engage attention. During this time, the pupil camera was directed toward the infant's left eye. After the infant's left eye was in view, the ET experimenter placed the eye-tracking computer in automatic mode, during which the camera remained directed at the pupil despite small displacements of the infant's head (via an algorithm built into the eye tracker). If the infant moved his or her head more quickly than the camera could follow, such that the pupil was lost from view, the ET experimenter changed the computer to manual mode, located the pupil in the camera, and resumed automatic control.

Following acquisition of the pupil image, as the infant watched the cartoon, adjustments were made on the eye tracker to maximize robustness of the POG. This varied somewhat among infants with respect to reflectance of infrared and visible illumination (corneal and pupil reflection, respectively). The infant was then shown a series of looming cartoon movies that made cartoon-style musical noises. This was done to attract the infant's attention so calibration could occur. The eye tracker was calibrated on each infant's left eye according to a two-point calibration routine (i.e., the POG for upper left and lower right locations were used; other locations were interpolated by the computer) and then checked against two different points at the end of the calibration.

If the calibration check was successful, the first block of the experiment began. Each block consisted of 11 exposure trials, 2 disambiguation trials, and 1 test trial. Between blocks of trials, we presented a dynamic attention-getter stimulus to return the infant's POG to the center of the screen. Blocks were presented repeatedly until the infant lost interest in the stimuli and did not return his or her gaze to the screen when the attention-getter appeared. This infant-controlled procedure was used instead of a fixed-block procedure in which all infants receive the same number of blocks to maximize the data collected. This procedure required both good calibration on the part of the experimenter and reasonably stationary posture on the part of the baby.

Stimuli. Stimuli were presented in four frames, which were always present on the screen. Each frame consisted of a box with white outlines presented on a black background. The two frames in the middle of the screen (A and B) were on top of each other. They subtended approximately 11.4° of visual angle, and each was centered approximately 6° from the midline. Two additional frames were on the right and left side of the screen (C and D) and also subtended approximately 11.4° of visual angle. Each was centered approximately 10.3° from the midline. The stimuli consisted of three Quicktime movies of moving objects. One movie was of an object that moved in time to the same repeated sound. This movie could appear in either of the two side frames (C and D). The other two movies were of two different gray objects that moved in silence. These movies appeared in the two central frames (A and B). One always appeared in the A frame, the other in the B frame. All the movies were 8 s in duration.

For each infant, the object appearing in C and D was randomly chosen from a set of four toys and four different sounds. Each object subtended approximately 8° of visual angle. These elements were combined in two Quicktime movies, during which the objects moved within their frame in time with their sound. For example, one sound was a telephone-like ringing sound that occurred once a second; in time to this, the object vibrated. Another sound was a melodic "de-dump" sound that occurred twice a second, during which the objects rotated 45° back and forth. For each infant, the sounds associated with the C and D events were the same. The intention was to ensure that (a) in each event, visual and auditory elements were synchronous, and (b) the two side-occurring events (C and D) were highly attractive to the infant. The central stimuli (A and B) were two different gray shapes that spun slowly without accompanying sound.

Design. Figure 3 presents a schematic of the experimental design. Infants were shown a sequence of events on a projection screen. The initial (i.e., exposure) spatiotemporal sequence was designed such that the two central events (AB) together could be understood to predict an exciting side event (C). Then infants observed two trials in which one of those central events alone did or did not predict the C event. Afterward, infants were shown the other central event alone, and their eye gaze was measured.

The events were organized so that a central event consisting of the two rotating gray shapes (A and B) always preceded the presentation of the brightly colored musical dancing shape occurring on one side of the screen (C). The C side was counterbalanced between participants. A fourth event (D, the same dancing shape) could also occur on the other side of the monitor. The transitional probability between the AB events and the C event was 1.0. All other transitional probabilities among events were 0.5.

After being familiarized to this statistically probable sequence (a total of 11 events, which always included four AB→C sequences), infants were shown one of two disambiguating trials. In the indirect screening-off condition, infants observed one of the two ambiguous events (A or B, counterbalanced) occur alone, followed by the D event. In the backward blocking condition, this event was followed by the C event. After two of these disambiguating sequence pairs, children were shown the other event from the ambiguous pair alone followed by a measurement trial, which consisted of a blank screen (except for the four white frames) and the sound that had accompanied both the C and D events. The eye tracker recorded the infants' eye movements during the entire experiment. Eye movements were recorded for 8 s following the onset of the measurement trial.

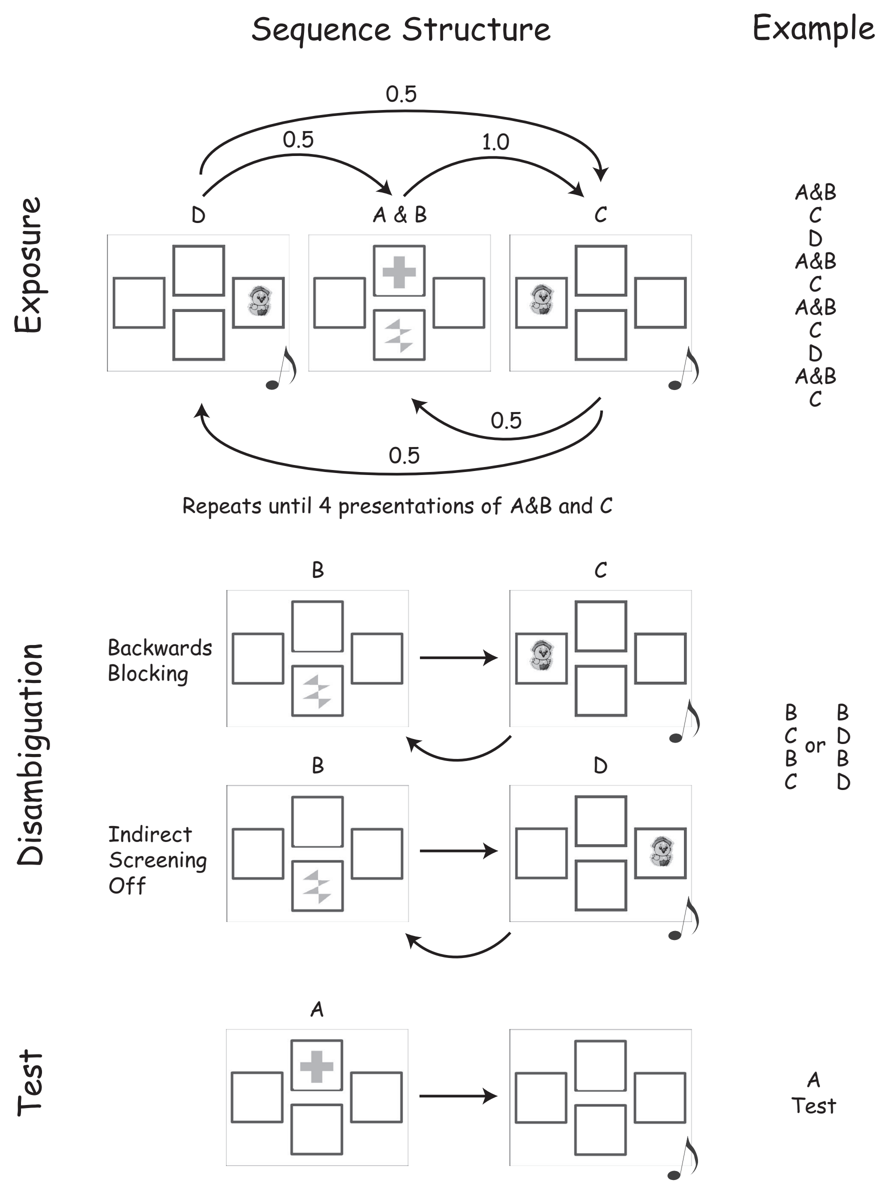

Figure 3. Schematic of Experiment 2 design. In the experiment, the disambiguating trials—A or B alone—were counterbalanced, as was the location of Events C and D.

The key aspect of the event sequence is the spatial location of the target events (i.e., C and D). Instead of manipulating the occurrence of the target event (as in Experiment 1), we manipulated the location of its occurrence. Both the indirect screening-off condition and the backward blocking condition have two central predictive events (A and B) that predict a third (C). Infants are then shown that one of those events predicts either the third event or a fourth event (D), and their looking time is then measured after they are shown the other event. Because this procedure isolates infants' inferences about the C event, these conditions were more similar to the preschooler experiments than the procedure used in Experiment 1.

Results

The eye tracker recorded the duration of gaze along with the location of the gaze on the screen. Infants were also video recorded and coded for fussiness and calibration errors (i.e., if the eye tracker moved to the incorrect eye). If the infant attended to fewer than 50% of the familiarization events, that trial was not included in data analysis. Infants successfully completed between two and seven blocks of trials (*Mdn* = 4). The dependent variable in our analysis was the average amount of time spent looking in the C and D frames over the test trials. Looking times (rather than fixation counts, which are normally used in adult eye-tracking experiments) were analyzed because infant eye position signals can be noisier than those from adults (e.g., Richardson & Kirkham, 2004).

Location of the infants' first look was analyzed but revealed no significant effects. This is not surprising given the nature of the task. In this paradigm, infants were asked to process a sound in absence of a visual stimulus. It is possible that the infants were predicting the location of the test trial solely on the basis of the prior visual stimulus. However, it is also possible that the infants were using the test sound to finalize their predictive decision, and therefore we were interested in looking behavior once they had processed the sound stimulus. Given the time needed to do this in addition to the time needed to program a saccade (on average, 133 ms; cf. Canfield, Smith, Brezsnyak, & Snow, 1997), we would expect the amount of looking over the course of the test trial to be more informative than the infants' first look. In methodologically similar work using a sound stimulus in absence of a visual event, total looking times rather than first looks provided more robust results because of the noise created by saccadic programming time (Richardson & Kirkham, 2004; Richardson & Spivey, 2000).

In our first analysis of the average looking times to the C and D frames, we included gender and location of the predictor event (i.e., A or B frame). We did not analyze looks to the A and B frames during the test trial, as those locations had never been paired with the sounds used. A 2 (gender) × 2 (condition: backward blocking vs. indirect screening off) × 2 (location of predictor event: A vs. B) × 2 (frame: C vs. D) analysis of variance (ANOVA) revealed a significant interaction between condition and frame, $F(1, 17) = 4.97$, $p = .04$. Infants in the backward blocking condition looked longer at the D frame than at the C frame during the test trials, whereas infants in the indirect screening-off condition looked longer at the C frame than at the D frame during the test trials (see Figure 4). There were no significant main effects or interactions involving gender. There was, however, a significant main effect of location of predictor event, $F(1, 17) = 6.54$, $p = .02$. Infants looked longer at the frames overall when Event B was played alone in the disambiguating part of the sequence than when Event A was played alone. However, this factor did not significantly interact with any other factor, nor were there any significant three-way interactions. It seems to be the case that infants' attention to the entire display was increased when they had previously seen a predictor event in the B location. This preference did not affect their looking behavior to either expected or unexpected events and therefore seems to be of no theoretical significance.

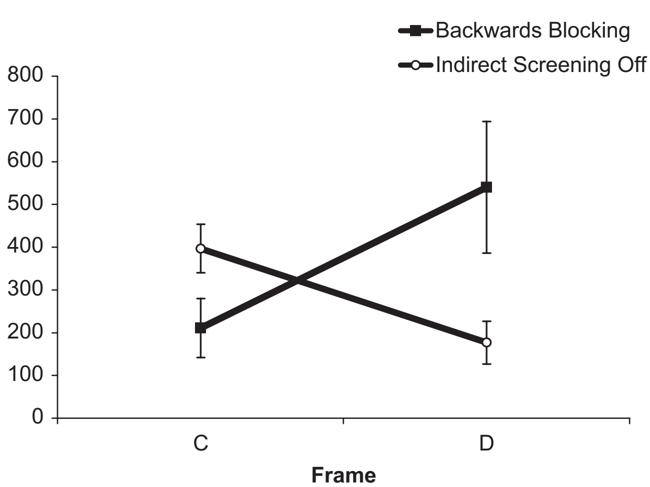

Figure 4. Average looking time to the C and D frames during the test trials across both conditions in Experiment 2.

In planned comparisons of the individual conditions, looks to the C and D frames were coded as expected or unexpected dependent on condition. During the test trials in the indirect screening-off condition, the expected frame was the one that housed the C event, and the unexpected frame was the one that housed the D event. In the backward blocking condition, these were reversed: The expected frame was the one that housed the D event, and the unexpected frame was the one that housed the C event. Average looking time during the test trials was necessarily less than the entire 8-s test display, as only looks within the frames were analyzed.

We compared data from the two conditions separately. In the indirect screening-off condition, a 2 (absolute location: left vs. right) × 2 (frame type: expected vs. unexpected) repeated measures ANOVA revealed a significant main effect of frame type, $F(1, 12) = 6.99$, $p = .02$. During test trials, infants looked longer at the expected frame (i.e., the frame previously associated with the C event) than at the unexpected frame (i.e., the frame previously associated with the D event; $M = 396$ ms vs. 177 ms, respectively; see Figure 5A). Effect size analysis shows that this is a very large effect ($d = 1.1$; J. Cohen, 1988). There was no effect of absolute location; looking times to either side of the screen were not reliably different ($M_{\text{left}} = 269$ ms, $M_{\text{right}} = 302$ ms), $F(1, 12) = 0.70$, *ns.* This suggests that infants did not have an a priori preference to look to one frame over another. Furthermore, preference for the expected frame appeared to emerge quickly in the experiment. Analyses of just the first two blocks also showed a significant advantage of the expected frame, $F(1, 11) = 10.20$, $p = .009$.[2] This suggests that performance did not improve over the experiment but was constant throughout.

In the backward blocking condition, a similar 2 (absolute location: left vs. right) × 2 (frame type: expected vs. unexpected) repeated measures ANOVA revealed a nonsignificant trend in the expected direction. During test trials, infants looked at the expected frame an average of 540 ms and at the unexpected frame an average of 211 ms. Although these means appear to be different from each other in the predicted direction, there was a large degree of variance in looking time, and the difference was only a nonsignificant trend, $F(1, 11) = 3.43$, $p = .09$ (see Figure 5B). There was no effect of absolute location; looking times to either side of the screen were not reliably different ($M_{\text{left}} = 242$ ms, $M_{\text{right}} = 470$ ms), $F(1, 11) = 2.56$, *ns.* Analyses of only the first two blocks also revealed no significant effects.

Similar results were found when looking times were expressed as the proportion of time infants spent looking at the expected frame. In the indirect screening-off condition, infants spent 70.1% of the codable time looking at the expected frame, which was significantly higher than chance, $t(12) = 3.16$, $p = .008$. Although in the backward blocking condition infants looked at the expected frame for 68.7% of the codable time, this was not significant, $t(11) = 1.52$, $p = .16$.

Discussion

During test trials in the indirect screening-off condition, 8-month-olds looked reliably longer at the expected frame (i.e.,

[2] One infant did not provide data during the first two blocks and therefore was not included in these analyses.

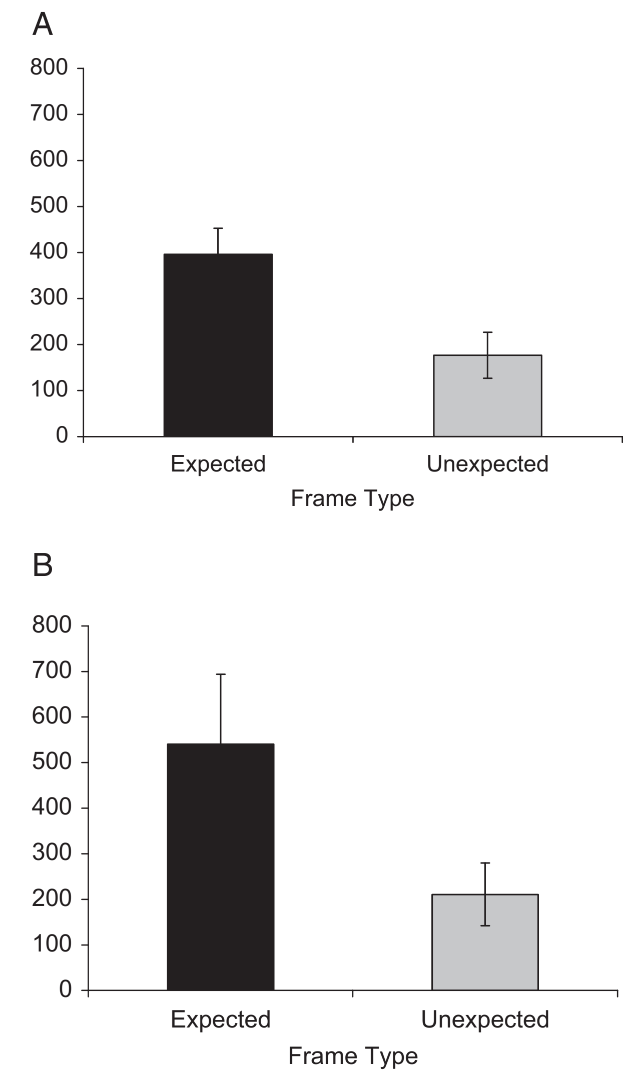

Figure 5. A: Average looking time to the expected (C) and unexpected (D) frames during the test trials in the indirect screening-off condition in Experiment 2. B: Average looking time to the expected (D) and unexpected (C) frames during the test trials in the backward blocking condition in Experiment 2.

where they had previously seen the C event) than at the unexpected frame (i.e., where they had previously seen the D event). In the backward blocking condition, infants showed a different pattern: The advantage of the expected (i.e., the D location) over the unexpected (i.e., the C location) was a nonsignificant trend. It is important to note that there was a significant interaction between condition and location of event (C vs. D). Infants clearly treated the two locations differently between the two conditions. These results suggest that the 8-month-olds were sensitive to the sequence information provided and not only anticipated the location of the visual event but also differentiated between the two locations associated with the auditory cue.

These results are consistent with findings by Richardson and Kirkham (2004), which showed that 6-month-olds could associate a multimodal event with a location and access it with an auditory cue. In that paradigm, infants were given two distinct sounds associated with two separate locations. The present data suggest that, under some circumstances, 8-month-olds could associate one sound with two locations and predict the appearance of an event on the basis of previously presented spatiotemporal information.

One concern with the present procedure is that responses in the indirect screening-off condition might not be the result of a retrospective inference but could be due to *inhibition of return* (IOR; Posner & Cohen, 1983). According to this view, after infants saw the disambiguation trial at the D location, their processing of visual stimuli at that location would be inhibited relative to the contralateral position. This would cause them to stare longer at the C location. However, IOR generally occurs in infants if there is more than 300 ms and less than 1,500 ms between offset of one stimulus (e.g., the cue) and onset of the next (e.g., the target; Butcher, Kalverboer, & Geuze, 1999). The design of this study was such that there was a stimulus inserted between the disambiguation trial at the D location and the presentation of the test trial. The inserted stimulus (A or B, occurring in the central location) was displayed for 8 s, significantly longer than the 1,500 ms window within which IOR occurs. In addition, in this study, we are examining duration of looking responses, not selection.

Another hypothesis consistent with these results is that infants use an alternation strategy: They look first to the right and then to the left throughout the task. This could explain infants' behavior without ascribing to them any actual knowledge of the sequence. During familiarization, for example, the infant might see the C event on the left side, followed by either the AB events in the center or the D event on the right side. Thus, after seeing the C event, regardless of where the next event occurs (center or right), the infant moves his or her eyes to the right. Therefore, an alternation strategy (look to the left, look to the right, look to the left, etc.) would predict a look to the right. Our dependent measure (duration of looks within the small, square, side-located frames) offers support for our original hypothesis. For the infants to succeed, they had to fixate on the frame where the C and D events had previously occurred. In other words, the infant would have to have enough actual knowledge of the sequence to predict the appearance of the object in that exact location, not just move his or her eyes left or right of the previous visual element. In addition, an alternation strategy would predict a significant "first looks" result, which was not found, and would not predict the significant duration of looks finding, which was present.

These results suggest that 8-month-olds can make clear predictions when the information that resolves ambiguity in a sequence of events does so explicitly (in the indirect screening-off condition). When two events predicted a third and one of those two events did not, infants predicted that the other event must predict the third. When that information did not resolve the ambiguity (in the backward blocking condition), infants responded in a somewhat dissimilar way (at least at the level of a statistical trend). When, for example, the AB events predicted C and then the A event alone predicted C, infants tended to respond that the B event alone would predict C. This statistical trend suggests that infants might respond similarly to preschoolers in previous experiments (Sobel et al., 2004) and that the method of investigating backward blocking in Experiment 1 failed to demonstrate toddlers' understanding of the phenomenon. However, the more conservative conclusion that we wish to advocate is that infants simply treat these two sequences of events (indirect screening off and backward blocking) differently. This suggests only that their mechanisms for predictive inference are not based on calculations of associative strength.

General Discussion

Previous research has suggested that a particular computational model—causal graphical models—provides a description of how children represent causal knowledge. Although there is a great deal of evidence suggesting that this computational model describes preschoolers' causal inferences (reviewed in Gopnik et al., 2004), there is almost no evidence investigating younger children. The two experiments in this article examine whether toddlers' and infants' inferences are consistent with the Markov assumption—a fundamental assumption underlying this computational model. In particular, we examined how children reason about conditional independence information and whether they can do so retrospectively.

Experiment 1 demonstrated that 19- and 24-month-olds made causal inferences consistent with the Markov assumption, using a procedure based on Gopnik et al. (2001). When all the evidence was directly observed, children did not treat an object that only activated the blicket detector in the presence of another object as having causal efficacy. The ability to make this inference from retrospective evidence appeared to develop between the ages of 19 and 24 months. Twenty-four-month-olds had no difficulty making these inferences from retrospective data and discriminated between such an inference and another pattern of data in which an object had the same associative strength. Younger children did not discriminate between these trials.

However, Experiment 1 might have underestimated toddlers' knowledge because of the motor demands involved. To test this hypothesis, in Experiment 2, we introduced 8-month-olds to an anticipatory eye gaze paradigm. Infants were shown sequences of events that represented similar retrospective inferences to Experiment 1. In the indirect screening-off condition, 8-month-olds inferred that the test event (Event B presented alone) predicted the target event (Event C) after seeing that Event A alone did not predict Event C. They did this even though they had never seen Event B alone directly predict the occurrence of Event C. Infants in the backward blocking condition, in which Event A did predict the target event (C), did not make this inference, even though the associative strength between Events B and C was the same as in the indirect screening-off condition. These data suggest that when eye gaze is used as a dependent measure, 8-month-olds can recognize condition independence information.

A concern is whether the types of inferences being measured are the same across Experiments 1 and 2. In Experiment 1, children's interventions revealed their knowledge of the causal relation between objects and the blicket detector. Given the experimental set-up, children's interventions should reflect their beliefs about what objects and actions would activate the machine. In Experiment 2, the inference does not necessarily indicate a causal relation between the events, merely a predictive one. It is possible that such predictive abilities reflect a deeper understanding—for example, that there must be a mechanism or reason why these events occur in this particular order. However, such deeper knowledge might only be constructed later in development. What we have demonstrated is that infants' predictive inferences are consistent with the Markov assumption, which suggests that their predictive inferences can be represented by the same computational description of older children's causal knowledge and reasoning abilities.

Furthermore, the present data also suggest that such predictive abilities cannot be represented by certain alternative models designed to represent causal knowledge. In particular, the hypothesis that these inferences are made on the basis of the recognition of associations does not seem to provide the proper framework to explain these data. Models that rely primarily on calculations of associative strength that do not distinguish among forms of retrospective inference, such as the Rescorla and Wagner (1972) equation and other contemporary models based on it (e.g., Cramer et al., 2002), seem inconsistent with the present data. These models do not take into account the associative strength of potential causes that are absent when an effect occurs and thus do not change the associative strength of those events. This suggests that the object not placed on the detector alone (in Experiment 1) and the event not paired with the C event alone (in Experiment 2) will have the same associative strength across both the indirect screening-off and the backward blocking trials. Twenty-four-month-olds in Experiment 1 and 8-month-olds in Experiment 2 did not treat these patterns of data the same.

However, unlike other findings (e.g., the rare–common experiment presented in Sobel et al., 2004) that showed that preschoolers' causal inferences are not likely to be represented by associative models, the present data are consistent with a variety of other learning models. For example, the pattern of results we present could be explained by models that rely on causal strength and were designed with retrospective inferences in mind (e.g., Kruschke & Blair, 2000; Wasserman & Berglan, 1998) as well as various parameter estimation models (e.g., Allan, 1980; Cheng, 1997; Shanks, 1995), which rely on calculations based on the frequency with which events co-occur. For instance, Wasserman and Berglan (1998) posited that when learners observe positive evidence, they decrease the associative strength of any other potential cause not present. Thus, in the backward blocking trial, after infants observed that both events predicted the outcome, the positive relation between one of those events and the outcome decreased the associative strength between the other event and the outcome, making it lower than in the indirect screening-off condition. Similarly, in the ΔP model (Allan, 1980; Shanks, 1995), the causal parameter representing the strength of the relation between a potential cause and effect is the difference between the probability of the effect given the presence of the potential cause and the probability of the effect given the absence of that cause. These models also predict a higher parameter value for the object in the indirect screening-off trial than the backward blocking trial.[3]

[3] To demonstrate this, we note that in the indirect screening-off trial in Experiment 1, whenever Object A was on the detector, the detector activated, and in the one trial it was not on the detector, the detector failed to activate, making its ΔP value = 1. In the backward blocking trial, whenever Object A was on the detector, the detector activated, and in the one trial it was not on the detector, the detector also activated, making its ΔP value = 0.

There are two aspects of these data, however, that are inconsistent with these alternative explanations. As mentioned previously, the first is that children appear capable of making inferences on the basis of a very small sample of data. The alternate models we have discussed all require large amounts of data to make meaningful estimations. In the present experiments, toddlers made proper causal inferences given only 3 or 4 data points, and infants could do so with only 15. Second, many of these models were not designed with these data in mind. As such, one has to describe exactly how they apply. For instance, in Experiment 1, it is not unreasonable to assume that if children were using an associative model, they would do so by choosing to place the object with the higher causal strength or parameter value on the blicket detector. In the backward blocking procedure, these models would all make the same prediction: Because Object A has only been shown to activate the machine and Object B's efficacy is unknown, Object A must have a higher value than Object B and should be chosen. This was not the case. Both 19- and 24-month-olds chose between these objects at chance levels. Although little should be made of a null finding, it is inconsistent with these models. A stronger argument would be to present infants with inferences that are inconsistent with the models we have listed, parallel to the method used in previous research on preschoolers' causal inference (Sobel et al., 2004). We are currently engaged in this endeavor.

Although the present data suggest that 8-month-olds possess retrospective inferential abilities, an open question is whether these abilities develop or are present in even younger infants. Researchers in infant causal perception have documented a developmental difference between 6- and 7-month-olds regarding infants' capacity to view launching events in terms of their causal or spatiotemporal features (e.g., L. B. Cohen & Amsel, 1998; Leslie & Keeble, 1987). It is possible that younger infants would not demonstrate the same pattern of response to the indirect screening-off and backward blocking trials as the 8-month-olds in the present study. Kirkham et al. (2002) found that infants as young as 2 months old could recognize the importance of transitional probabilities when perceiving sequences of shapes. Whether this translates to an ability to make such inferences retrospectively is an open question we are investigating.

In conclusion, 24-month-olds appeared to make causal inferences about activating a physical machine (i.e., a blicket detector) consistent with a particular computational description—causal graphical models. They made these inferences in a manner that some models of associative strength do not explain. Younger children (19-month-olds) showed similar inferential abilities when all the data were directly presented to them but lagged behind the older children in terms of their retrospective inference abilities. However, this lag could be due to the response demands in the procedure; when these demands were eliminated through the use of eye gaze instead of behavioral reenactment as the dependent measure, 8-month-olds appeared no different from 24-month-olds and were similar to older children. Although these data do not eliminate the possibility that children use other models of causal inference, they are consistent with the hypothesis that causal graphical models provide a computational-level description of how children represent their causal knowledge and extend this hypothesis into the 1st year of life.

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15,* 147–149.

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

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209–254). New York: Academic Press.

Butcher, P. R., Kalverboer, A. F., & Geuze, R. H. (1999). Inhibition of return in very young infants: A longitudinal study. *Infant Behavior and Development, 22,* 303–319.

Canfield, R. L., Smith, E. G., Brezsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life: A longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development, 62*(2, Serial No. 250), 1–145.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367–405.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. *Infant Behavior and Development, 21,* 713–731.

Cramer, R. E., Weiss, R. F., Williams, R., Reid, S., Nieri, L., & Manning-Ryan, B. (2002). Human agency and associative learning: Pavlovian principles govern social process in causal relationship detection. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 55*(B), 241–266.

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

Diamond, A., Cruttenden, L., & Neiderman, D. (1994). AB with multiple wells: 1. Why are multiple wells sometimes easier than two wells? 2. Memory or memory + inhibition? *Developmental Psychology, 30,* 192–205.

Dickinson, A. (2001). Causal learning: Association versus computation. *Current Directions in Psychological Science, 10,* 127–132.

Fiser, J., & Aslin, R. N. (2003). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA, 99,* 15822–15826.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111,* 1–30.

Gopnik, A., Sobel, D. M., Schulz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and co-variation. *Developmental Psychology, 37,* 620–629.

Haith, M. M. (1993). Future-oriented processes in infancy: The case of visual expectations. In C. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 235–264). Hillsdale, NJ: Erlbaum.

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., Cole-Davies, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Developmental Psychology, 39,* 61–70.

Hume, D. (1978). *A treatise of human nature.* Oxford: Oxford University Press. (Original work published 1739)

Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind-body distinction. *Child Development, 64,* 1534–1549.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., & Foster, K. (2002). Young infants' perception of unity and form in occlusion displays. *Journal of Experimental Child Psychology, 81,* 358–374.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development, 74,* 94–108.

Kalish, C. W. (1998). Young children's predictions of illness: Failure to recognize probabilistic causation. *Developmental Psychology, 34,* 1046–1058.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Keen, R. (2003). Representation of objects and events: Why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science, 12,* 79–83.

Kirkham, N. Z., Cruess, L. M., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6,* 449–467.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy. *Cognition, 83,* B35–B42.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review, 7,* 636–645.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science, 16,* 678–683.

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 856–876.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25,* 265–288.

Lewkowicz, D. J. (2000). Development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin, 126,* 281–308.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31,* 838–850.

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

Pearl, J. (2000). *Causality.* New York: Oxford University Press.

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Piaget, J. (1929). *The child's conception of the world.* London: Routledge and Kegan Paul.

Posner, M. I., & Cohen, Y. (1983). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 531–556). Hillsdale, NJ: Erlbaum.

Reichenbach, H. (1956). *The direction of time.* Berkeley: University of California Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64–99). New York: Appleton-Century-Crofts.

Richardson, D. C., & Kirkham, N. Z. (2004). Multi-modal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General, 133,* 46–62.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition, 76,* 269–295.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development, 74,* 1807–1822.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology, 40,* 162–176.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 37*(B), 1–21.

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 257–279.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol. 21* (pp. 229–261). San Diego, CA: Academic Press.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28,* 303–333.

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

Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science, 7,* 337–342.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27,* 453–489.

Tenenbaum, J. B., Sobel, D. M., Griffiths, T. L., & Gopnik, A. (2006). *Bayesian inference in causal learning from ambiguous data: Evidence from adults and children.* Manuscript submitted for publication.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Uzgiris, I. C., & Hunt, J. M. V. (1975). *Assessment in infancy: Ordinal scales of psychological development.* Champaign: University of Illinois Press.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82,* 27–58.

Wang, S., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. *Cognition, 93,* 167–198.

Wasserman, E. A., & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgment: The role of within-compound associations. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 51*(B), 121–138.

Wellman, H. (1990). *The child's theory of mind.* Cambridge, MA: MIT Press.

Wentworth, N., Haith, M. M., & Hood, R. (2002). Spatiotemporal regularity and interevent contingencies as information for infants' visual expectations. *Infancy, 3,* 303–321.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11,* 37–63.

Received January 7, 2005
Revision received August 2, 2005
Accepted August 16, 2005 ■