

Running head: CAUSAL INFERENCES

The development of causal learning based on indirect evidence: More than associations

David M. Sobel

Department of Cognitive and Linguistic Sciences, Brown University

Joshua B. Tenenbaum

Department of Brain and Cognitive Sciences, MIT

Alison Gopnik

Department of Psychology, University of California at Berkeley

This work was made possible in part by a National Research Service Award (#F31-MH-12047) to DMS, and NSF grant DLS01322487 to AG. We wish to thank Sara Baldi, Sierra Beck, Meghan Harris, and Khara Ramos for help with the data collection and analysis and Clark Glymour, Tom Griffiths, Tamar Kushner, Molly Losh, Laura Schulz, Lani Shiota, and Dan Slobin, for helpful discussion and feedback on this manuscript. Address correspondence to: D. Sobel, Department of Cognitive and Linguistic Sciences, Box 1978, Brown University, Providence, RI 02912-1978. Phone: (401) 863-3038. Fax: (401) 863-2255. Email: [Dave\\_Sobel@brown.edu](mailto:Dave_Sobel@brown.edu)

Abstract

Previous research on causal reasoning suggests that children can infer causal relations from patterns of events. However, certain associative models suggest that what appear to be cases of causal inference may simply reduce to children recognizing relevant associations among events, and responding based on those associations. In the present experiments, young children were asked to make inferences that relied on indirect evidence. Critically, associative models either made no predictions, or made incorrect predictions about these inferences. In Experiments 1-3, children were introduced to a “blicket detector”, a machine that lit up and played music when certain objects were placed upon it. Children observed patterns of contingency between objects and the machine’s activation that required them to use indirect evidence to make causal inferences. In general, children were able to make these inferences, but a developmental difference between 3- and 4-year-olds was found. We suggest that children’s causal inferences are not based on recognizing associations, but rather that children develop a mechanism for Bayesian structure learning. Experiment 4 explicitly tests a prediction of this account. Children were asked to make an inference about ambiguous data based on the base-rate of certain events occurring. Four-year-olds, but not 3-year-olds were able to make this inference.

The development of causal learning based on indirect evidence: More than associations

### Introduction

As adults, we know a remarkable amount about the causal structure of the world. Developmentalists have shown that much of this causal knowledge is acquired at a relatively early age. For instance, before their fifth birthday, children are able to understand complex causal relations in folk physics (e.g., Baillargeon, Kotovsky, & Needham, 1995; Bullock, Gelman, & Baillargeon, 1982), folk biology (e.g., Gelman & Wellman, 1991; Inagaki & Hatano, 1993), and folk psychology (e.g., Gopnik & Wellman, 1994; Perner, 1991; Wellman, 1990).

These research programs have shown that young children know a great deal about causality and that their causal knowledge changes with age. However, this research has not explained how that causal knowledge is represented, and more significantly, how it is learned. The experiments presented here attempted to discriminate among potential mechanisms for causal learning. To do this, we used an experimental set-up in which children were exposed to particular patterns of evidence about a new causal relation, and were asked to make inferences about that relation. The way children make inferences in this novel experimental setting could shed light on what mechanism for causal learning they have in place. In turn, these mechanisms might be involved in the causal learning we see in the development of everyday physical, psychological, and biological knowledge.

There are two dominant hypotheses as to how children acquire an understanding of causal relations. The first is that children's causal knowledge is primarily composed of a few innate, domain-specific knowledge structures (Keil, 1995; Leslie & Keeble, 1987; Premack, 1990; Spelke, Breinlinger, Macomber, & Jacobson, 1992). On this view, development of

causal knowledge in these domains is reflected by the maturation and/or elaboration of these basic structures. For instance, Leslie and Keeble (1987) argue that the perception of physical causality (such as the perception of “launching” [Michotte, 1962]) is mediated by an innate visual mechanism, which remains unchanged throughout development.

An alternative view, often associated with the “theory theory” of cognitive development, is that children learn causal relations from patterns of evidence, in the same way that an adult scientist might infer causal relations from data (Carey, 1985; Gopnik, 1988; Gopnik & Wellman, 1994; Gopnik & Meltzoff, 1997; Keil, 1989; Perner, 1991; Wellman, 1990). To engage in this learning, children might draw on two sources of knowledge. First, children might apply their existing knowledge of causal mechanisms to observed evidence. For example, if children think that beliefs and desires cause actions, then when they see a new action they might try to understand that action in terms of beliefs and desires. This substantive causal knowledge can range from specific prior knowledge applicable to the current situation (as in the case of beliefs and desires above), to relatively general assumptions about causal structure (e.g., that causes precede effects – a piece of knowledge that potentially discriminates which events are causes and which events are effects).

There is some evidence that children do use their existing knowledge of causal mechanisms to learn new causal relations and make causal inferences (Ahn et al., 1995, 2000; Bullock, Gelman, & Baillargeon, 1982; Shultz, 1982). For example, in all of Shultz’s (1982) “general transmission” studies – traditionally cited as paradigmatic of children using mechanistic information – participants were first trained about what potentially caused the effect (e.g., that in general, flashlights caused spots of light on the wall). Children seemed to be relying on this prior physical knowledge to make new causal inferences.

In addition, children might also rely on a second, more formal kind of knowledge. This formal causal knowledge would allow children to use patterns of data to infer the causal structure among events. In the literature on causal learning in adults, a number of specific mechanisms that might underlie this sort of formal causal inference have been proposed (e.g., Cheng, 1997, 2000; Dickinson, 2001; Shanks, 1995; Tenenbaum & Griffiths, 2001; Van Hamme & Wasserman, 1994). Adults, however, often have extensive experience and explicit education in causal inference and learning. We do not know whether such mechanisms might also be involved in the development of children's causal knowledge. The goal of this paper is to discriminate among a set of possible mechanisms, which potentially provide an account of children's causal learning.

This project assumes that young children can accurately relate patterns of data to causal structure. Some investigations, however, have suggested that children have difficulty understanding how patterns of data relate to causal relations (Klahr, 2000; Kuhn, 1989). In these experiments, children were required to construct experimental manipulations or state what kinds of evidence would be necessary to falsify a hypothesis. Schauble (1996), for example, found that both fifth graders and naïve adults had difficulty designing unconfounded experiments to learn how a set of features (e.g., body shape, engine size, the presence of a tailfin) influenced the speed of a racecar. Although learning did occur over time, both children and adults were relatively impaired in their hypothesis testing and were not able to quickly learn new causal structures.

These studies suggest that young children might have difficulty generating explicit, unconfounded experiments to learn new causal relations. However, young children might still be able to accurately draw causal conclusions from patterns of evidence. Children might implicitly recognize causal relations from data without having the metacognitive capacity to

decide whether a particular piece of evidence violates a causal hypothesis. The formal mechanisms that we will describe, therefore, should not be taken as a metacognitive description of children's causal learning and inference, but rather as a description of how this implicit causal knowledge develops.

We will outline four classes of mechanisms for learning causal relations from patterns of evidence. These mechanisms range from the relatively simple mechanisms of association found in the classical conditioning literature (e.g., Rescorla-Wagner, 1972) to mechanism based on the general learning algorithms used in contemporary artificial intelligence (e.g., Pearl, 1988, 2000; Spirtes, Glymour, & Scheines, 1993, 2001):

1) Learning associations. Children might simply associate causes and effects, in the same way that animals associate conditioned and unconditioned stimuli in classical conditioning (Rescorla & Wagner, 1972; Mackintosh, 1975). On this view, there is nothing to understanding causality beyond recognizing associations. For instance, these models make no predictions about how one could use causal knowledge to generate interventions to elicit effects.

2) Learning causal relations from associations. Some have suggested that causal learning takes place by calculating the associative strength among events, based on an associative model such as the Rescorla-Wagner equation, but then translating that associative strength into a measure of causal strength. That measure of causal strength might then be combined with other types of information to make causal inferences or generate new interventions (see e.g., Cramer, Weiss, & Williams et al., 2002; Dickinson, 2001; Dickinson & Shanks, 1995). In response to the discovery of a set of learning paradigms that the Rescorla-Wagner model has difficulty accounting for, others have suggested alternative mechanisms for calculating causal strength (e.g., Krushke & Blair, 2000; Pearce, 1987; Van Hamme &

Wasserman, 1994; Wasserman & Berglan, 1998). These alternative associative mechanisms also would provide a basis for causal inference and intervention.

3) Rational parameter estimation models. Other investigators have proposed that causal learning relies on the estimation of causal parameters based on the frequency with which events co-occur (e.g., Allan, 1980; Cheng, 1997, 2000; Shanks, 1995). These models calculate an estimate of the maximal likelihood value of the strength of a presumed causal relationship for any amount of data (Tenenbaum & Griffiths, 2001). This distinguishes these models from those described in #2 above: they estimate the strength of a causal model – the probability that an effect occurs given a cause and some background information. The models described in #2 only estimate these strength parameters weakly or at asymptote, as in Shanks' (1995) treatment of RW as an estimation of the  $\Delta P$  parameter (see Griffiths & Tenenbaum, 2001, for a further discussion).

4) Learning causal graphs. Children might construct a “causal graph”: an abstract representation of the causal structure of a set of variables, based on evidence about the conditional probability of those variables (Glymour, 2001; Gopnik, 2000; Gopnik & Glymour 2002; Gopnik, et al., in press; Tenenbaum & Griffiths, in press; Tenenbaum, Griffiths, & Steyvers, 2002). Mathematical theories of such causal graphs, often called “causal Bayes nets”, have been developed in computer science, philosophy, and statistics (Glymour, 2001; Pearl, 1988, 2000; Spirtes, Glymour, & Scheines, 1993, 2001). Several different algorithms for learning causal graphs have been proposed within this general framework. In this paper, we will consider a specific graph structure-learning algorithm – one that relies on Bayesian inference to construct a causal graph, and that takes into account particular types of prior causal knowledge (Tenenbaum & Griffiths, in press; Tenenbaum, Griffiths & Steyvers, 2002).

Which of these mechanisms is responsible for children’s causal learning? Previous experiments have demonstrated that children can infer causal relations from patterns of data. Children use what the philosopher Reichenbach (1956) called “screening-off” reasoning (Gopnik, Sobel, Schulz, & Glymour, 2001). Children’s inferences in these experiments went well beyond the simple sort of classical conditioning algorithms, as in #1 above. In particular, children were able to use screening-off information to construct a novel intervention that they had never experienced before. However, the results of these experiments could be explained by any of the other mechanisms described above, including #2, #3, or #4.

In the current series of experiments, we expand on these earlier findings to differentiate among these potential mechanisms. In Experiment 1, we show that children will not only design a new causal intervention to bring about an effect, but also that they prefer that intervention to one that is based on an observed association. This provides additional evidence that children go beyond simple associative models such as those described in #1. Experiments 2-3 investigate how children reason in a “backwards blocking” paradigm. Classical associative models, such as those described in #1 and a subset of #2 above, have difficulty explaining such inferences, though they can be accommodated by certain modified associative models (#2), by parameter estimation models (#3), and by a range of graphical structure-learning algorithms (#4). Then, in Experiment 4, we present an experiment motivated by the predictions of a particular Bayesian structure learning account formulated by Tenenbaum and Griffiths (in press). In this model, learners systematically take into account baseline prior probabilities when inferring causal structure. We test whether young children will in fact do this. To our knowledge, none of the existing

mechanisms in #1-3 makes predictions about this case, nor do other classes of structure-learning algorithms in the causal graphical model framework.

### An example of causal learning: Screening-off

First, we will use the simple screening-off case to describe the basic reasoning behind graphical structure-learning algorithms. In the process of learning new causal relations, we often assume that contingency is an indicator of causality. However, as any introductory statistics course teaches, there are problems inferring a causal relation when we see that event A is simply correlated with event B. One problem is that a third event, C, might be the common cause of both A and B. For example, you may notice that when you drink wine in the evening, you have trouble sleeping. Temporal priority would suggest that drinking wine (A) causes insomnia (B). However, you might also notice that you only drink wine when you go to a party. The excitement of the party (C) might cause you to drink wine and to lose sleep. This would explain the correlation between wine drinking and insomnia without concluding a causal relation between them. Figure 1a and 1b depict the two potential causal structures that would produce this pattern of observation.

-----  
 Insert Figures 1a and 1b approximately here  
 -----

It is necessary, therefore, to have some way of examining the probability of events A and B relative to the probability of event C. Reichenbach (1956) proposed a natural way of doing this, which he called “screening off”. To determine the cause of your insomnia, for example, you must observe the conditional probabilities of the three events. If you observe that you only have insomnia when you drink wine at parties, but not when you drink wine

alone, you could conclude that the parties are the problem. Likewise, if you observe that you have insomnia when you drink wine at parties, but not when you abstain at parties, you could conclude that wine is the problem. It is also possible that both factors independently contribute to your sleeplessness. By examining the conditional probabilities among events, you can infer which of the two causal structures, shown in Figures 1a and 1b is most likely.

This kind of reasoning goes beyond simply recognizing the associations among the three events; it considers the pattern of dependence and independence among them. This sort of causal inference from patterns of data is ubiquitous in scientific reasoning, and underlies many basic statistical and experimental techniques. Studies with adults have shown that they are capable of screening-off reasoning (Cheng & Novick, 1990; Shanks, 1985; Shanks & Dickinson, 1987; Spellman, 1996). In the artificial intelligence literature, powerful normative causal learning algorithms have been proposed that are generalizations of this basic logic (Pearl, 1988, 2000; Spirtes et al., 1993, 2001).

Gopnik, Sobel, Schulz, and Glymour (2001) investigated whether young children could engage in this form of reasoning. They introduced children to a “blicket detector” (see also Gopnik & Sobel, 2000). The blicket detector lights up and plays music when certain objects (“blickets”) are placed upon it. Thus, the detector presents children with a novel, non-obvious causal property of objects. Children were told that the blicket detector was a “blicket machine” and that “blickets make the machine go”. Children quickly learned this relation. Children were then given two conditions. In the experimental one-cause condition, two objects (A and B) were each placed on the machine individually. Object A activated the machine and object B did not. The objects were then placed on the machine together twice, activating it both times. In contrast, in a control two-cause condition, the experimenter placed object A on the machine by itself three times. Object A activated the

machine each time. Then the experimenter placed object B on the machine by itself three times. Object B did not activate the machine the first time, but did activate it the next two times. Children were then asked whether each object was a blicket.

In both conditions, object A was associated with the machine's activation 100% of the time; object B 66% of the time. The difference between the conditions is that in the one-cause condition, the pattern of dependent and independent relations between the objects and the machine's activation should screen off A from B as a cause of the detector's activation. In this condition, object B only activates the detector dependent on the presence of object A. If children are using screening-off reasoning, then they should categorize only object A as a blicket in the one-cause condition, but should categorize both objects as blickets in the two-cause condition, because in this condition both objects independently activated the machine. Children as young as 30 months old behaved in this manner. Gopnik et al. (2001) concluded that children can use the pattern of independence and conditional independence among events to determine screening-off relations – whether Figure 1a or Figure 1b is the correct causal structure. To do this, children engage in the process of learning a particular causal structure from observation of events.

An objection: Associative models can explain children's screening off reasoning

However, there are alternative accounts for children's behavior in Gopnik et al.'s (2001) "screening-off" experiments. These experiments can be described as a variant of the "blocking" paradigm (Kamin, 1969), which the Rescorla-Wagner (1972) associative model was designed to explain. Children might not recognize anything about causal structure (i.e., that object A being placed on the machine causes it to activate and object B being placed on

the machine does not), but instead might choose object A as a blicket simply because it is more strongly associated with the detector activating.

To test this hypothesis, Gopnik et al. (2001, Experiment 3) examined whether children would generate a previously unobserved intervention that reflected causal knowledge beyond recognizing associations. Children were shown two objects (A and B). Object A was placed on the detector by itself and nothing happened. Object A was removed and object B was then placed on the detector, which activated. Without removing object B, object A was then placed on the machine. Children were asked to make the machine stop, an effect they had never observed. Children used the observed dependence and independence information correctly to design such an intervention: the majority of the children removed only object B from the detector. While this experiment casts doubt that mechanisms described in #1 are responsible for children's causal inference, it says little about the other potential mechanisms.

In the following experiments, we expanded on this basic experimental technique to further explore children's causal learning, and to further discriminate among potential causal learning mechanisms. In Experiment 1, we examined whether children could make an inference about the causal effects of an object when they did not directly observe those effects, and whether children would use this information to generate a novel intervention to elicit an effect, rather than imitating another intervention they had observed would accomplish this goal.

### Experiment 1

In Experiment 1, we investigated children's ability to engage in an indirect "screening-off" inference as well as their ability to make causal interventions based on that

inference. In the one-cause condition in Gopnik et al.'s (2001) experiments, children directly observed that one object activated the detector by itself and that the other object did not. A possible objection to this study is that children might have simply ignored all the trials in which both blocks were placed on the machine, and only paid attention to the independent effects of each object (see e.g., Cheng & Novick, 1992). We need to eliminate this possibility to show that children were genuinely capable of causal inference. Asking children to make inferences about data they did not observe addresses this concern.

Further, in real life, we often have to draw causal conclusions from more complex and indirect patterns of evidence, and to design new interventions in the world based on these inferences. Can young children generate such interventions? Because most associative models have their origins in the Pavlovian conditioning literature, they make few predictions about how learners use associative strength between events to intervene on the world. However, a simple associative account should imply that in designing their interventions, child would reproduce the associations they had just witnessed – that is, they would imitate actions that had been associated with effective results.

In this experiment, we investigated children's performance on a measure of screening-off in which they do not directly observe whether each object is effective. In the Gopnik et al. (2001) experiments, children directly observed the independent effects of each object on the machine. In their one-cause condition, they saw that object A activated the machine by itself and that object B did not. Here, we examine whether children can make causal inferences when the causal effects of objects are not directly observed. Further, when asked to produce an action that elicits the effect: can children produce a novel action, consistent with their inference, or will they simply imitate what they previously observed to be effective?

## Method

Participants. Twenty-one 3-year-olds were recruited from a list of hospital births provided by an urban area university. Five children were excluded for failing control questions (see below), leaving a sample of 16 children between the ages of 37 and 47 months (mean age 42 months).<sup>1</sup> Approximately equal numbers of boys and girls participated in the experiment. While most children were from white, middle-class backgrounds, a range of ethnicities resembling the diversity of the population was represented. No child had ever been a participant in any previous experiment in the lab.

Materials. The “blicket detector” machine used by Gopnik and Sobel (2000) was used in this experiment. The detector was 5” x 7” x 3,” 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, the other was attached to a switchbox. If the switchbox was in the “on” position, the detector would light up and play music when an object was placed upon it. If the switchbox was in the “off” position, the detector would do nothing if an object was placed upon it. The switchbox wire ran to a confederate who surreptitiously controlled whether an object would activate the machine. The wire, switchbox, and confederate were hidden from the child’s view. The apparatus was designed so that when the switch was on, the detector turned on as soon as the object made contact with it and continued to light up and play music as long as the object remained in contact. The detector turned off as soon as the object ceased to make contact with it. This provided a strong impression that something about the object itself caused the effect.

Nineteen wooden blocks of various colors and shapes were used as trial objects. No two blocks were identical. One block – a white square block – was referred to throughout

the experiment as a “blicket”. No other block was white or was of that particular shape. The remaining 18 blocks were grouped into three pairs, used in the training and control trials, and four triples, used in the test phase. Each pair differed from each other in color and shape. An informal pilot experiment with adults suggested that no member of each pair or triple was naturally more similar to the white block. Two white ceramic knobs (approximately 1.5” in diameter) and two small metal tee-joints (approximately 1.5” in length) were used in the pretest screening.

Procedure. All children were tested by a male experimenter with whom they were familiar. Children were brought into a private game room at the university and sat facing the experimenter at a table. Children first received the pretest used by Gopnik and Sobel (2000). Two knobs and two tee-joints were placed in front of the child. Children were told that one of the knobs was a “dax” and were asked to give the experimenter the other dax. After they responded, children were told that one of the metal tee-joints was a “wug” and were asked to give the experimenter the other wug. The pretest ensured that children would extend novel names to objects and would interact with the experimenter.

As in Gopnik et al. (2001, Experiment 2), the blicket detector was then brought out and children were told that the machine was a “blicket machine” and that “blickets make the machine go”. The experimenter took out the white square block and said, “This is a blicket”. The “blicket” was put on the machine, which activated. The experimenter said, “See, it makes the machine go”. Children were told to play with the blicket; they all did so by taking it off and putting it on the machine, which always activated. They were told that the blicket “always makes the machine go”.

Training Phase. Children were then shown a pair of blocks. Each was placed on the detector for approximately three seconds, one at a time. One of them activated the detector

and the other did not. This was demonstrated twice. The white block was then placed on the machine and children were reminded that the white block “is a blicket”. The white block was then held out and the child was asked which one of the two blocks was “like the blicket”. If the child chose the block that had activated the machine, the experimenter moved on. If not, the two blocks were again demonstrated on the machine and the question was repeated. This training trial was repeated with another pair of blocks to ensure the children understood the nature of the machine and the test question.

Test Phase. During the test phase, children were given three types of tasks. In the *one-cause* tasks, they were shown three blocks (A, B, and C). Two of them (A and B, chosen randomly) were placed on the detector together, which activated it. This was demonstrated twice, and each time the blocks were then removed from the detector together. Children were then shown one of the two blocks (A) on the machine by itself, which did not activate it. Children were never shown the effect of the other block (B) by itself. The white block was then brought out and placed on the machine, which activated it. Children were reminded, “The blicket makes the machine go”. Children were asked to give the experimenter the block that was “like the blicket”. After children responded, the object they chose was returned to its original position and they were asked to “make the machine go”. Children received two of these tasks, counterbalanced for spatial location. Different blocks were used on each trial for each child.

In the *two-cause* tasks, children were also shown three blocks (D, E, and F), two (D and E, randomly chosen) were each placed on the machine individually. One block (D) was placed on the machine by itself three times, activating it the first two times and not the third. The other (E) was placed on the detector by itself twice, activating it both times. The white block (the “blicket”) was then brought out and placed on the machine, which activated.

Children were reminded, “the blicket makes the machine go” and were asked to give the experimenter the one that was “like the blicket.” As in the one-cause task, after children responded, the object they chose was returned to its original position and they were asked to “make the machine go”. Children received two of these trials counterbalanced for the spatial location, with different blocks used on each trial.

If children were merely responding to the number of times each object activated the machine, then their performance on the one-cause and two-cause tasks should be similar. Objects A and D each activated the machine 66% of the time, objects B and E 100% of the time, and objects C and F, which were never placed on the machine, never activated the machine at all. Alternatively, if children can make inferences about an object’s causal properties from indirect evidence, then their response to the categorization question should differ in the two tasks. In the one-cause tasks, they should respond that object B, but not A or C, is like the blicket. In contrast, in the two-cause tasks, children should respond that D and E are both like the blicket, but F is not. Even though object E activates the machine 100% of the time and object D activates it 66% of the time, each object activates the machine independently. In a similar condition in Gopnik et al. (2001), children responded that objects were blickets as long as they activated the machine independently, even if they did so probabilistically. In those experiments, children did not show a preference for an object that activated the machine three out of three times over an object that activated the machine two out of three times.

In both the one-cause and two-cause tasks, three objects were always present to control for a possible preference for novelty. If the children choose object B in the one-cause condition, it could be interpreted simply as a preference for the object they had not yet

seen on the machine individually. The third objects (C and F) that are never placed on the machine serve as a control for this response.

The third type of task was a *control* task to ensure children were following the experiment. Children were shown two objects. Each was demonstrated individually on the detector. One activated it; the other did not. Children were given the same instructions regarding the white block and were again asked to give the experimenter the one that was “like the blicket” and were asked to “make the machine go” with only the two objects on the table. Only children who gave the experimenter the block that activated the machine in response to the categorization question were included in the experiment. This was done to ensure that the baseline categorization response to the “like the blicket” question was to place together objects that activated the machine. Five children were excluded for this reason (see note 1). The five trials were presented in random order, with the constraint that the first trial was never the control.

## Results

To examine whether there were order effects, McNemar’s  $\chi^2$  tests were performed on responses to the two one-cause and two two-cause trials. These tests revealed no significant differences between the responses on either trial type, so the data from these trials were combined. Further, initial analysis revealed no effect of order on performance; children who received a one-cause trial first performed in the same manner on both types of task as children who received a two-cause trial first.

### Categorization Question

Table 1 shows the mean responses to the “like the blicket” categorization question and the “make the machine go” intervention question for both the one-cause and two-cause trials. Responses were first examined as a function of task type. A within-subject  $t$ -test revealed that children chose the 100% effective object (objects B and E), as being like the blicket more often in the one-cause trials than in the two-cause trials: 82% vs. 38%,  $t(1, 15) = 2.91, p < .05$ . Further, children chose the 66% effective object (objects A and D), as being like the blicket less often in the one-cause trials than in the two-cause trials: 12% vs. 47%,  $t(1, 15) = -2.42, p < .05$ . Children showed no difference in their choices of the novel object (objects C and F) between the two tasks: 6% vs. 16%,  $t(1, 15) = -.90, ns$ .

-----  
 Insert Table 1 approximately here  
 -----

Children’s choices within each task were then analyzed. On the one-cause trials, children chose object B 82% of the time, significantly more frequently than they chose either object A (12% of the time, Binomial test,  $p < .001$ ) or object C, the novel object (6% of the time, Binomial test,  $p < .001$ ). In contrast, on the two-cause trials, children showed no significant difference between their choices of objects E and D (38% vs. 47%, Binomial test,  $ns$ ) or between their choices of objects E and F (38% vs. 16%, Binomial test,  $ns$ ).

Table 2 shows the number of children who chose the 100% causally effective object zero, one, and two times in response to both the categorization and intervention questions on the two one-cause and two two-cause tasks. Individual children’s choices of each object were analyzed across the two conditions. A Sign test revealed that more children chose the 100% causally effective object in the one-cause task than in the two-cause task:  $p < .05$ .

Likewise, a similar analysis revealed that more children chose the 66% effective object in the

two-cause task than in the one-cause task:  $p < .05$ . No difference was found for the novel object, which is indicative of its infrequent selection across the two conditions. Finally, responses were compared against chance performance. Chi-squared goodness-of-fit tests were run on children's choices of objects B and E, with chance responding set at 33%. On the one-cause trials, responses differed from what would be expected by chance:  $\chi^2(2) = 54.45, p < .0001$ . In contrast, on the two-cause trials, responses did not differ from what would be expected by chance:  $\chi^2(2) = 1.02, ns$ .<sup>2</sup>

-----  
 Insert Table 2 approximately here  
 -----

### Intervention Question

The mean responses to the “make the machine go” intervention question across the two task types are shown in Table 1. Responses to this question were categorized into one of five groups: children placed either the 100% effective, 66% effective, or novel object on the machine by itself, children placed the 100% and 66% effective objects on the machine together, or they placed some other combination of objects on the machine. Importantly, in the one-cause tasks, placing the 100% and 66% objects on the machine together corresponded to imitating the event children observed activate the detector. However, placing only the 100% object on the machine (or any combination involving the 100% object) would also activate it.

On both the one-cause and two-cause trials, the pattern of responses within each condition paralleled the pattern of responses to the categorization question. In the one-cause task, children placed object B on the machine by itself 59% of the time. This response

was more frequent than the response of placing object A on the machine by itself (13% of the time, Binomial test,  $p < .005$ ) or object C on the machine by itself (13% of the time, Binomial test,  $p < .005$ ). This response was also more frequent than the action the experimenter had demonstrated – placing both objects A and B together (9% of the time, Binomial test,  $p < .001$ ). In contrast, in the two-cause task, children placed object E on the machine by itself to make it go 44% of the time. This was not significantly different from their placing either object D or object F on the machine by itself (19% and 28% of the time respectively, Binomial tests, *ns*).

Unlike responses to the categorization question, in a nonparametric analysis, the distribution of responses to the intervention question was similar across the one-cause and two-cause conditions. Sign tests revealed that the distributions of each individual response did not differ between the two conditions. This finding, however, is due to children recognizing that in both conditions, the 100% effective object would make the machine go. In the one-cause trial, however, this was an inference the child had to make. In the two-cause condition, children could respond simply based on their direct observations.

## Discussion

Experiment 1 demonstrated that 3-year-olds could make screening-off inferences, even when they did not observe the direct effect of each object. Children observed that two objects activated the machine together, and then that one of those objects did not activate the machine by itself. From this, they inferred that the other object had the causal power to make the machine activate. In making this inference, children considered information from the presentations of both objects on the machine together; if they were simply ignoring these

presentations, there should have been no difference between their response to object B and to the distracter object C in the one-cause trials.

This experiment replicates and extends the finding that children can use screening-off reasoning to discriminate between an object that independently activates the machine and one that does so dependent on the presence of another object (Gopnik et al., 2001). Furthermore, children's interventions paralleled their inferences. On the one-cause trials, children observed two instances in which the experimenter elicited the effect by placing objects A and B on the machine together. However, children rarely imitated this response. Instead, they generated a novel intervention based on their causal inference. This suggests that the mechanism behind children's inferential abilities is not one of recognizing associations.

One concern with this experiment is the order in which the categorization and intervention questions were asked: the intervention question always followed the categorization question. It is possible that this order influenced children's responses – children might have been primed to select the same object in response to both questions. Other experiments using the blicket detector, however, have used similar categorization and intervention questions, counterbalanced for order across children, and have found no order effects (Kushnir, 2001; Nazzi, 2001; Gopnik & Nazzi, in press). Moreover, the intervention question was not presented as a forced-choice question, but rather as an open-ended request – “can you make it go?” If the order of the questions caused children to recognize that object B would activate the machine by itself, then they would have had to choose between two potentially effective actions. Response levels, however, were not at chance – instead, they clearly favored one action over the other.

The results of this experiment, combined with the intervention experiment in Gopnik et al. (2001), suggest that it is unlikely that children use a simple associative mechanism in their causal learning. However, these experiments do not rule out the possibility that children are translating a calculation of associative strength into a measure of causal efficacy, and designing their interventions accordingly (e.g., Dickinson, 2001; Dickinson & Shanks, 1995). According to the Rescorla-Wagner (1972) model, in the one-cause trials, object B is more strongly associated with the activation of the detector than object A. Thus, performance on both the categorization and intervention questions in Experiment 1 could be explained by the fact that object B simply has greater associative strength. What is necessary is a new paradigm that could discriminate among the computational models that could describe children's causal inferences.

Researchers studying adult causal judgments have described another kind of indirect causal inference relevant to this possibility: *backwards blocking*. In studies on backwards blocking, participants observed an outcome occurring in the presence of two potential causes (A and B). Participants then observed that event A independently caused the outcome. These participants were less likely to judge event B as a cause of the effect than those who did not observe the effect of event A by itself. These backwards blocking tasks require retrospective reevaluation, which models of associative strength like the Rescorla-Wagner equation have difficulty explaining (Shanks, 1985; Shanks & Dickinson, 1987). These models only allow for modification of associative strengths of potential causes that are present on a given trial – observing that event A is independently associated with the outcome should not have any effect on the association between event B and the outcome. Experiment 2 examined how children respond to backwards blocking trials.

## Experiment 2

Experiment 2 adapted a backwards blocking paradigm to the blicket detector. Children were again introduced to the detector and shown two objects (A and B), which together activated the machine. Then, they were shown that object A alone activated the machine. In the one-cause condition of Experiment 1, the logical response was that the object not placed on the detector would activate the detector (and hence was the basis for being categorized with the “blicket”). This is not true in this experiment: object A definitely is a blicket, because it independently activated the machine, but whether object B will activate the machine is uncertain – B might or might not be a blicket. If children showed backwards blocking, however, they would claim that object B was not a blicket.

A straightforward interpretation of the Rescorla-Wagner model would not predict such an effect. To demonstrate this, contrast this paradigm with the one-cause task from Experiment 1. According to the Rescorla-Wagner model, the associative strength of object B (i.e., the object not placed on the detector alone) should be the same in a one-cause and a backwards blocking condition. In both cases, the object appeared twice in conjunction with the other object, and the machine activated on both occasions. The separate association, or lack of association, between object A and the machine’s activation should be irrelevant. In Experiment 2, we included a version of the one-cause task to make this comparison. Would children’s judgments about the causal properties of objects differ between backwards blocking and one-cause tasks?

To make this comparison effectively, it was necessary to make a slight modification to the experimental procedure. On the backwards blocking trial, if we asked children to make a forced-choice response, choosing the object that unambiguously activated the machine would be uninformative – critically, we need to assess what inference children make

about the other object. Thus, instead of asking children a forced-choice question, as in Experiment 1, children were asked whether each object was a blicket. This allowed us to examine whether children believed both objects were “blickets” in the backwards blocking trials, or if they believed only object A was a “blicket”. Further, it allowed us explicitly to contrast children’s categorization of object B between the one-cause and backwards blocking trials.

## Method

Participants. Eighteen 3-year-olds and 16 four-year-olds were recruited from a university-affiliated preschool and from a list of hospital births provided by an urban area university. Two children in the 3-year-old group were excluded (see below). The remaining 3-year-old sample ranged in age from 40 to 48 months (mean age is 44 months). The 4-year-old sample ranged in age from 53 to 60 months (mean age is 55 months). Approximately equal numbers of boys and girls participated in the experiment. While most children were from white, middle class backgrounds, a range of ethnicities resembling the diversity of the population was represented. No child had been a participant in Experiment 1.

Materials. The same “blicket detector” from Experiment 1 was used. Sixteen wooden blocks, different in shape and color, were also used. As in Experiment 1, the blocks were divided into pairs. No pair of blocks was the same color or shape. The metallic knobs and tee-joints from Experiment 1 were also used.

Procedure. After a brief warm-up, children received the same familiarization, pretest, and introduction to the blicket detector as in Experiment 1. They were then shown two blocks. Each was placed on the machine separately. The first activated the detector, and the experimenter said, “See, it makes the machine go. It’s a blicket.” The second did

not activate the detector, and the experimenter said, “See, it does not make the machine go. It’s not a blicket.” This was repeated. The children were then told that the game was to “find the blickets”. Children were then given two training trials. In these trials, they were shown two blocks, only one that activated the machine. After observing each block’s effect on the machine, they were then asked whether each block was a blicket. Feedback was given if children answered these training trials incorrectly. Again, this warm-up and training phase established the idea that individual objects, rather than combinations of objects, caused the machine to activate, and that these objects were called blickets.

The test phase of the experiment involved three types of trials. In the *one-cause* trials, two objects (A and B) were put on the table. The demonstration was similar to the one-cause condition in Experiment 1. Objects A and B were placed on the machine together; the machine activated. This was demonstrated twice. Then, object A was placed on the machine by itself and the machine did not activate. Children were then asked if each object was a blicket. Following this, they were asked to “make the machine go”.

The *backwards blocking* trials were identical to the one-cause trials, except that when the individual object was placed on the blicket detector by itself, the detector activated. Two new objects (C and D) were placed on the detector together and it activated. This was demonstrated twice. Then, object C was put on the detector alone and it activated. Children were then asked whether each object was a “blicket”. After these questions, children were asked to “make the machine go”. The associative strength of object D in this trial is the same as the strength of object B in the one-cause trial. However, in this trial, it is uncertain whether object D has the causal property. If children are using a calculation of associative strength to make their causal inferences, then they should claim that object D does have causal efficacy (and hence, is a blicket). However, if children engage in backwards blocking,

similar to adult participants in previous experiments, then they should not respond in this manner.

The *control* trial was the same as in Experiment 1. To be included in the analysis, children must respond that the object that had activated the machine was a blicket and the object that had not activated the machine was not a blicket. This was done to ensure that children had learned to label objects that activated the machine independently as blickets and those that did not as not blickets, and were paying attention for the entire session. None of the 4-year-olds and two of the 3-year-olds were excluded for this reason. There were two one-cause trials and two backward blocking trials, counterbalanced for spatial location. Children were always asked whether the object that was placed on the detector alone was a blicket first, and then they were asked whether the other object was a blicket. Different block sets were used for each trial. The five total trials were presented in random order with the provision that the control trial was never first.

## Results

To examine whether there were order effects, McNemar's  $\chi^2$  tests were performed on responses on the two one-cause and two backwards blocking trials. These tests revealed no significant differences between the patterns of response on the two trials in each condition, so these data were combined. Further, initial analyses revealed no effect of order on performance: children who received a one-cause trial first performed in the same manner on both types of trials as children who received a backwards blocking trial first. Finally, there did not appear to be any change in performance across the session. Regressions between performance and the order each task was presented in revealed no significant findings.

Based on these analyses, we were convinced that there was no effect of order on children's performance, nor was there learning within the session.

### One-Cause Condition

Table 3 shows responses to the “is it a blicket” categorization question for both types of trials. Likewise, Table 4 shows responses to the “make the machine go” intervention question. Performance on the one-cause trials was similar to performance on the one-cause trials in Experiment 1. Children categorized object A as a blicket 6% of the time, while they categorized object B as a blicket 100% of the time, a significant difference:  $t(1, 31) = 25.18, p < .0001$ . No age differences were found. When asked to make an intervention to activate the machine, children in both age groups often placed object B on the machine by itself (81% and 87% of the time for the 3- and 4-year-olds respectively). This was significantly more frequent than their choice to place either object A on the machine by itself (6% and 0%) or both objects on the machine together (13% for each age group, Binomial tests, both  $p$ -values  $< .05$ ). As in Experiment 1, children made an inference about the causal properties of the individual objects, and did not imitate what the experimenter had done. In addition, in this experiment, children did not make a forced choice between the objects on either the judgment question or the intervention question, and so it seems unlikely that the earlier causal judgment determined the intervention.

-----  
 Insert Tables 3 and 4 approximately here  
 -----

These findings were supported by several nonparametric analyses. First, a Chi-squared analysis revealed no effect of age on responses to either the “is it a blicket”

categorization question or the “make the machine go” intervention question for both objects; 3- and 4-year-olds responded similarly to both questions. Second, a Sign test revealed that responses to the categorization question differed between the two objects. More children categorized the object not placed on the detector (object B) as a blicket than the object that was placed on the detector (object A):  $Z\text{-score} = -5.39, p < .001$ . Likewise, more children placed object B on the detector in response to the intervention question than object A:  $Z\text{-score} = -4.73, p < .001$ , or both objects together:  $Z\text{-score} = -3.97, p < .001$ .

#### Backwards blocking condition

Responses to the backwards blocking trials were quite different from responses to the one-cause trials. On the backwards blocking trials, children categorized object C as a blicket 99% of the time, while they categorized object D as a blicket 31% of the time:  $t(1, 31) = 8.78, p < .001$ . A significant effect of age was found here: on the backwards blocking trials, the 4-year-olds rarely said the D object was a blicket (only 12% of the time), whereas the 3-year-olds were significantly more likely to say so (50% of the time:  $t(1, 30) = 2.82, p < .01$ ). However, both age groups categorized the C object as a blicket more often than the D object: 97% vs. 50%,  $t(1, 15) = 4.39, p < .001$  for the 3-year-olds, and 100% vs. 12%,  $t(1, 15) = 10.25, p < .001$  for the 4-year-olds.

Responses to the “make the machine go” question were also considered on the backwards blocking trials, Chi-squared goodness-of-fit tests were also done on both age groups. Again, responses differed from what would have been expected by chance:  $\chi^2(2) = 6.20, p < .05$ , for the 3-year-olds and  $\chi^2(2) = 19.00, p < .001$  for the 4-year-olds. The 4-year-olds placed the causally consistent C object on the machine by itself (74% of the time), significantly more often than the ambiguous D object alone (7% of the time), or both

objects together (19% of the time): Binomial tests, both  $p$ -values  $< .05$ . The 3-year-olds placed the C object on the machine more often (53% of the time) than the D object, (30% of the time) and placed the D object on the machine more often than they placed both objects on together (17% of the time). However, only the difference between the C object response and the both object response was statistically significant: Binomial test,  $p < .05$ . This suggests that once the 4-year-olds found a single object with the causal property, the other object was blocked from also having the causal property. This pattern was slightly less clear, however, for the younger children.

Nonparametric analyses supported these findings. A Chi-squared analysis revealed no effect of age on responses to the “is it a blicket” categorization question for the object C, but a significant difference in the pattern of responses between the two ages for object D:  $\chi^2(2) = 11.55, p < .01$ . Similar analyses on the “make the machine go” intervention question, however, revealed no effect of age. The pattern of responses between objects C and D also differed. Overall, Sign-tests<sup>3</sup> revealed that more children claimed that object C was a blicket than object D:  $p < .001$ . Further, when asked to make the machine go, more children placed object C on the detector alone than object D:  $p < .05$ , or both objects together:  $p < .05$ . Because 3- and 4-year-olds’ responses differed, these analyses were also run on both age groups separately. Both 3- and 4-year-olds responses to whether objects C and D were blickets did significantly differ:  $p < .05$  and  $.001$ , respectively. However, only 4-year-olds significantly differed in their responses to the intervention question: they were more likely to place object C on the machine alone than object D:  $p < .05$ . Younger children did not show this difference.

#### Comparison between the one-cause and backwards blocking trials

In the backwards blocking trials, once the 4-year-olds found one object that reliably activated the machine on its own, they categorized that object as the only cause. This replicates the results of studies using adult participants (e.g., Shanks, 1985) and of a similar experiment using the blicket machine with adults (Tenenbaum, Sobel, & Gopnik, 2002). Three-year-olds, in contrast, sometimes categorized the other object as a potential cause. It is important to note, however, that even the 3-year-olds categorized object D in the backwards blocking trials at the level of chance and not below chance: 50%,  $\chi^2(2) = 1.00$ , *ns*. Finally, children overall categorized object D as a blicket in the backwards blocking trials significantly less frequently than they categorized object B a blicket in the one-cause condition:  $t(1,31) = 9.34$ ,  $p < .001$ . This was true even when only the 3-year-olds were considered:  $t(1,15) = 4.90$ ,  $p < .001$ . This also suggests that the children – even the 3-year-olds – were not making causal inferences simply on the basis of associative models, such as the Rescorla-Wagner (1972) equation, which predicts that the associative strength between these objects and the machine’s activation would be the same. Rather, it suggests that children’s causal inferences are better explained by a different mechanism.

Finally, individual responses were analyzed in a nonparametric fashion to supplement the parametric analyses above. Table 5 shows the distribution of children who claimed that the object not placed on the machine by itself (objects B and D) were blickets. A Sign test using the Binomial distribution revealed that the frequency of responses that object B was a blicket differed from the frequency of responses that object D was a blicket:  $p < .001$ . Likewise, more children placed object B on the detector in response to the intervention question than object D:  $p < .001$ .

-----  
 Insert Table 5 approximately here

---

## Discussion

Experiment 2 had several goals. The first was to replicate the findings of Experiment 1 using a different method. The one-cause trials here were similar to those in Experiment 1 except for the different manner in which the test question was asked. On these trials, children inferred that object B was a blicket, even though they did not observe its direct effect on the detector. Further, as in Experiment 1, when asked to activate the machine, they did not imitate the response they had observed to be effective (i.e., place both objects A and B on the machine). Instead, their most frequent response was to place only object B on the machine.

A second goal was to see how children would behave in a “backwards blocking” paradigm. On these trials, children were shown two objects that activated the machine together. Then, one of the two objects was demonstrated alone, which also activated the machine. The critical question was how children would categorize the object that had not been demonstrated on the machine alone. Older children did not categorize this object as a “blicket”. Younger children’s responses were less clear: half the time they categorized the object as a “blicket”, and half the time they did not. Importantly, however, children in both age groups categorized this object differently in the backwards blocking trials than in the one-cause trials. This suggests that they were not simply using the strength of the association between each object and the machine’s activation as an indicator of a causal relation. We will return to this issue in the general discussion.

Before we examine the question of whether responses to the backwards blocking trials discriminate among potential models of children’s causal learning, we must address a

methodological concern with this experiment. In the training, control, and one-cause trials, children were always shown two objects – one that activated the detector and one that did not. Because of this exposure, children might have interpreted the “is it a blicket?” question as if it were a forced choice. Since children were always first asked whether the object placed on the detector by itself was a blicket, this might have influenced their responses in both conditions. Thus, it is possible that children were making a much simpler inference: that on every trial, exactly one object was a blicket and exactly one was not. On the one-cause trials, because object A is clearly not a blicket, object B must be. Likewise, on the backwards blocking trials, because object C unambiguously activated the machine, and clearly was a blicket, under a forced choice interpretation, object D must not be a blicket. Experiment 3 explored this possibility.

### Experiment 3

Experiment 3 replicated Experiment 2 with several modifications to address the methodological concern above. First, during the training trial, children saw three objects at a time, instead of two, and more than one object activated the machine on each trial. Thus, the result of the training was that children learned that more than one object in a trial could be a blicket. The one-cause and backwards blocking trials were similar to Experiment 2; we also added trials in which the two objects differed in associative strength, but both clearly activated the machine independently. This was done by presenting children with two objects that both independently activated the detector – one just did so more often than the other. If children interpreted the “is it a blicket?” question as a forced choice, and answered by choosing the object with the higher associative strength, then they would categorize only one of these objects as a blicket. Alternatively, if they were making a causal inference, then since

both objects activated the machine independently, both should be blickets. Given the increased number of trials, pilot work suggested that children became bored with the procedure if they were asked to make the machine go during each trial. Thus, the intervention question was eliminated, since it was not relevant to the methodological concern. Finally, because a developmental difference between the 4-year-olds and 3-year-olds was found in Experiment 2, only a 4-year-old sample was considered here.

## Method

Participants. Sixteen 4-year-olds were recruited from a university affiliated preschool and from a list of hospital births provided by an urban area university. The sample ranged in age from 51 to 63 months (mean age was 58 months). Approximately equal numbers of boys and girls participated in the experiment. While most children were from white, middle class backgrounds, a range of ethnicities resembling the diversity of the population was represented. No child had been a participant in Experiments 1 or 2.

Materials. The same “blicket detector” as in Experiments 1 and 2 was used. Twenty wooden blocks, different in shape and color, and assembled in the same manner as in Experiment 2, were also used. The metallic knobs and tee-joints from Experiments 1 and 2 were also used.

Procedure. After a brief warm-up, children received the same familiarization, pretest, and introduction to the blicket detector as in Experiment 2. They then received a single training trial. Three objects were placed in front of them. Each was placed on the machine one at a time. Two activated it and one did not, determined randomly. Children were asked whether each was a “blicket”. All children correctly answered these questions.

Children then received four types of tasks. The *one-cause* and *backwards blocking* tasks were identical to those in Experiment 2, except that children were not asked to activate the machine after they categorized each object. In addition, children received two trials of an *association* task. In this task, two objects (E and F) were placed in front of the child. Object E was placed on the detector by itself once, activating it. Object F was placed on the detector by itself twice, activating it both times. Children were asked whether each object was a blicket. If children were treating the question as a forced choice and choosing the object with the greater associative strength, then they should choose only object F as a blicket. If children were not interpreting the procedure this way, they should categorize both objects E and F as blickets.<sup>4</sup>

Finally, children were given a *control* task, which was identical to the training trial. Three objects were placed on the machine one at a time; two activated the machine and one did not. Children were asked if each object was a blicket. Children had to correctly categorize the objects that activated the machine as blickets and the one that did not as not a blicket to be included in the analysis. Again, this was done to ensure that children had learned that only objects that activated the detector individually were labeled “blickets”. No children were excluded for this reason. The seven trials were presented in a random order, with the constraint that the first trial was never the control.

## Results

Initial McNemar's  $\chi^2$  tests revealed no difference between responses on the one-cause, backwards blocking, and association trials, so these data were combined. Initial analyses, similar to those done in Experiment 2, also revealed no effect of order. Table 6 shows responses to the “is it a blicket?” question for the three types of trials.

-----  
 Insert Table 6 approximately here  
 -----

On the one-cause trials, children never categorized object A as a blicket. They categorized object B as a blicket significantly more often (94% of the time):  $t(1, 15) = 21.96$ ,  $p < .001$ . This replicates the findings of the previous two experiments: children were able to infer that an object possessed the causal property even when they did not directly observe its influence. Likewise, examining individual performance, a Sign-test using the binomial distribution revealed that the distribution of responses to objects A and B was different:  $p < .001$ .

On the backwards blocking trials, children said that object D was a blicket 34% of the time. This was significantly less than the 100% of the time they categorized object C as a blicket:  $t(1, 15) = 6.01$ ,  $p < .001$ . This replicates the findings of Experiment 2; once the 4-year-olds discovered an object that independently activated the machine, they did not postulate the presence of another cause. Analysis of individual responses revealed a similar finding: a Sign-test using the Binomial distribution revealed that the pattern of responses to objects C and D differed among participants:  $p < .001$ .

Further, responses to object B and D clearly differed between the one-cause and backwards blocking trials; even though these two objects had the same level of associative strength, children categorized object B as a blicket more often than object D: 94% vs. 34% of the time, respectively,  $t(1, 15) = -4.54$ ,  $p < .001$ . Nonparametric analysis supported this finding as well; a Sign-test using the Binomial distribution revealed that the pattern of responses to objects B and D also differed among participants:  $p < .01$ . This suggests that the causal inferences children were engaging in were not based on recognizing associations.

Finally, children showed no difference in their categorization of the two objects in the association condition. Overall, children categorized both the object that activated the machine once and the object that activated the machine twice as a blicket: 94% and 100% of the time respectively,  $t(1, 15) = 1.46, ns$ . A similar Sign test on individual responses was also not significant. This suggests that children did not interpret “is it a blicket?” as a forced choice question.

### Discussion

The goal of Experiment 3 was to replicate Experiment 2 while controlling for the possibility that children interpreted the “is it a blicket?” question as a forced choice. The results of this experiment paralleled the previous one. When children were presented with trials on which both objects clearly activated the machine independently, they categorized both as blickets. Children did not believe that there was always only one blicket on each trial, and that they did not base their answers to the categorization question on the order in which the questions were asked.

### What can backwards blocking tell us about potential mechanisms of causal learning?

The conclusion of Experiments 1-3 is that children’s causal inferences cannot be explained by models that rely only on calculating the associative strength among events, such as the Rescorla-Wagner (1972) model. However, there are other associative mechanisms, based on modifications to the Rescorla-Wagner equation, which can account for backwards blocking (e.g., Wasserman & Berglan, 1998). Similarly, rational parameter estimation models, such as Cheng’s (1997) “power PC” model, generate a strength parameter that is undefined for the blocked object in the standard backwards blocking paradigm – thus

according to Cheng’s model, children should be uncertain about whether it is a blicket. While there is general doubt that mechanisms described above in #1, and some of those described in #2 can account for children’s causal learning, these experiments do not discriminate between other mechanisms described in #2, or the models from #3 and #4.

There is a general problem, however, with appealing to the accounts described in #2 and #3 to explain these data. In Experiments 1-3, children see only two or three trials involving each object. Associative models and parameter estimation models typically assume many more observations of data than the number that occur in our experiments. One advantage of appealing to certain causal graph structure-learning accounts (i.e., #4) is that they would allow children to integrate information about the prior probability of particular kinds of causal relations with the currently observed data. This would allow children to make accurate inferences based on only a few trials, as in our experiments, as well as the previous “screening-off” experiments by Gopnik et al. (2001). We now turn to a description of these models in general, and also of a particular model that might account for children’s causal learning.

#### Varieties of causal graph structure-learning algorithms

The backwards blocking paradigm allows one to illustrate the difference in causal structure learning between associative and rational parameter estimation mechanisms (#2 and 3) and between those models and graphical causal structure-learning mechanisms (#4): Figures 2a and 2b depict the two potential causal structures consistent with the data in the backwards blocking trial. Associative models and models of parameter estimation begin by assuming a fixed causal structure – namely Figure 2a – that both objects being placed on the machine (A and B) are each potential causes of the machine activating (E). According to

these models, 4-year-olds' responses in Experiments 2 and 3 indicate that their calculation of the strength of the link between B and E is zero, or below a threshold level at which objects are labeled "blickets".

-----  
 Insert Figures 2a and 2b approximately here  
 -----

Graphical structure-learning algorithms interpret the backwards blocking data in a different manner. These algorithms do not calculate the strength values of an established causal structure. Instead, they use the observed data to determine what that causal structure is – in this case, whether Figure 2a or 2b is the causal structure. They can also determine the strength values of the causal relations. On this view, the 4-year-olds in Experiments 2 and 3 recognized that the more plausible causal structure given the observed data was Figure 2b, and responded in that manner.

Traditionally, such structure-learning algorithms have been divided into two classes: constraint-based learning algorithms and Bayesian learning algorithms. Constraint based algorithms construct a structure consistent with dependence and independence relations that are found in the observed data. For example if event A causes event B, and event B causes event C, then A and C will be statistically dependent. However, events A and C will (in general) be statistically independent conditioned on the state of B. The statistical dependence between A and C suggests a causal link between them, but the conditional independence of A and C given B shows that this link is not a direct causal connection. Constraint-based algorithms, such as TETRAD (Scheines, Spirtes, Glymour & Meek, 1994) used statistical tests of independence to determine these relations. These algorithms potentially serve as a mechanism for children's causal learning. Glymour (2001) has

described how this type of algorithm can model the screening-off experiments of Gopnik et al. (2001), and Gopnik et al. (in press) describe a similar model of the backwards blocking data from Experiments 2 and 3.

In contrast, Bayesian structure-learning algorithms start with a set of hypotheses that could have generated the observed data. Each of those hypotheses is assigned a prior probability. Given the data actually observed, the prior probability of each hypothesis is updated by an application of Bayes' rule to yield a posterior probability that each hypothesis is the actual causal structure of the system (for a general reference to these algorithms, see Heckerman et al., 1995). These algorithms also potentially serve as a mechanism for children's causal learning: Tenenbaum and Griffiths (in press; Tenenbaum, Griffiths, & Steyvers, 2002) describe how a particular Bayesian structure-learning algorithm could account for both the screening-off and backwards blocking data.

In general, both types of algorithms can be evaluated on two dimensions that are particularly relevant to children's causal learning: the ability to use prior knowledge, and the ability to make inferences from small sample sizes. First, as we have seen, children's causal learning relies not only on the formal mechanism for learning structure from observed data, but also on various types of prior knowledge that might affect causal learning. Structure-learning algorithms differ in the extent to which they integrate prior knowledge to make inferences, and in the methods used to integrate such knowledge. In general, this is a qualitative difference between the constraint-based and Bayesian approaches. The constraint-based approach often relies on no prior knowledge, or certainly less prior knowledge than the Bayesian approach. Some kinds of prior knowledge, however, such as the knowledge that causes precede effects can be used by these algorithms. This prior knowledge can be used in two ways. First, once a set of causal relations is determined, it can

constrain the possible structures generated by the algorithm (by, for example, eliminating structures in which effects would have to precede causes). Second, it can also be used to alter the significance level that is used in testing for dependence.

In contrast, Bayesian algorithms integrate prior knowledge in a more natural and elegant manner. Bayesian learning algorithms compare the prior probabilities of particular causal hypotheses to their posterior probabilities, given the evidence. These algorithms must rely on prior knowledge to constrain the initial hypothesis space and generate initial prior probabilities on those hypotheses. These prior probabilities, therefore, can easily be adjusted to reflect existing knowledge. For example, in the blicket detector experiments, knowledge of causal mechanisms might influence what initial hypotheses are under consideration (see e.g., Ahn et al., 1996). Tenenbaum and Griffiths's (in press) Bayesian structure-learning account of the backwards blocking data explicitly relies on children's knowledge of causal mechanisms (see below).

In addition to the role of prior knowledge, structure-learning algorithms also differ with regard to their treatment of small sample sizes. As noted above, the results of Experiments 2 and 3, as well as the screening-off experiments of Gopnik et al. (2001) can be explained by mechanisms from classes #2 and #3, as well as both types of structure-learning algorithms in class #4. However, the methods in #2 and #3 have difficulty dealing with the very small sample sizes in these experiments. Different structure-learning algorithms also treat small samples differently. In particular, most constraint-based algorithms rely on statistical tests of independence, like the chi-squared or  $G^2$  test, to generate inferences about causal structure (Spirtes, et al., 1993, 2001; see also Griffiths & Tenenbaum, 2001). In practice, these tests usually require large number of observations to make inferences. In order to deal with small samples, these algorithms have to assume that these samples are

representative of the data, for example by assuming that each sample is multiplied by some large constant when testing significance (see Glymour, 2001; Gopnik et al., in press, for further discussion).

In contrast, Bayesian models can deal with small samples by using prior knowledge to adjust the prior probabilities of different causal structures. Tenenbaum and colleagues (Tenenbaum & Griffiths, 2001, in press; Tenenbaum et al., 2002) has proposed a particular type of Bayesian structure-learning algorithm that relies heavily on prior knowledge, and which uses this prior knowledge to make accurate causal inferences based on a minimal number of observations. They have shown that a related Bayesian structure learning model makes more accurate predictions about adult causal structure learning data than either a probabilistic contrast model (Lober & Shanks, 2000; Shanks, 1995) or Cheng's (1997) "power PC" model (Griffiths & Tenenbaum, 2001).

These models rest on several assumptions: (1) children recognize that the presence of each block on the machine (events A and B) either is sufficient to cause the machine's activation (E) or is causally unrelated to E; (2) the causal status of A and B are independent; and (3) E does not occur unless A or B occurs. Bayesian updating then licenses the following inferences. After observing events A, B, and E (i.e., observing that blocks A and B activate the machine together), the posterior probability that A is a cause of E increases above its prior probability, and similarly for B. However, after observing that B alone activates the detector, these quantities diverge. The probability that B is cause of E becomes 1, because otherwise this event could never be observed. The probability that A is a cause of E returns to its baseline prior probability, because knowing that B is surely a cause of E makes the compound case, in retrospect, uninformative about the causal status of A. In Experiments 2 and 3, the assumption is that this prior is relatively low, and the data that the

children observe before being asked the test question in the backwards blocking trial is that approximately 50% of the objects are blickets (for a more detailed description of this model, see Tenenbaum & Griffiths, in press; Tenenbaum et al., 2002).

This reliance on prior probabilities motivates a prediction about performance on a backwards blocking trial: if the prior probability that objects are blickets is high enough, and this probability is explicitly understood by the children, then the observed data should not be sufficient to justify a backwards blocking response. In Experiment 4, we manipulated the prior probability that objects activated the detector (and hence, were blickets) to test whether this class of algorithm can explain children's causal learning. To our knowledge, no associative or parameter estimation model predicts a difference on backwards blocking given this manipulation of the prior probabilities.

#### Experiment 4

In Experiment 4, children were given a variant of the backwards blocking task from the previous experiments. The critical difference was that before being shown the backwards blocking trial, children were first given a training phase in which the frequency of blickets was varied – blickets were either rare or common. According to a Bayesian structure-learning algorithm, if blickets were rare, then children should reason as in Experiments 2 and 3, and infer that the uncertain object is not a blicket. This is because a model with only one causal relation should have a higher posterior probability than a model with two causal relations, given the observed data. However, if blickets were common, then participants should infer that the uncertain object is a blicket. In this case, the prior probability that the graph shown in Figure 2a is the correct model should be relatively high. The resulting posterior probability based on the observed backwards blocking data should

not provide enough evidence to override the information about the prior probability of the model in which both objects are blickets.

Thus, the Bayesian approach makes a set of explicit predictions regarding children's sensitivity to prior probabilities:

- 1) Regardless of whether children were trained that blickets were rare or common, they should categorize object A as a blicket in the backwards blocking trial.
  - 2) Children in both conditions will be less likely to categorize object B as a blicket in the backwards blocking trial than object A. The difference between their categorization of objects A and B, however, will be greater in the rare condition than in the common condition.
  - 3) Children will be more inclined to categorize objects as blickets in the baseline trials when blickets are common – that is, children will recognize that when blickets are rare, fewer objects are likely to be blickets.
- Differences in responses on the backwards blocking trial will still be significant when this baseline measure is taken into account.

## Method

Participants Thirty-eight 3-year-olds and 33 four-year-olds were recruited from two suburban area preschools and from a list of hospital births provided by an urban area university. One 3-year-old was excluded because of experimental error. Five 3-year-olds and one 4-year-old were excluded for failing control questions (see below), leaving a sample of 32 children in each age group. The 3-year-olds ranged in age from 35 to 47 (mean age 42 months) and the 4-year-olds ranged in age from 47 to 63 months (mean age 53 months).

Approximately equal number of boys and girls participated in the experiment. While most children were from white, middle-class backgrounds, a range of ethnicities resembling the diversity of the population was represented. No child had ever been a participant in any previous experiment in the lab.

Materials. The same “blicket detector” as in the previous experiments was used. Eighteen blue wooden cylindrical blocks were used. These blocks were held in a 12” x 12” x 4” white cardboard box. Two smaller 6” x 12” x 2” white cardboard boxes were also used. One had the word “Blickets” printed on it. The other had the words “Not Blickets” printed on it. The metallic knobs and tee-joints from Experiments 1 and 2 were also used.

Procedure. Children were first given the same “daxes” and “wugs” pretest as in previous experiments. Children were then introduced to the blicket detector, by being told that it was a “blicket machine” and that “blickets made the machine go.” The box containing the blocks was brought out and children were told, “I have this whole box of toys and I want to know which are the blickets.” Children were randomly assigned to one of two conditions. In the Rare condition, children were told, “It’s a good thing we have this machine because only a few of these are blickets. Most of these are not. It’s very important to know which are which.” In the Common condition, children were told the opposite: “It’s a good thing we have this machine, because most of these are blickets, but a few of them are not. It’s very important to find out which are which.”

Two blocks were then taken out of the box and the experimenter said, “Let’s try these two”. The blocks were placed on the machine together and the machine activated. The experimenter said, “Look, together they make it go. Now let’s try them one at a time.” One of the two blocks was then placed on the machine and the machine activated. The experimenter said, “Wow. Look, this one makes the machine go by itself. It’s a blicket. I

have this box and it says ‘blickets’ on it. Let’s put the blicket in the blicket box.” The experimenter put the block that just activated the machine into the box labeled “blickets”. The experimenter then said, “Now let’s try this other one.” The other object was put on the machine, which did not respond. The experimenter said, “Wow. Look, it did not make the machine go by itself. It is not a blicket. I have another box that says ‘Not Blickets’ on it. Let’s put this one in the ‘Not blicket’ box.” The experimenter then did so.

Next, the experimenter said, “Remember, when we did them together – together they made the machine go.” This was demonstrated with the two blocks. “But that’s because the blicket <hold up the blicket> made the machine go, but this one <hold up other object> did not make the machine go.” Each block was demonstrated individually with their proper effect on the machine. This was done to make sure that children understood the machine as having an “or” function: the machine would activate even if only one of the blocks on it was a blicket.

Two new blocks were taken out of the box and each object was placed on the machine. After children saw the effects of each object, the experimenter asked, “Where do these go?” After the child made their response, the experimenter confirmed it by asking, “Just to make sure, is this one a blicket/not a blicket?” for each block. Five such pairs were demonstrated (10 blocks in all). In the rare condition, only one out of the ten made the machine go (randomly determined). In the common condition, nine out of the ten made the machine go (randomly determined).

After the ten blocks were demonstrated, the machine and box of remaining blocks were removed. Children were asked to look at the “blicket” and “not blicket” boxes. In the rare condition, children were told that, “Most of the blocks we saw were not blickets. A few of them were, but almost all of the ones we tried were not blickets.” In the common

condition, children were told the opposite. This was done to remind children about the base rate.

*Test Phase.* The machine and remaining blocks were then placed back on the table for the test trials. In the first trial (backwards blocking), two blocks (A and B) were taken out of the box. The experimenter placed the two blocks on the detector together, which activated. Then block A was put on the detector alone and the detector also activated. Then, children were asked which box each block should go into. If the child said that they did not know, the experimenter asked the child to take a guess.

After this trial, children were given a baseline trial. Two more blocks were brought out; children saw that they activated the machine together. Children were then asked into which box these blocks should be placed. Finally, a control trial was done to ensure that children were on task. Two more blocks were brought out. Each was placed on the machine, one at a time. One made it go and one did not (randomly determined). Children were then asked to put the blocks into the box where they belonged. If the child did not correctly categorize these blocks (place the block that made the machine go in the “blicket” box and the one that did not make the machine go in the “not blicket” box), they were not included in the analysis. Five 3-year-olds and one 4-year-old were excluded for this reason.

## Results

In line with the first prediction above, in the backwards blocking trial, all children placed the block that unambiguously activated the machine alone in the blicket box. Table 7 shows the probability that children in the rare and common conditions placed the uncertain block in the blicket box on the backwards blocking trial, the probability that children placed

either block in the blicket box on the base rate trial, and the difference between those two probabilities.

-----  
 Insert Table 7 approximately here  
 -----

In line with the second prediction above, categorization of the uncertain B object on the backwards blocking trial was subjected to an Analysis of Variance with condition and age group as between-subject factors. This analysis revealed that children placed the uncertain B object in the blicket box more often in the common condition than in the rare condition: 84% vs. 53%,  $F(1, 60) = 9.74, p < .01$ . Further, a significant effect of age was found; across the conditions, more 3-year-olds categorized this object as a blicket than 4-year-olds: 84% vs. 53%,  $F(1, 60) = 9.74, p < .01$ . Finally, a significant age x condition interaction was found:  $F(1, 60) = 6.23, p < .05$ .

Individual *t*-tests revealed that on the backwards blocking trial, 4-year-olds placed the uncertain B block in the blicket box more often in the common condition than in the rare condition: 81% vs. 25%,  $t(1, 30) = -3.74, p < .001$ . When blickets were rare, children responded that the cause of the machine's activation was the block that independently activated the machine. When blickets were common, 4-year-olds were less likely to respond in this manner. Instead, the most probable account was that the B object was a blicket. Three-year-olds, in contrast, did not make this distinction. For the most part, 3-year-olds categorized the uncertain B object as a blicket, regardless of whether they were trained that blickets were rare or common: 81% vs. 87% respectively,  $t(1, 30) = -0.47, ns$ .

Two nonparametric procedures were also conducted to examine the effect of prior probability information on individual response patterns. First, a chi-squared analysis was run

on children's pattern of response. In the 4-year-old sample, four of the sixteen children in the rare condition placed the B object in the blicket box in the backwards blocking condition. In contrast, thirteen of the sixteen children in the common condition did so:  $\chi^2(1) = 10.17, p < .001$ . Second, a Mann-Whitney  $U$  procedure was performed on the probability that 4-year-olds would place object B in the blicket box in the backwards blocking trial, and either object in the blicket box on the baseline trial. For the backwards blocking trial, a significant difference was found between the rare and common condition ( $U = 56, p < .01$ ). This was not the case for performance on the baseline trial; performance on this trial only showed a tendency to differ ( $U = 80, p < .10$ ). Similar analyses were run on the 3-year-old sample with no significant results. Four-year-olds' responses demonstrated that they used the prior probability an object is a blicket in order to resolve the ambiguity of the backwards blocking trial.

Four-year-olds' recognition of the prior probability of blickets extended to the baseline trial. In the rare condition, 4-year-olds were less likely to categorize blocks as blickets than in the common condition: 78% vs. 97%,  $t(1,15) = -2.63, p < .05$ . This difference alone, however, does not account for the difference in the backwards blocking trial: in the rare condition, 4-year-olds did not just put fewer objects into the blicket box, than children in the common condition. The difference between the probability that children categorized the B object as a blicket in the backwards blocking trial and the probability that children categorized a block in the base rate trial as a blicket differed between the rare and common conditions: 53% vs. 16%,  $t(1,15) = -2.87, p < .01$ . This is consistent with the third prediction above. The difference in responses between the rare and common conditions shows that children used information about prior probabilities to

resolve the uncertainty of the backward blocking condition; the difference in conditions was not due to children simply placing fewer blocks in the blicket box in the rare condition.

### Discussion

Children were introduced to the blicket detector and given the same introduction as in the previous experiments. Children were then trained that the occurrence of blickets was either rare or common. The experiment tested whether children could use this information to guide their inferences about ambiguous data. Four-year-olds seemed capable of using this information: when blickets were rare, they categorized objects whose causal properties were uncertain as not a blicket; when blickets were common, they categorized the same objects as blickets. Three-year-olds, however, seemed insensitive to this manipulation of prior probabilities. Further, regardless of training, they seemed inclined to categorize all the objects as blickets, even the ambiguous objects – that is, in this new condition, unlike the previous experiments, they rarely demonstrated backwards blocking.

The first goal of this experiment was to examine the predictions of causal learning mechanisms other than the Rescorla-Wagner (1972) equation, which were designed to account for the standard backwards blocking paradigm. The data from 4-year-olds in the present experiment go beyond the scope of the predictions of these models. As far as we know, no associative model or parameter estimation model takes into account a measure of the prior probability of the outcome occurring, and uses that information to disambiguate observed data.

The second goal of Experiment 4 was to test the predictions of a particular Bayesian structure learning account of children's causal inferences. Four-year-olds' responses – but not those of younger children – were in line with all of the predictions of this account.

Four-year-olds categorized the object that unambiguously activated the detector as a blicket, regardless of whether blickets were rare or common. Four-year-olds were less likely to categorize object B as a blicket than object A in both conditions, but this difference was significant greater in the rare condition. Finally, in the baseline trial, 4-year-olds were more inclined to categorize objects as blickets in the common condition than the rare condition – but this difference did not account for the difference in the backwards blocking trial.

In addition to what it means for mechanisms of causal learning, 4-year-olds' sensitivity to base rates is interesting in light of recent research on children's ability to reason probabilistically. Classical research suggested that even adults have trouble with probabilistic reasoning, and often have great difficulty using base rates in their decision-making (see e.g., Kahneman & Tversky, 1973). Previous researchers have also suggested that it is only around 6-years-old that children can recognize probabilities and use that understanding to evaluate gambles (Acredolo, O'Connor, Banks, & Horobin, 1989; Piaget & Inhelder, 1975; Schlottman, 2000). However, this research did not examine whether children can implicitly use prior probability information in their categorization and causal learning. Children might be able to do this even if they were unable to explicitly reason about probabilities or base-rates or use this information to evaluate gambles. The present experiment shows that younger children can, in fact, attend to this information and use it to make inferences about the causal properties of objects (see also Gutheil & Gelman [1997] for similar results on slightly older children).

### General Discussion

Four experiments examined children's abilities to make causal inferences based on indirect or ambiguous evidence. In each experiment, children were introduced to a "blicket

detector” – a machine that activated when certain objects (“blickets”) were placed upon it. In the first experiment, 3-year-olds were shown that two objects together activated the blicket detector, and then that one of those objects by itself did not. They inferred that the other object was causally effective and was “like” an established “blicket”. Further, when asked to elicit the effect, the modal response was to place that object on the machine by itself. Children did this, even though they had never seen this action previously; they did not simply imitate the action they had observed activate the machine.

The second experiment replicated the findings of the first, using a different method. In addition, children were tested on a variation of the “backwards blocking” paradigm. In this task, children saw two objects activate the machine together, and then observed that one of those objects activated the machine by itself. The crucial question was how they would categorize the other object. Like adult participants in previous “backwards blocking” experiments (e.g., Shanks & Dickinson, 1987), 4-year-olds judged that this second object was not causally effective; younger children, in contrast, were unclear about the causal status of this object. A third experiment replicated this “backwards blocking” effect on a new group of 4-year-olds, controlling for a potential methodological problem in Experiment 2.

Finally, Experiment 4 examined the predictions of a particular account of causal inference – Bayesian structure learning – and demonstrated that 4-year-olds, but not younger children, responded in accordance with the predictions of such an account. The older children were sensitive to training about the base rate of blickets, and used that information to respond to a backwards blocking trial. When blickets were rare, children did not categorize the ambiguous object as a blicket; when blickets were common, children did categorize it as a blicket.

The results of these experiments suggest that young children can make inferences about the causal properties of objects, even when they do not directly observe those properties. They can also use those inferences to produce new interventions to elicit events, rather than simply imitating the effective actions of others. This suggests that children are not simply associating actions with effects. In addition, the backwards blocking data show that the Rescorla-Wagner (1972) model cannot account for children's causal inferences. This is true even if we assume that children are not simply responding associatively, but are converting measures of association strength into measures of causal strength (see e.g., Cramer, Weiss, Williams, et al., 2002; Dickinson, 2001). Moreover, the fact that young children, as well as adults (see Shanks, 1985), responded in this way suggest that these "backwards blocking" results are not due to extensive experience or education. Finally, the results of Experiment 4 go beyond the predictions of other associative accounts (see e.g., Dickinson, 2001; Van Hamme & Wasserman, 1994; Wasserman & Berglan, 1998) as well as parameter estimation accounts (Cheng, 1997; Shanks, 1995) of children's causal learning; none of these models take into account the base rate probability of events.

It is, of course, possible that one of these models could be modified, or some other even more complex associative model could be constructed in a way that would account for both the present data and the adult data. However, we believe that the use of a Bayesian structure-learning algorithm is a better account of children's causal inferences. Experiment 4 provides preliminary, but by no means conclusive, evidence for such an account. Further research, should investigate this issue. For instance, the Bayesian account presented by Tenenbaum and Griffiths (in press) suggests that this sensitivity to priors could be used to resolve cases in which no unambiguous data is present. Consider the following experiment: children are trained that the prior probability of blickets is rare. They are then shown three

objects (A, B, and C). A and B activate the detector together, as do A and C. Even though all three objects activate the machine, and there is no direct evidence that A activates the machine by itself, the model would predict that A would be categorized as a blicket with greater probability than B and C, which should be categorized as blickets with equal frequency. Further, this should still be the case if children were shown that A and B activated the detector together 10 times, and then A and C activated the detector together once. These investigations are currently underway in the lab. Preliminary data suggest that children's responses are in line with the Bayesian model.

#### The development of a causal learning mechanism

Three and 4-year-olds responded differently in Experiments 2 and 4. Four-year-olds showed substantial backwards blocking, and considered prior probabilities when making a causal inference; three-year-olds, in contrast, did not. Do 3-year-olds use associative mechanisms while 4-year-olds use Bayesian structure learning? This seems unlikely, as the 3-year-olds in Experiment 2 did engage in some backwards blocking – they were less likely to categorize the A object as a blicket in the backwards blocking than the one-cause tasks, and so did not behave as the RW model would predict.

Gopnik and colleagues (Gopnik & Glymour, 2002; Gopnik et al., in press) suggest that causal graphical models might provide the best representation of children's developing causal knowledge, and that graphical structure-learning algorithms provide models of children's causal learning in general, regardless of age. The present data suggest more specifically that Bayesian structure learning algorithms may provide a particularly apt model of the causal learning that leads to backwards blocking. The data also suggest that 4-year-olds can integrate knowledge of the prior probabilities of events with new information.

We believe that the differences between the older and younger children reflect different capacities to encode, store, and operate on their representations and relevant knowledge about prior probabilities, rather than differences in their fundamental causal representations or learning methods. There is, at least, some indirect evidence for this hypothesis. The intervention data in Gopnik et al. (2001) and in the one-cause trials in Experiments 1 and 2 suggest that even 3-year-old children will generate novel actions based on causal inferences, even when they could imitate an action they had previously seen to be effective. Furthermore, even 3-year-olds made correct screening-off inferences based on very small sample sizes, which also speaks against associative accounts.

The developmental differences in the present data could be due to several information processing factors. First, the Bayesian approach relies on calculating a posterior probability for each hypothesis by an application of Bayes' rule based on the prior probability that each hypothesis is the true causal structure and the observed data. This requires that children encode and store information about prior probabilities long enough to use it for their inferences.

A second factor that could explain the developmental difference is to examine whether young children have the relevant prior knowledge to construct meaningful hypotheses. A strength of the Bayesian approach is its ability to incorporate various pieces of information in constructing priors over existing hypotheses. A crucial piece of information that could be represented in those priors is a concept of a parameterization of the causal models. For example, in Figure 2a, an "or" parameterization is assumed – if at least one blicket is on the machine, the machine will activate. In Experiment 4, we explicitly point this out to children, but it is not clear that the younger children understood this concept. Do 3-year-olds recognize there is something about being a "blicket" that makes the

machine go? Among other factors, either a failure to remember the base-rate information, or an inability to understand the parameterization, could lead the three-year-olds to have difficulty with these tasks.

### Conclusion

These experiments show that young children can engage in various kinds of causal inference that involve indirect and/or ambiguous evidence. Children's inferential abilities could not be predicted by simple associative models. While it is possible that some more complex associative model could explain these results, it seems more likely that a mechanism for causal learning will be found among recent computational models of causal inference, such as the Bayesian structural inference mechanism we described and tested here. Determining the development of this mechanism, and the development of other cognitive abilities relevant to this mechanism, should be a goal of future research.

References

- Acredolo, C., O'Connor, J., Banks, L., & Horobin, K. (1989). Children's ability to make probability estimates: Skills revealed through application of Anderson's functional measurement methodology. *Child Development, 60*, 933-945.
- Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*, 299-352.
- Ahn, W., Gelman, S. A., Amsterlaw, J. A., Hohenstein, J., & Kalish, C. W. (2000). Causal status effect in children's categorization. *Cognition, 76*, 35-43.
- Baillargeon, R., Kotovsky, L., & Needham, A. (1995). The acquisition of physical knowledge in infancy. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 79-116). New York: Clarendon Press/Oxford University Press.
- 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.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press/Bradford Books.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367-405.
- Cheng, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive power. In F. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 227-253). Cambridge, MA: MIT Press
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545-567.

- Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychology Review*, *99*, 365-382.
- 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*, *55B*, 241-266.
- Dickinson, A. (2001). Causal learning: Association versus Computation. *Current Directions in Psychological Science*, *10*, 127-132.
- Gelman, S. A., & Wellman, H. M. (1991). Insides and essence: Early understandings of the non-obvious. *Cognition*, *38*, 213-244.
- Glymour C. (2001). *The Mind's Arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.
- Gopnik, A. (1988). Conceptual and semantic development as theory change: The case of object permanence. *Mind and Language*, *3*, 197-216.
- Gopnik, A. (2000). Explanation as orgasm and the drive for causal understanding: The function, evolution, and phenomenology of the theory-formation system. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 299-324). Cambridge, MA: MIT Press.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (in press). A theory of causal learning in children: Causal maps and Bayes nets. *Psychology Review*.
- Gopnik, A., & Glymour, C. (2002). Causal maps and Bayes nets: A cognitive and computational account of theory-formation. In P. Carruthers & S. Stich (Eds.), *The cognitive basis of science* (pp. 117-132). New York, NY: Cambridge University Press.
- Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

- Gopnik, A., & Nazzi, T. (in press). Words, kinds, and causal powers: A theory theory perspective on early naming and categorization. In D. Rakison & L. Oakes (Eds.), *Early categorization*. Oxford: Oxford University Press
- Gopnik, A., & Sobel, D. M. (2000). Detectingblickets: How young children use information about causal properties in categorization and induction. *Child Development*, *71*, 1205-1222.
- 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.
- Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirschfield & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257-293). New York: Cambridge University Press.
- Gutheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology*, *64*, 159-174.
- Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind-body distinction. *Child Development*, *64*, 1534-1549.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell, & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition, a multidisciplinary debate* (pp. 234-267). Oxford: Clarendon Press.

- Klahr, D. (2000). *Exploring science : the cognition and development of discovery processes*. Cambridge, MA: MIT Press.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674-689.
- Kushnir, T. (2001, April). *Action at a distance: How spatial contiguity affects preschool children's causal categorization*. Poster presented at the 2001 Biennial meeting of the Society for Research in Child Development, Minneapolis, MN.
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25, 265-288.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-298.
- Miller, S. A. (1987). *Developmental research methods*. Pretence-Hall: Englewood Cliffs, N.J.
- Nazzi, T. (2001, April). *Children's causal sorting*. Poster presented at the 2001 Biennial meeting of the Society for Research in Child Development, Minneapolis, MN.
- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychology Review*, 94, 61-73.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufman.
- Pearl, J. (2000). *Causality*. New York: Oxford University Press.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. New York: W. W. Norton.
- Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, 36, 1-16.
- Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: 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.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*, 102-119.
- Schlottman, A. (2000). Children's judgments of gambles: A disordinal violation of utility. *Journal of Behavioral Decision Making, 13*, 77-89.
- Scheines, Spirtes, Glymour & Meek, 1994
- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*, (1, Series No. 194).
- Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology, 37b*, 1-21.
- Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*, 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.
- 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.
- Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search* (Springer Lecture Notes in Statistics). New York: Springer-Verlag.
- Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (Springer Lecture Notes in Statistics, 2<sup>nd</sup> edition, revised). Cambridge, MA: MIT Press.

- Tenenbaum, J. B. (1999). *A Bayesian Framework for Concept Learning*. Doctoral Dissertation, Massachusetts Institute of Technology.
- Tenenbaum, J. B., & Griffiths, T. L. (2001). *Structure learning in human causal induction*. Proceedings of the 13<sup>th</sup> Annual Conference on the Advances in Neural Information Processing Systems.
- Tenenbaum, J. B., & Griffiths, T. L. (in press). *Theory-based causal induction*. Proceedings of the 14<sup>th</sup> Annual Conference on the Advances in Neural Information Processing Systems. see [www-psych.stanford.edu/~jbt/temp/nips-causal02.pdf](http://www-psych.stanford.edu/~jbt/temp/nips-causal02.pdf).
- Tenenbaum, J. B., Griffiths, T., & Steyvers, M. (2002). *A theory of Bayesian causal inference*. Manuscript in preparation, Massachusetts Institute of Technology.
- Tenenbaum, J. B., Sobel, D. M., & Gopnik, A. (2002). *Learning causal structure: Adults and children use Bayesian reasoning to make inferences about ambiguous causal events*. Manuscript in preparation, Massachusetts Institute of Technology.
- Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, 25, 127-151.
- 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 & Physiological Psychology*, 51, 121-138.
- Wellman, H. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.
- White, M. A., & Duker, J. (1973). Suggested standards for children's samples. *American Psychologist*, 28, 700-703.



Figure 1

Two different causal models depicting the relationship between parties, wine drinking, and insomnia

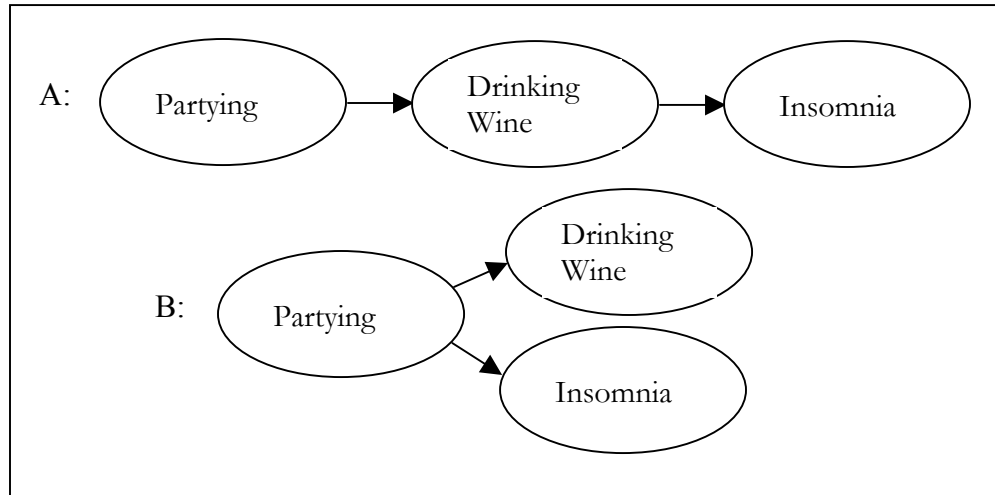


Figure 2:

Two models of the causal structure of the backwards blocking paradigm from Experiments 2 and 3.

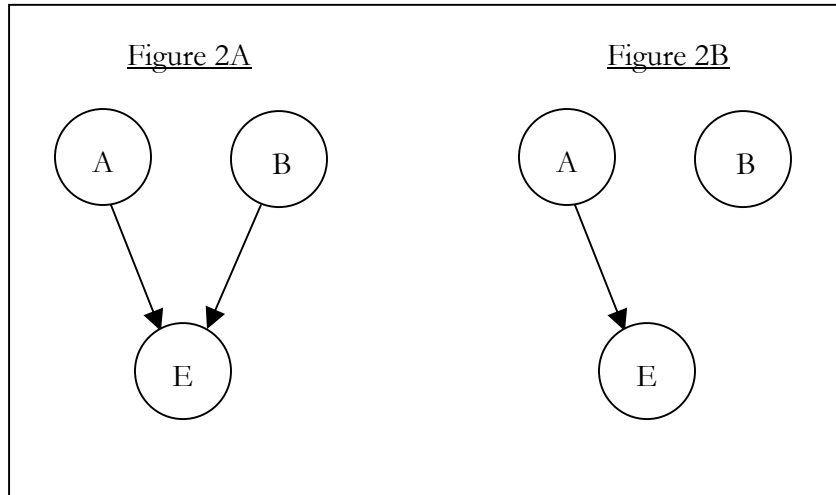


Table 1

Frequency of responses to the “like the blicket” categorization question and “make the machine go” intervention questions Experiment 1 for each task

---

<u>Categorization Question</u>		
<u>Object choice</u>	<u>One-cause</u>	<u>Two-Cause</u>
100% effective object (B/E):	1.63 (81.5)	0.75 (37.5)
66% effective object (A/D):	0.25 (12.5)	0.94 (47.0)
Novel Object (C/F):	0.13 (6.5)	0.31 (15.5)
<u>Intervention Question</u>		
<u>Object choice</u>	<u>One-cause</u>	<u>Two-Cause</u>
100% effective object only (B/E)	1.19 (59)	0.88 (44)
66% effective object only (A/D)	0.25 (13)	0.39 (19)
Novel object only (C/F)	0.25 (13)	0.56 (28)
Both 100% and 66% objects	0.18 (9)	0.18 (9)
Other	0.12 (6)	0.00 (0)

---

Note. Percentages in parentheses. Maximum response is 2.

Table 2

Number of children who chose the 100% causally effective object in response to the “like the blicket” categorization question and “make the machine go” intervention question on the two types of trials in Experiment 1.

---

<u>Categorization Question</u>				
	<u>Two-Cause Trials</u>			
<u>One-Cause Trials</u>	<u>Never</u>	<u>One</u>	<u>Both</u>	<u>Total</u>
Never chose 100% object	0	0	1	1
Chose 100% on 1 of 2 trials	1	2	1	4
Chose 100% on both trials	6	4	1	11
Total	7	6	3	16

<u>Intervention Question</u>				
	<u>Two-Cause Trials</u>			
<u>One-Cause Trials</u>	<u>Never</u>	<u>One</u>	<u>Both</u>	<u>Total</u>
Never chose 100% object alone	1	3	0	4
Chose 100% alone on 1 of 2 trials	1	2	2	5
Chose 100% alone on both trials	3	3	1	7
Total	5	7	3	16

---

Table 3

Frequency of “yes” responses to the “is it a blicket?” question in Experiment 2


---

<u>Age and object</u>	<u>One-cause</u>	<u>Backward blocking</u>
3-year-olds:		
Object demonstrated alone (A/C)	0.25 (0.58)	1.94 (0.25)
Object not demonstrated alone (B/D)	2.00 (0.00)	1.00 (0.82)
4-year-olds:		
Object demonstrated alone (A/C):	0.00 (0.00)	2.00 (0.00)
Object not demonstrated alone (B/D):	2.00 (0.00)	0.25 (0.68)

---

Note. Standard deviation in parentheses. Maximum response is 2.

Table 4

Responses to the “make the machine go” intervention question in Experiment 2 for each task

---

<u>Age and object choice</u>	<u>One-cause</u>	<u>Backward blocking</u>
3-year-olds		
Object demonstrated alone (A/C) only:	81%	53%
Object not demonstrated alone (B/D) only:	6%	30%
Both objects together:	13%	17%
4-year-olds		
Object demonstrated alone (A/C) only:	87%	74%
Object not demonstrated alone (B/D) only:	0%	7%
Both objects together:	13%	19%

---

Table 5

Number of children in each age group who chose the object not demonstrated on the machine by itself (objects B/D) in response to the “is it a blicket” categorization question on the two types of trials in Experiment 2.

3-year-olds

	<u>Backwards blocking Trials (Chose object D)</u>			
	<u>Never</u>	<u>Once</u>	<u>Both</u>	<u>Total</u>
<u>One-Cause Trials</u>				
Never chose object B	0	0	0	0
Chose B on 1 of 2 trials	0	0	0	0
Chose B on both trials	5	6	5	16
Total	5	6	5	16

4-year-olds

	<u>Backwards blocking Trials (Chose object D)</u>			
	<u>Never</u>	<u>Once</u>	<u>Both</u>	<u>Total</u>
<u>One-Cause Trials</u>				
Never chose object B	0	0	0	0
Chose B on 1 of 2 trials	0	0	0	0
Chose B on both trials	14	0	2	16
Total	14	0	2	16

Table 6

Frequency of “yes” responses to the “is it a blicket?” question in Experiment 3


---

<u>Object choice</u>	<u>One-cause</u>	<u>Backward blocking</u>
Object demonstrated alone (A/C)	0.00 (0.00)	2.00 (0.00)
Object not demonstrated alone (B/D):	1.88 (0.17)	0.69 (0.44)
<u>Object choice</u>	<u>Association</u>	
Object with more associative strength (F)	2.00 (0.00)	
Object with less associative strength (E)	1.88 (0.17)	

---

Note. Standard deviation in parentheses. Maximum response is 2.

Table 7

Probability that children placed the B block in the backwards blocking trial and either block in the base rate trial in the “blicket” box as a function of trial.

---

3-year-olds

	<u>Backwards Blocking Trial</u>	<u>Baseline Trial</u>	<u>Difference</u>
Rare Condition	0.81 (0.40)	0.94 (0.17)	-0.13
Common Condition	0.88 (0.34)	1.00 (0.00)	-0.13

	<u>Backwards Blocking Trial</u>	<u>Baseline Trial</u>	<u>Difference</u>
Rare Condition	0.25 (0.45)	0.78 (0.26)	-0.53
Common Condition	0.81 (0.40)	0.97 (0.13)	-0.16

---

Notes. Standard deviation in parentheses

## Endnotes

---

<sup>1</sup> While the percentage of our overall sample who failed controls was relatively high on this experiment (24%), it was not significantly different from the 19% of similar aged children who failed controls on previous research using the blicket detector (Gopnik, Sobel, Schulz, & Glymour, 2001). Further, in an investigation of how child development studies are reported, White and Duker (1973, cited in Miller, 1987) showed that when a study on child participants excluded subjects, 70% of those studies exclude 25% or fewer of their overall sample. We take this as a standard for research, and believe that this experiment falls within that standard.

<sup>2</sup> It is not clear, however, that this analysis is correct. There were three objects, thus setting chance levels at 33% seems appropriate. However, one of those objects was not manipulated. A second analysis was done excluding children's choices of the novel object. This analysis revealed that children's choices between the 100% and 66% objects differed from a chance level of 50% in the one-cause condition  $\chi^2(2) = 16.50, p < .001$ , but not in the two-cause condition  $\chi^2(2) = 3.00, ns$ .

<sup>3</sup> Because the number of ties between the groups eliminated enough subjects to bring  $N < 25$  (see Cohen, 1992), these Sign tests used the Binomial distribution, and thus no Z-statistic was generated.

<sup>4</sup> It is possible that these two objects had the same associative strength – if one trial was sufficient to condition to asymptote. However, few models ever make such a parameter setting. Further, even if this were the case, this condition would still resolve whether children interpreted the “is it a blicket” question as a forced choice.