Running head: CHILDREN'S CAUSAL INFERENCES

Children's causal inferences from indirect evidence:

Backwards blocking and Bayesian reasoning in preschoolers

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

Cognitive Science in press.

Abstract

Previous research suggests that children can infer causal relations from patterns of events. However, what appear to be cases of causal inference may simply reduce to children recognizing relevant associations among events, and responding based on those associations. To examine this claim, in Experiments 1 and 2, 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. Critically, associative models either made no predictions, or made incorrect predictions about these inferences. In general, children were able to make these inferences, but some developmental differences between 3- and 4-year-olds were 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 3 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.

Children's causal inferences from indirect evidence:
Backwards blocking and Bayesian reasoning in preschoolers

As adults, we know a remarkable amount about the causal structure of the world. Much of this causal knowledge is acquired at a relatively early age. Before their fifth birthday, children understand complex causal relations in folk physics (e.g., Baillargeon, Kotovsky, & Needham, 1995; Bullock, Gelman, & Baillargeon, 1982), folk biology (e.g., Gelman & Welman, 1991; Inagaki & Hatano, 1993), and folk psychology (e.g., Gopnik & Wellman, 1994; Perner, 1991; Wellman, 1990). This research has demonstrated that young children know a great deal about causality and that their causal knowledge changes with age. However, this research has not fully explained how 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. These mechanisms might be involved in the causal learning we see in the development of everyday physical, psychological, and biological knowledge.

We propose that children must draw on two sources of knowledge to make causal inferences. First, children apply their existing knowledge of causal mechanisms to observed data. 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 causality (e.g., that causes precede effects – a piece of knowledge that potentially discriminates which events are causes and which events are effects). Some of this knowledge may be innate (such as certain aspects of physical causality, see Spelke et al., 1992); other pieces of knowledge may be acquired throughout development.

There is 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) generative 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 cause spots of light to appear on the wall). Children relied on this prior physical knowledge to make new causal inferences.

Children must also rely on a second source of knowledge when making causal inferences: a mechanism that infers the causal structure of events from data. How might this mechanism work? The literature on causal learning in adults proposes a number of specific models that might underlie this sort of learning (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 children use such mechanisms to acquire causal knowledge. The goal of this paper is to discriminate among these possible mechanisms. By doing so, we potentially provide an account of children’s causal learning and their representation of causal knowledge.

This project, however, assumes that young children can accurately relate patterns of data to causal structure. Some investigations have suggested that children have difficulty understanding how patterns of data relate to causal relations (Klahr, 2000; Kuhn, 1989; Schauble, 1990, 1996). In these experiments, children were required to construct experimental manipulations or state what kinds of evidence would be necessarily 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 scientific reasoning studies suggest that young children have difficulty explicitly stating what evidence would be required to test particular hypotheses. However, young children might still be able to draw accurate causal conclusions from patterns of evidence. Children might be able to recognize causal relations from data without having the metacognitive capacity to recognize that this particular piece of data was necessary to make a conclusion among several causal hypotheses. Likewise, children might be able to draw accurate conclusions from unconfounded experiments, even if they are unable to design those experiments in the first place. The learning mechanism 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 children might interpret data that they observe.

Classes of Causal Learning Mechanisms

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 mechanisms 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. These models assume that candidate causes and effects have been identified, typically based on substantive information such as temporal priority, and output the strength of each cause-effect association. Since these models only output associative strengths, they make no predictions about how people can use causal knowledge to generate interventions to elicit effects.

2) Inferring causal relations from associations. In generalizing the associative approach outlined in #1 above, 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, advocates of this approach have suggested alternative mechanisms for calculating causal strength (e.g., Krushke & Blair, 2000; Van Hamme & Wasserman, 1994; Wasserman & Berglan, 1998). Because of the transformation of associative strength to causal strength, these alternative associative mechanisms do 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. Two prominent proposals in this category are the □P model (Allan, 1980; Jenkins & Ward, 1965; Shanks, 1995) and the Power PC model (Cheng, 1997, 2000). These models calculate an estimate of the maximal likelihood value of the strength of a presumed causal relationship given a set of data (Glymour, 2001; Tenenbaum & Griffiths, 2001). This distinguishes these models from those described in #2 above; they estimate the strength of a particular causal model – the probability that an effect occurs given a cause and some background information. The models in #2 only estimate these strength parameters weakly or at asymptote. For instance, under certain conditions, Rescorla-Wagner can be shown to converge to □P in the limit of infinitely many, randomly intermixed trials (Cheng, 1997; Danks, in press; Shanks; 1995).

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, Glymour, Sobel, Schulz, Kushnir, & Danks, in press; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; Tenenbaum & Griffiths, 2003). Mathematical theories of 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. We will not review the entire literature on causal learning using this framework in this paper (see Gopnik et al., in press, for one such review). Instead, we will focus on a specific graph-learning algorithm that relies on Bayesian inference and particular pieces of substantive causal knowledge (Tenenbaum & Griffiths, 2003).

Research on Children's Causal Learning

We believe that a particular structure-learning algorithm from the fourth class of models listed above best accounts for at least some kinds of causal learning in children. The present experiments are motivated by predictions of this algorithm. In order to motivate this particular algorithm, we must first describe the basic reasoning behind structure-learning algorithms, and state how they differ from the other three classes of models outlined above. Further, we must consider whether previous work on causal learning is consistent with the prediction of this model and other alternatives.

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 potential 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 (a substantive assumption you bring to the learning process) 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.

Thus, it is necessary 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, 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 these events, you can infer which of the two causal structures shown in Figures 1a and 1b is more 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; it 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 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. Children observed that object A activated the machine and that object B did not. The objects were then placed on the machine together twice, activating it both times. Children were then asked whether each object was a blicket.

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. Again, children were asked whether each object was a blicket.

In both conditions, object A was associated with the machine's activation 100% of the time and object B was associated with the machine's activation 66% of the time. The difference between the conditions was that in the *one-cause* condition, the pattern of dependent and independent relations between the two 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 on the detector. If children use screening-off reasoning, then they should categorize only object A as a blicket in the *one-cause* condition. However, in the *two-cause* condition, they should categorize both objects as blickets; in this condition, both objects independently activated the machine – one just does so with greater frequency. Three- and 4-year-olds, and in a slightly modified version of this experiment, 30-month-olds, behaved in this manner. Gopnik et al. (2001) concluded that children could use the pattern of independence and conditional independence among events to determine screening-off relations. To do this, they claimed that children engage in the process of learning causal structure from observation of events.

An Objection: Associative Models Explain 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 – so that now both objects were on the machine. Children were asked to make the machine stop, an intervention 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 children simply recognize associations (model class #1 described above), 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 inferences, and to further discriminate among potential causal learning mechanisms. In Experiments 1 and 2, we examined how children made inferences about the effects of an object based on different pieces of indirect evidence. These experiments were designed to suggest that a simple version of the Rescorla-Wagner (1972) model was a problematic account of children's causal inference. In Experiment 3, we present a particular mechanism of causal inference, based on Bayesian reasoning and substantive causal knowledge. Experiment 3 is motivated by the predictions of this mechanism.

Experiment 1

Experiment 1 investigated children's ability to make two types of indirect inferences. The first is an indirect screening-off inference, which uses a variant of the Gopnik et al. (2001) procedure. 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 Gopnik et al.'s (2001) 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). Asking children to make inferences about data they did not directly observe addresses this concern.

We also presented children with an inference problem based on findings from the adult reasoning data: *backwards blocking*. In studies on backwards blocking, adult participants observed an outcome occurring in the presence of two potential causes (A and B). Participants then observed that event A independently causes the outcome. Participants were less likely to judge event B as a cause of the outcome than those who only observe A and B cause the outcome together (Shanks, 1985; Shanks & Dickinson, 1987). Associative models, such as the Rescorla-Wagner (1972) equation, have difficulty explaining these data.

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.

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 caused the machine to activate.

Sixteen wooden blocks, different in shape and color, were also used. The blocks were divided into pairs. No pair of blocks was the same color or shape. 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.

Procedure. All children were tested by a male experimenter with whom they were familiar. Children were brought into a private game room 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 1), 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”. Children 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.” These demonstrations were performed twice. 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. One activated the machine, and the other did not. 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 questions during these training trials incorrectly. The 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 conditions. In the *indirect screening-off* condition, two objects (A and B) were put on the table. The two objects were placed on the machine together, and 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". Children were given two trials of this condition.

The *backwards blocking* condition were identical to the indirect screening-off tasks, except that when the individual object was placed on the blicket detector by itself, the detector did activate. Thus, two new objects (A and B) were placed on the detector together and it activated. This was demonstrated twice. Then, object A was put on the detector alone and it did activate. Children were then asked whether each object was a "blicket". After these questions, children were asked to "make the machine go". Children were also given two trials of this condition.

A straightforward interpretation of the Rescorla-Wagner model states that the associative strength of object B in the backwards blocking condition is the same as the associative strength of object B in the indirect screening-off 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. If children are using a calculation of associative strength to make their causal inferences, then they should be as likely to claim that object B in the backwards blocking condition has causal efficacy (and hence, is a blicket) as object B in the indirect screening-off condition. However, if children engage in backwards blocking, similar to adult participants in previous experiments, then they should not respond in this manner.

Finally, children were given one trial of a *control* condition. In this condition, children were shown two objects. Each was demonstrated individually on the detector. One activated it; the other did not. Children were asked whether each was a blicket. To be included in the analysis, children must respond that the object that activated the machine was a blicket and the object that did not activate 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. This was also done to ensure children were paying attention for the entire session. None of the 4-year-olds and two of the 3-year-olds were excluded from the data analysis for this reason.

In both the indirect screening-off and backwards blocking conditions, the object that was independently placed on the machine was counterbalanced for spatial location across the two tasks (on one task, it was the object on the left, on the other it was the object on the right). Across both conditions, children were first asked whether the object that was placed on the detector alone was a blicket, and then they were asked whether the other object was a blicket. Different pairs of blocks were used for each of the five tasks. The five tasks were presented in a randomly generated order with the provision that the control task was never first. Thus, in the analyses below, condition is a within-subject factor, and age is a between-subject factor.

Results

To examine whether there was differences in responses within a testing session based on order, McNemar's χ^2 tests were performed on responses to the two indirect screening-off tasks and two backwards blocking tasks. These tests revealed no significant differences between the responses on either the "is it a blicket" question or "make the machine go" question on the two tasks in either condition. Thus, the data from the two tasks were combined in both conditions. Further, initial analyses revealed no effect of order on performance: children who received a indirect screening-off trial first performed in the same manner in both conditions 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. Table 1 shows responses to the "is it a blicket" categorization question for both types of trials. Likewise, Table 2 shows responses to the "make the machine go" intervention question.

Indirect screening-off condition. On the indirect screening-off tasks, children observed that objects A and B together made the machine activate, and then that object A alone did not. In response to the "is it a blicket" question, children responded that object A was a blicket on an average of 0.13 out of 2 trials (6% of the time), and that object B was a blicket on an average of 2.00 out of 2 trials (100% of the time), a significant difference: $t(31) = 25.18$, $p < .0001$. Similarly, on the "make the machine go" question, children placed the A object on the machine on an average of 0.32 out of 2 trials, while they placed the B object on the machine on average 1.94 out of 2 trials (97% of the time), a significant difference: $t(31) = 11.60$, $p < .001$. No age differences were found in response to either question.

Patterns of individual responses were then analyzed. In response to both the "is it a blicket" and "make the machine go" questions, children could respond in one of four ways. They could have chosen only object A, only object B, both objects A and B, or neither objects A or B. Children never responded that both objects were not blickets, and they always placed at least one object on the machine in response to the intervention question.

In response to the categorization question, 4-year-olds chose only object B at every opportunity. Thirteen of the sixteen 3-year-olds did so as well. Three 3-year-olds chose both objects as blickets on both trials or both objects on one trial and only object B on the other trial. There was a non-significant trend for the older children to respond in this manner more often than the younger children: $\square^2(1) = 3.31$, $p < .07$. Looking at responses over individual trials, 4-year-olds chose only object B on 100% of the trials. Three-year-olds did so on 28 out of 32 trials (87.5% of the time). On the remaining 12.5% of the trials, they chose both objects as blickets. Thus, every child stated that object B was a blicket. The number of times each child said the A object was a blicket was analyzed with a Mann-Whitney test, which revealed a non-significant trend between the 3- and 4-year-olds: $U = 104$, z = -1.79, $p < .08$.

In response to the intervention question, twelve 3-year-olds and thirteen 4-year-olds placed only object B on the machine by itself on both trials, not a significant difference: $\square^2(1) = 0.18$, *ns*. Looking at individual trials, the 4-year-olds placed only the B object on the machine on 28 out of 32 trials (87.5% of the time). On the remaining 12.5% of the trials, they placed both objects on the machine together. The 3-year-olds placed only the B object on the machine on 26 out of 32 trials (81.5% of the time). They placed both objects on the machine on 4 trials (12.5% of the time) and only object A on the machine on 2 trials (6% of the time). Mann-Whitney *U* tests revealed that the number of times children made each of these responses did not differ between the two age groups[1].

Finally, an analysis was done to examine whether children were consistent in their responses to the categorization and intervention questions in each presentation. Children made the same response to the two questions on 58 of the 64 trials, significantly more often than all other responses put together: Binomial test, based on a *z*-approximation, $p < .001$. The most frequent response was to choose only object B as a blicket, and place only object B on the machine to activate it. Across the two ages, children did so on 54 out of the 64 trials (80% of the time), which is significantly greater than all other responses put together: Binomial test, based on a *z*-approximation: $p < .001$. Thus, both age groups were able to infer that object B had the causal efficacy, and were able to make a novel intervention consistent with that inference.

Backwards blocking condition. On the backwards blocking condition, children observed that objects A and B together made the machine activate, and then that object A alone activated the machine. In response to the "is it a blicket" question, children categorized object A as a blicket on an average of 1.97 out of 2 trials. They categorized object B as a blicket on an average of 0.62 out of 2 trials. This was a significant difference: $t(31) = 8.78$, $p < .001$. A significant effect of age was also found in their categorization of object B: the 4-year-olds rarely said object B was a blicket (on an average of 0.25 out of 2 trials), whereas the 3-year-olds were significantly more likely to say so (on an average of 1.00 out of 2 trials): $t(30) = 2.82$, $p < .01$. There was no effect of age on children's categorization of object A. Although this effect of age was present, both age groups categorized object A as a blicket more often than object B: $t(15) = 4.39$, $p < .001$ for the 3-year-olds, and $t(15) = 10.25$, $p < .001$, for the 4-year-olds.

Patterns of individual responses were then analyzed in response to both questions. As in the indirect screening-off condition, children could respond in one of four ways in response to both questions. They could have chosen only object A, only object B, both objects A and B, or neither objects A or B. Children never claimed that both objects were not blickets, nor did they fail to place at least one object on the machine in response to the intervention question.

In the backwards blocking condition, fourteen of the 4-year-olds chose only object A as a blicket on both trials. Only five of the 3-year-olds responded in this manner, a significant difference: □^{2}(2) = 10.49, *p* < .01. Looking at responses on individual trials, the 4-year-olds chose only object A as a blicket on 28 of the 32 trials (87.5% of the time). On the remaining 12.5% of the trials, they categorized both objects as blickets. The 3-year-olds categorized only object A as a blicket on 16 out of 32 trials (50% of the time). They categorized both objects as blickets on 15 out of 32 trials (47% of the time), and object B as the only blicket once (3%). Mann-Whitney tests revealed that 3- and 4-year-olds did not differ in their categorization of object A – both age groups often claimed it was a blicket: *U* = 120, *z* = -1.00, *ns*. Three- and 4-year-olds differed in the number of times they categorized object B as a blicket: *U* = 62, *z* = -2.83, p < .01.

On the intervention questions, ten 4-year-olds and seven 3-year-olds placed only the A object on the machine on both trials – not a significant difference: □^{2}(1) = 1.13, *ns*. However, seven 3-year-olds placed object B on the machine by itself on at least one of the two trials, while only two 4-year-olds did so, which was significant: $\chi^2(1) = 3.87$, $p < .05$. This suggests that when asked to make an intervention, the 4-year-olds recognize that placing object A or both A and B on the machine together was efficacious, while the 3-year-olds were more likely to believe that object B would activate the machine by itself. Looking at individual trials, the 4-year-olds placed only the A object on the machine on 23 out of 32 trials (72% of the time). On seven trials, they placed both objects on the machine together (22% of the time), and they placed only object B on the machine on the remaining two trials (6% of the time). The 3-year-olds placed only the A object on the machine on 17 out of 32 trials (53% of the time). On five trials, they placed both objects on the machine together (15.5% of the time), and they placed only object B on the machine on the remaining ten trials (31.5% of the time). A Mann-Whitney *U* test revealed that the 3-year-olds were more likely to place only object B on the detector than the 4-year-olds: $U = 85$, $z = -2.06$, $p < .05$. Similar Mann-Whitney tests showed no difference in the other two response types[2].

Were children consistent in their answers to the two questions? If children engaged in backwards blocking, and categorize object A, but not object B as a blicket, they should not place object B on the machine in response to the intervention question. Children categorized only object A as a blicket on 44 out of the 64 trials (69% of the time). When they did so, they placed only object A on the machine in response to the intervention question on 34 of those trials (77% of the time). A chi-squared goodness of fit test on these 44 trials indicated that the distribution of responses was different from what would be expected by chance: $\chi^2(2) = 38.77$, $p < .001$. Even when 3-year-olds' and 4-year-olds' responses were considered separately, these findings held. On the 16 trials in which the 3-year-olds categorized only object A as a blicket, they placed only object A on the machine on the intervention question on 12 of those 16 trials (75% of the time). Of the 28 trials in which the 4-year-olds categorized only object A as a blicket, they placed only object A on the machine on the intervention question on 22 of those 28 trials (79% of the time). Chi-squared goodness of fit tests revealed that both distributions of responses differed from what would be expected by chance: $\chi^2(2) = 12.50$ and 26.64 for the 3- and 4-year-olds respectively, both *p*-values < .01.

Thus, in the backwards blocking condition, once children observed that one object activated the machine on its own, they often categorized that object as the only cause. This was especially true of the 4-year-olds. 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, 2003). 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 B in the backwards blocking trials at the level of chance and not below chance: $\chi^2(2) = 1.00$, *ns*.

Comparison between the indirect screening-off condition and backwards blocking conditions. The Rescorla-Wagner (1972) model predicts that the associative strength of object B is the same across the indirect screening-off and backwards blocking conditions. Did children treat this object the same way across the two conditions? Overall, children categorized object B as a blicket in the backwards blocking trials less frequently than they categorized object B a blicket in the indirect screening-off condition: $t(31) = 9.34$, $p < .001$. This was true even when only the 3-year-olds were considered: $t(15) = 4.90$, $p < .001$. This suggests that the children were not making causal inferences simply based on associative models, such as the Rescorla-Wagner (1972) equation. Rather, these data suggest that children's causal inferences are better explained by a different mechanism.

Individual responses were analyzed in a nonparametric fashion to supplement the parametric analyses above. Table 3 shows the distribution of responses on the categorization question between the two conditions. In particular, if children made a screening-off inference, then on the indirect screening-off tasks, they should choose only object B as a blicket on both trials. If children made a backwards blocking inference, then on the backwards blocking tasks, they should choose only object A as a blicket on both trials. Four of the 16 three-year-olds and 14 of the 16 four-year-olds responded in exactly this manner. A McNemar's chi-squared analysis revealed that children were more likely to respond in this manner on the indirect screening-off tasks than the backwards blocking tasks: $\chi^2(1) = 6.75$, $p < .01$. However, when this analysis was considered across the two age groups, this difference was only true of the 3-year-olds: $\chi^2(1) = 4.90$, $p < .05$. The older children were equally likely to make an indirect screening-off inference as they were to make a backwards blocking inference.

Discussion

Experiment 1 had several goals. The first was to examine whether young children could make a screening-off inference based on indirect evidence. In the indirect screening-off condition, 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. Further, children's interventions paralleled their inferences. 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 when asked to make the machine go. 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.

A second goal of Experiment 1 was to see how children responded to a backwards blocking paradigm. In this condition, children were shown two objects that activated the machine together. Then, one of those 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. Four-year-olds 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.

Responses differed across the two conditions. Both 3- and 4-year-olds categorized the object not demonstrated on the detector by itself differently. In the indirect screening-off condition, they often claimed it was a blicket. In the backwards blocking condition, they were less likely to do so. This suggests that they were not simply using the output of the Rescorla-Wagner (1972) model as an indicator of a causal relation, as this model posits that these two objects have the same associative strength.

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 every trial in Experiment 1, children were 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?" categorization 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. On the indirect screening-off trials, because object A is clearly not a blicket, object B must be. Likewise, on the backwards blocking trials, because object A unambiguously activated the machine, and clearly was a blicket, under a forced choice interpretation, object B must not be a blicket. Experiment 2 explored this possibility.

Experiment 2

Experiment 2 replicated Experiment 1 with several modifications to address the methodological concern above. First, during the training trials, 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 indirect screening-off and backwards blocking conditions were similar to Experiment 1; 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. If they did not interpret this question as asking them to choose between the two objects, 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 also 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 1, only a 4-year-old sample was considered here. Since the younger children did not unambiguously demonstrate backwards blocking, we felt it was only necessary to examine an older sample to address this methodological concern.

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.

Materials. The same "blicket detector" as in Experiments 1 was used. Twenty wooden blocks, different in shape and color, and assembled in the same manner as in Experiment 1, were also used. The metallic knobs and tee-joints from Experiments 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 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 conditions. The *indirect screening-off* and *backwards blocking* conditions were identical to those in Experiment 1, except that children were not asked to make the machine go after they categorized each object. Children received two trials of each of these conditions, counterbalanced for spatial location. In addition, children received two trials of an *association* condition. In this condition, two objects (A and B) were placed in front of the child. Object A was placed on the detector by itself once, activating it. Object B 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 B as a blicket. If children were not interpreting the procedure this way, they should categorize both objects as blickets.[3]

Finally, children were given a *control* condition, 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. Thus in the analyses below, condition is a within-subject factor, and age is a between-subject factor.

Results

Initial McNemar's chi-squared tests revealed no difference between responses on the two indirect screening-off, two backwards blocking, and two association tasks. An analysis similar to that performed in Experiment 1 revealed no effects of order. Table 4 shows responses to the "is it a blicket?" question for the three types of tasks.

In the indirect screening-off condition, children never categorized object A as a blicket. They categorized object B as a blicket on an average of 1.88 out of 2 trials. This is a significant difference: $t(15) = 21.96$, $p < .001$. This replicates the findings of the previous experiment: children inferred that an object possessed the causal property even when they did not directly observe its influence. Examining individual performance, 14 of the 16 children categorized object B as the only blicket on both trials, and children responded in this manner on 30 of the 32 total trials.

In the backwards blocking condition, children always categorized object A as a blicket. They categorized object B was a blicket on an average of 0.69 out of 2 trials. This is a significant difference: $t(15) = 6.01$, $p < .001$. This replicates the findings of Experiment 1: once 4-year-olds observe an object that independently activates the machine, they do not postulate the presence of another cause. Examining individual differences, nine of the 16 children categorized object A as the only blicket on both trials, and overall children categorized only object A as a blicket on 21 out of the 32 total trials. On the remaining 11 trials, children categorized both objects as blickets. A chi-squared goodness of fit test revealed that this distribution of responses differed from what we would expect if children were simply answering these questions at chance: $\square^2(2) = 20.71$, $p < .001$.

Responses to object B in the indirect screening-off and backwards blocking tasks clearly differed. Even though these two objects had the same level of associative strength, children categorized object B as a blicket more often in the indirect screening-off condition than in the backwards blocking condition: 94% vs. 34% of the time, respectively, $t(15) = -4.54$, $p < .001$. Of the sixteen children in the experiment, nine of them chose only object A as a blicket in the two backwards blocking tasks and only object B as a blicket in the two indirect screening-off tasks. Overall, more children chose only object B as a blicket in the two indirect screening-off tasks than chose only object A in the two backwards blocking tasks, but this was a non-significant trend: McNemar's $\square^2(1) = 3.20$, $p < .10$.

Finally, children showed no difference in their categorization of the two objects in the association condition. Overall, children categorized the object that activated the machine once as a blicket on 1.88 out 2 trials, and the object that activated the machine twice on 2.00 out of 2 trials. This was not a significant difference: $t(15) = 1.46$, *ns*. When children were shown two objects that each unambiguously activated the blicket detector, they categorized both as "blickets". Looking at individual responses, children categorized both objects as blickets on 30 of the 32 trials. This suggests that children did not interpret "is it a blicket?" as a forced choice question.

Discussion

The goal of Experiment 2 was to replicate Experiment 1 while controlling for the possibility that children interpreted the "is it a blicket?" question as a forced choice. Responses to the indirect screening-off and backwards blocking conditions in this experiment paralleled those of Experiment 1. Further, when children were presented with cases in which both objects clearly activated the machine independently, children categorized both as blickets. Children did not respond that there was only one blicket on each trial. This suggests that they did not interpret the test question as asking for a forced-choice response.

A Structure-Learning Account: Bayesian Inference and Substantive Knowledge about Causal Mechanisms

The results of Experiments 1 and 2 suggest that children's causal inferences cannot be explained by certain models that rely solely 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.

These data shed doubt on the possibility that children are simply recognizing associations (model class #1 described above). These data also shed doubt on the possibility that children are using some calculation of associative strength based on the Rescorla-Wagner (1972) equation to make causal inference (a subset of model class #2). However, these experiments do not discriminate between other contemporary mechanisms described by #2, parameter estimation models (#3), and structure learning algorithms (#4).

Tenenbaum and Griffiths (2003) have offered an account of both the screening off data from Gopnik et al. (2001) and the current backwards blocking data. Below, we describe this account and an experimental test of one of its predictions. One advantage of this approach is that, to our knowledge, no other model of causal learning from classes #2, #3, or #4 can naturally explain these new data.

The model is motivated by an observation from the present and previous blicket detector experiments. In these experiments, children make inferences based on observing a small number of trials involving each object. This in itself is a problem for associative models and parameter estimation models. Associative models typically assume many more observations of data than the number that occur in these experiments. For instance, any model based on the modification of the Rescorla-Wagner equation (models in class #2, such as Wasserman & Berglan's [1998] model) does not allow a single positive association to condition a stimulus to asymptote. In the backwards blocking condition, these models do not predict that object A *must be* a blicket based on the one time it activated the blicket detector independently. Parameter estimation models (class #3) suffer from related problems because they also rely on frequency data to estimate the relevant probabilities. In their basic version, both the □P and Power PC models produce the asymptotic value of object A, but an undefined value for object B (see Glymour, 2001). One advantage of appealing to certain causal graph structure-learning accounts (class #4) is that they allow children to integrate information about the prior probability of particular kinds of causal relations with the currently observed data. This machinery supports inferences about causal structure based on very limited observations – including situations where the direct efficacy of an object is not observed, such as occurs with object B in both the indirect screening off and backwards blocking conditions.

As an example, Figure 2a and Figure 2b depict two potential causal structures consistent with the data in the backwards blocking trial. Associative models and parameter estimation models begin by assuming a fixed causal structure – namely Figure 2a – in which either object being placed on the machine (A and B) is a potential cause of the machine activating (E). According to these models, 4-year-olds' responses in Experiments 1 and 2 indicate that their calculation of the strength of the link between B and E is near zero, or below a threshold level at which objects are labeled "blickets".

Graphical structure-learning algorithms process 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 represented in the graph. On this view, the 4-year-olds in

Experiments 1 and 2 recognized that the more likely causal structure given the observed data was Figure 2b, and categorized and intervened in that manner.

There are several types of graph structure-learning algorithms (see Gopnik et al., in press, for a review). The model we would like to propose – based on Bayesian inference – starts with a set of hypotheses that could have generated the observed data. Figure 2a and 2b represent two such hypotheses. 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). However, in order for Bayesian learning to be effective given such a small set of observations, one must specifically confine the initial hypothesis space. Tenenbaum and Griffiths (2003) describe three assumptions children might use to constrain the initial hypothesis space.

Assumption 1: Temporal priority. An object's position may affect the detector activating. However, the detector's activating does not affect an object's position. Put simply, objects being placed on the machine can cause the machine to activate, but the machine's activation does not cause the experimenter to place an object on the machine. This assumption constrains the hypothesis space by eliminating any hypothesis in which the detector's activation is the cause of an objects' position (e.g., E → A or E → B).

Assumption 2: Object positions are independent. An object being placed on the detector does not cause another object to be placed on or removed from the detector. This assumption constrains the hypothesis space by eliminating any hypothesis in which any object's position causes any other object's position (e.g., A → B or B → A).

Assumption 3: The "Activation Law". The blicket detector itself constrains the models to a particular parameterization: the detector operates if at least one blicket is placed upon it. Unlike the previous two assumptions, which constrain what types of hypotheses are considered, this assumption constrains how all the hypotheses are parameterized. In particular, this assumption states that each cause, if present, is fully sufficient on its own to produce the effect. Thus, the detector will activate if object A is on the detector and the hypothesis specifies A → E or object B is on the detector and the hypothesis specifies B → E (or both). This activation law could be generalized to allow for a small level of noise (e.g., situations in which occasionally the detector fails to activate when a blicket is placed on it), but we consider only the simpler deterministic case here. In the previous experiments, this assumption is consistent with the instructions given to the children (i.e., "blickets make the machine go"). In addition, this assumption is consistent with our intuitions about – and children's experience with – how most machines work. In Experiment 3, however, we presented children with explicit instructions about this activation law.

Given these assumptions, Bayesian updating then licenses the following inferences about the backwards blocking paradigm. After observing that objects A and B activate the machine together, the posterior probability that A is a blicket increases above its prior probability, and similarly for B. However, after observing that object A alone activates the detector, these quantities diverge. The probability that object A activates the machine (and hence, is a blicket) becomes 1, because otherwise the event could never be observed. The probability that B activates the machine returns to its baseline prior probability, because knowing that A surely causes the machine to activate makes the compound case, in retrospect, uninformative about the causal status of B. In Experiments 1 and 2, the assumption is that this base rate 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.

This reliance on baseline prior probabilities motivates a prediction about performance on a backwards blocking trial: if the prior probability that an object is a blicket is high enough and children recognize this high probability, then the observed data should not be sufficient to justify a strong backwards blocking response. Likewise, if the prior is low enough and children recognize this low probability, then backwards blocking should be strong. In Experiment 3, we manipulated the prior probability that objects activated the detector (and hence, were blickets) to test whether this kind of algorithm can explain children's causal learning. The Bayesian account outlined above predicts a significant difference based on this manipulation. Standard associative models or parameter estimation models do not: regardless of whether blickets are rare or common, these models carry out the same computations for both objects in the backwards blocking paradigm.

Experiment 3

In Experiment 3, 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 the Bayesian model outlined above, if blickets are rare, then children should reason as in Experiments 1 and 2, and infer that the uncertain object is not a blicket. This is because, given the observed data, a model with only one causal relation should have a higher posterior probability than a model with two causal relations. However, if blickets are common, then participants should reason that the uncertain object is in fact 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, this 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. Recall that this object unambiguously activates the detector independently.
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. Its 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 individually. 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 from the table. 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 new 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 (object A) in the blicket box. Thus, regardless of whether blickets were rare or common, children's categorization of the unambiguous object was at ceiling levels – it was always placed in the blicket box. Table 5 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.

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*(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*(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: □^{2}(1) = 10.17, *p* < .001. Second, a Mann-Whitney *U* test 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 only 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%, $t(15) = -2.63$, $p < .05$. 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%, $t(15) = -2.87$, $p < .01$. 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 used this information to guide their inferences about the uncertain information in a backwards blocking condition. Four-year-olds seemed capable of using this information: when blickets were rare, they categorized an object whose causal property was uncertain as not a blicket; when blickets were common, they categorized the same object as a blicket. 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 object. In this new paradigm, unlike Experiment 1, they rarely demonstrated backwards blocking.

One 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. Each model categorizes the causal strength of an individual object, based on its association or co-occurrence with the effect. None of these models takes into account the difference in the overall number of causes that are present. Thus, none of these models would predict the different in responses observed by the 4-year-olds.

A further goal of Experiment 3 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 significantly 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. Some research has 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 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 related results on slightly older children).

General Discussion

Three experiments examined children's abilities to make causal inferences based on indirect 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- and 4-year-olds were shown two indirect inference problems. In the first, they were shown that two objects activated the blicket detector together, and then that one of those objects did not activate the machine by itself. Children inferred that the other object was causally efficacious and labeled it a "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.

In the second inference problem, children were tested on a "backwards blocking" paradigm. In this task, children observed that two objects activate the machine together, and then that one of those objects did activate the machine by itself. This object was always categorized as a blicket. The critical question was how children would categorize the other object. Four-year-olds judged that this second object was not causally efficacious. Younger children, in contrast, were unclear about the causal status of this object. A second experiment replicated this backwards blocking effect on a new group of 4-year-olds, controlling for a potential methodological problem in Experiment 1. These findings demonstrate that retrospective reevaluations in causal inference can occur not only in a standard backwards blocking paradigm (Shanks & Dickinson, 1987), but also with many fewer trials and with much younger learners.

Several classes of causal learning mechanisms may account for these data. We propose a model of Bayesian structure learning that uses substantive prior knowledge to confine and parameterize the initial hypothesis space. In Experiment 3, we examined a prediction of this model: when children observe the backwards blocking data, Bayesian inference allows them to reason that the probability of the ambiguous object (object B) being a blicket is a function only of the baseline probability of any object being a blicket. If that baseline is high, and children recognize that probability, then they should categorize the object as a blicket. If that baseline is low, then they should not categorize the object as a blicket. We demonstrated that 4-year-olds, but not younger children, responded in accordance with the predictions of such an account.

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 causal outcomes, 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., Dickinson, 2001). Moreover, the fact that young children responded in this way suggests that these backwards blocking results are not due to extensive experience or education. Finally, the results of Experiment 3 go beyond the predictions of other associative accounts (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 takes into account the base rate of causal outcomes. In Experiment 3, four-year-olds were quite sensitive to this information.

It is 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 the present data. However, we believe that the use of a Bayesian structure-learning algorithm is a better account of children's causal inferences. Experiment 3 provides preliminary, but by no means conclusive, evidence for such an account. Further research, should investigate this issue. For instance, Tenenbaum, Sobel, and Gopnik (2003) have replicated the rare-common backwards blocking manipulation on adult learners. By using adults, who provide multiple ratings of each object's causal efficacy (not just whether each is a blicket at the end of the trial), we can test some of the quantitative predictions of the Bayesian model. Indeed, preliminary data suggests that 4-year-olds and adults seem to make similar inferences.

The Bayesian account suggests that this sensitivity to priors could also be used to resolve cases in which only ambiguous data is present. Consider another follow-up 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 Bayesian 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.

This program of research may not only help to elucidate the particular types of causal learning we describe here but also to help explain other kinds of conceptual and cognitive change. In particular, causal learning mechanisms may play an important role in the development of everyday theories. Over the last twenty years a number of investigators have argued that children develop theories – coherent, abstract causal models of the world – in a number of domains including psychology, biology and physics (see e.g., Carey, 1985; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1994; Wellman, 1990). Psychologists have charted changes in those theories over development. However, we know very little about how these theories are formed or why they change. Causal knowledge seems to play a central role in theories, both in science and in everyday life. We would speculate that the kinds of causal learning mechanisms we described here might underpin children's ability to learn about many aspects of the world around them. For instance, Schulz and Gopnik (in press) have recently demonstrated that children use "screening-off" reasoning to make inferences about causal relations in psychological and biological domains, as well as the physical causal relations described in the "blicket detector" experiments.

The Development of a Causal Learning Mechanism

Three- and 4-year-olds responded differently in Experiments 1 and 3. Four-year-olds showed substantial backwards blocking, and considered prior probabilities when making a causal inference. Three-year-olds showed a weaker backwards blocking response in Experiment 1, and were insensitive to priors in Experiment 3. What accounts for this difference? Do 3-year-olds use associative or parameter estimation mechanisms and develop a mechanism for Bayesian structure learning by age 4? Or, are all children using Bayesian inference, but with younger children less able to make strong inferences, due to either domain-general (e.g., information processing) limitations or insufficient domain-specific substantive knowledge (to constraint the hypothesis space)? The present data are unable to answer these questions. Three-year-olds in Experiment 1 did engage in some backwards blocking – they were less likely to categorize the object not directly placed on the detector as a blicket in the backwards blocking than the indirect screening-off tasks. Three-year-olds in Experiment 1 were also able to generate appropriate interventions, and did not imitate the experimenter in the indirect screening-off tasks. Three-year-olds were also able to make these inferences given a relatively small number of observations. All of this speaks against the possibility that Bayesian causal inference only develops between ages three and four.

One might suggest that the developmental difference in Experiment 3 is even more striking since children were given explicit training about the assumptions necessary to engage in Bayesian inference. They were explicitly told that blickets make the machine go, were given corrective feedback when they categorized incorrectly on the training, and were explicitly told about and given an example of the "the activation law". However, from an information-processing standpoint, Experiment 3 might have been more difficult than Experiment 1. Children had to observe and keep track of much more data. This might have simply overwhelmed the younger children. It is possible that a more sensitive measure would elicit backwards blocking from younger children. Such a method is currently being explored in the laboratory.

The developmental differences in the present data could be due to several information-processing factors. First, to implement the Bayesian computations directly, children would have to keep track of multiple hypotheses and their associated probabilities. The younger children may simply not have this capacity. Second, young children may not use the assumptions outlined in the discussion of Experiment 2 to constrain their initial hypothesis space. For instance, do young children recognize that the blicket detector operates under an "activation law"? Three-year-olds might not recognize there is something about being a "blicket" that makes the machine go. Several researchers suggest that children's knowledge of internal mechanisms is developing between ages 3 and 4 (Bullock et al., 1982; Gelman & O'Reilly, 1988; Shultz, 1982). In Experiment 3, we explicitly pointed out the activation law to children, but it is not clear that the younger children necessarily understood this concept. If 3-year-olds lack this knowledge, it might not matter if we explicitly pointed out the relationship. We are currently exploring this possibility in the lab.

Conclusions

These experiments show that young children can engage in various kinds of causal inference that involve indirect and/or ambiguous evidence. Simple associative models are not able to predict children's inferential abilities. 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.

Mapping out the development of this inference machinery, and other cognitive capacities that support it, is an important goal for 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.

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

Baillargeon, R., 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: Comparitive and Physiological Psychology, 55B*, 241-266.

Danks, D. (in press). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology.*

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

Dickinson, A., & Shanks, D. (1995). Instrumental action and causal representation. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate. Symposia of the Fyssen Foundation* (pp. 5-25). New York, NY: Clarendon Press/Oxford University Press.

Gelman, S. A., & O'Reilly, A. W. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development, 59,* 876-887.

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. (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., & Sobel, D. M. (2000). Detecting blickets: 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.

Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning, 20*, 197-243.

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

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological monographs: General and applied, 79 (No. 594).*

Kahneman, D., & Tversky, A. (1973). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430-454.

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.

Klahr, D. (2000). *Exploring science: the cognition and development of discovery processes*. Cambridge, MA: MIT Press.

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

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674-689.

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

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.

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. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31-57.

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.

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.

Shultz, T.R. (1982). Rules of causal attribution. *Monographs of the society for research in child development, 47(1)*, 1-51.

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, 2nd edition, revised)*. Cambridge, MA: MIT Press.

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

Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. *Proceedings of the 12th Annual Conference on the Advances in Neural Information Processing Systems.*

Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal inference. *Proceedings of the 14th Annual Conference on the Advances in Neural Information Processing Systems*.

Tenenbaum, J. B., Sobel, D. M., & Gopnik, A. (2003). *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.

Author Note

David M. Sobel, Department of Cognitive and Linguistic Sciences, Brown University; Joshua B. Tenenbaum, Department of Brain and Cognitive Science, Massachusetts Institute of Technology; Alison Gopnik, Department of Psychology, University of California at Berkeley.

This work was made possible in part by a National Research Service Award (F31-MH12047) to DMS while he was at the University of California, 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, Michael Ramscar, Laura Schulz, Lani Shiota, and Dan Slobin, for helpful discussion and feedback on this manuscript.

Correspondence concerning this article should be addressed to D. Sobel, Department of Cognitive and Linguistic Science, Box 1978, Brown University, Providence, RI, 02912. Email: David_Sobel_1@brown.edu.

Table 1

Average Number of "Yes" Responses to the "Is it a Blicket?" Question in Experiment 1

Age and object	Indirect screening-off	Backward blocking
3-year-olds:		
Object demonstrated alone (A)	0.25 (0.58)	1.94 (0.25)
Object not demonstrated alone (B)	2.00 (0.00)	1.00 (0.82)
4-year-olds:		
Object demonstrated alone (A):	0.00 (0.00)	2.00 (0.00)
Object not demonstrated alone (B):	2.00 (0.00)	0.25 (0.68)

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

Table 2

Average Number of Times Combinations of Objects were Placed on the Machine in Responses to the "Make the Machine Go" Question in Experiment 1

	Indirect screening-off	Backward blocking
Age and object choice		
3-year-olds:		
Only object demonstrated alone (A)	0.13 (0.34)	1.06 (0.93)
Only object not demonstrated alone (B)	1.63 (0.72)	0.63 (0.81)
Both objects together	0.25 (0.58)	0.31 (0.60)
4-year-olds:		
Object demonstrated alone (A):	0.00 (0.00)	1.44 (0.81)
Object not demonstrated alone (B):	1.75 (0.58)	0.13 (0.34)
Both objects together	0.25 (0.58)	0.44 (0.73)

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

Table 3

Distribution of Responses between the Two Conditions in Experiment 1

3-year-olds	Backwards blocking			
	A on both	B on both	Both on both	Other
Indirect screening-off				
Chose A on both trials	0	0	0	0
Chose B on both trials	4	0	2	6
Chose Both objects on both trials	0	0	1	0
Other Response	1	0	1	1

4-year-olds	Backwards blocking			
	A on both	B on both	Both on both	Other
Indirect Screening-off				
Chose A on both trials	0	0	0	0
Chose B on both trials	14	0	2	0
Chose Both objects on both trials	0	0	0	0
Other Response	0	0	0	0

Notes. A Bayesian structure learning account predicts children will be more inclined to choose object B in the indirect screening-off than the backwards blocking condition. This response pattern is in bold.

Table 4

Frequency of "Yes" Responses to the "Is it a Blicket?" Question in Experiment 2

Object choice	Indirect Screening-off	Backward blocking
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)

Object choice	Association
Object with more associative strength (F)	2.00 (0.00)
Object with less associative strength (E)	1.88 (0.17)

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

Table 5

Probability that Children Placed the B block in the Backwards Blocking Trial and Either Block in the Baseline Trial in the Blicket Box across the Two Conditions in Experiment 3

3-year-olds

	Backwards Blocking Trial	Baseline Trial	Difference
Rare Condition	0.81	0.94	- 0.13
	(0.40)	(0.17)	
Common Condition	0.88	1.00	- 0.13
	(0.34)	(0.00)	
	Backwards Blocking Trial	Baseline Trial	Difference
Rare Condition	0.25	0.78	- 0.53
	(0.45)	(0.26)	
Common Condition	0.81	0.97	- 0.16
	(0.40)	(0.13)	

Notes. Standard deviation in parentheses

Figure Captions

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.

Endnotes

[1] For both the categorization and intervention questions, these analyses were supplemented by a Chi-squared analysis. For the categorization question, there was a significant difference between the responses of the 3- and 4-year-olds: $\square^2(1) = 4.27$, $p < .05$. However, since children never gave two of the four response types, and since 4-year-olds always choose only object B, we also analyzed these data using a Fisher Exact test, which was not significant. Further, analyzing the responses of only the first indirect screening-off trial also revealed no difference between 3- and 4-year-olds' performance: $\square^2(1) = 1.03$, *ns*. For the intervention question, these analyses revealed no significant differences between the two age groups.

[2] For both the categorization and intervention questions, analyses were supplemented by a Chi-squared analysis. For the categorization question, there was a significant difference between the distribution of responses of the 3- and 4-year-olds: $\square^2(2) = 10.64$, $p < .01$. Further, analyzing the responses of only the first backwards blocking trial revealed a similar finding: $\square^2(1) = 3.87$, $p < .05$. For the intervention question, there was a significant difference between the distributions of responses of the two age groups: $\square^2(2) = 6.57$, $p < .05$, which also held when only the first backwards blocking trial was considered $\square^2(2) = 8.23$, $p < .05$.

[3] 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.