Preschool children learn causal structure from conditional interventions

Laura E. Schulz and Alison Gopnik

University of California, Berkeley

Clark Glymour

Carnegie Mellon University

and

Institute for Human and Machine Cognition, Pensacola, Fl

Submitted to *Developmental Psychology* July 8th, 2004

This research was supported by a National Science Foundation Graduate Research Fellowship Award (#99-136) to L.S and a National Science Foundation grant (DLS0132487) to A.G. . C.G.'s work was supported by grant NN00014-03-1-0516 from the Office of Naval Research to IHMC We are grateful to Catherine Griffith, Jeri Janowsky, Tamar Kushnir, Lori Markson, and Christine Schaeffer for helpful comments and suggestions. Thanks also to the children and staff at Fruit and Flower, Northeast Community Child Development Center, Helen Gordon Child Development Center, the Goddard School, and the Berkeley Child Study Center. Address for correspondence: L. Schulz & A. Gopnik, Psychology Department, University of California, 3210 Tolman Hall #1650, Berkeley, CA 94720-1650. Phone: 503-225-0799. Fax: 510-642-5293. E-mail: laurasch@socrates.berkeley.edu.

Preschool children learn causal structure from conditional interventions

Abstract

The conditional intervention principle is a formal principle that relates patterns of interventions and outcomes to causal structure. It is a central assumption of the causal Bayes net formalism. Four experiments suggest that preschoolers can use the conditional intervention principle both to learn complex causal structure from patterns of evidence and to predict patterns of evidence from knowledge of causal structure. Other theories of causal learning do not account for these results.

Preschool children learn causal structure from conditional interventions

By the time children are five years old, they understand a great deal about the causal structure of the world. Research on children's intuitive theories suggests that preschool children know some of the causal principles of physics, biology and psychology (Flavell, Green, & Flavell, 1995; Gelman & Wellman, 1991; Gopnik & Meltzoff; 1997; Gopnik & Wellman, 1994; Kalish, 1996). Across domains, young children are able to generate appropriate causal predictions, explanations and even counterfactual inferences (Bullock, Gelman & Baillargeon, 1982; Harris, German, & Mills, 1996; Sobel, 2001; Wellman, Hickling, & Schult, 1997). Moreover, children's causal theories change over time (Bartsch & Wellman, 1995) and in response to new evidence (Slaughter & Gopnik, 1996; Slaughter, Jaakola, & Carey, 1999).

However, much less is known about the process of causal learning. We know that children's knowledge changes but we do not know how or why it changes. The few experimental studies of children's causal reasoning have focused largely on the role that substantive, domain-specific principles play in children's causal judgments. Studies show for instance, that even infants are sensitive to spatiotemporal cues about causal relations (Leslie & Keeble, 1987; Oakes & Cohen, 1990) and that children use information about causal mechanisms involving contact and exchange of force when making judgments about physical causality (Bullock, Gelman & Baillargeon, 1982; Shultz, 1982). Accordingly, some researchers have suggested that children's early causal understanding might originate in domain-specific modules (Leslie & Keeble, 1987) or from innate concepts in core domains (Carey & Spelke, 1994; Keil, 1995; Spelke, Breinlinger, Macomber & Jacobson, 1992). In adult cognitive psychology by contrast, researchers have focused primarily on domain-general causal learning mechanisms. Studies show, for example, that adults make causal inferences based on the strength of association between events (Shanks, 1985; Shanks & Dickinson, 1987; Spellman, 1996) and based on patterns of covariation (Cheng, 1997, 2000).

Intuitively, however, it seems possible to have genuinely causal knowledge without necessarily knowing the mechanisms, spatiotemporal relations, and force dynamics that might underlie a causal relationship. On the other hand, understanding causation seems to require more than just understanding patterns of covariation or association. Recently, a number of philosophers and computer scientists have suggested that the crucial piece that is missing from covariation accounts is the notion of intervention.

Specifically, philosophers of science have suggested that knowing that X directly causes Y means knowing that, all else being equal, intervening to change X can change Y (see e.g., Woodward, 2003, Pearl 2000, Spirtes et al. 1993). Conversely, if intervening to change the value of X, changes the value of Y, all else being equal, we can conclude that the relationship between X and Y is causal. Moreover, we can infer the causal link between X and Y even if we don't know the mechanisms underlying the relationship. ~~This is why controlled experiments have become invaluable for causal inference in science.~~ On the other hand, if, controlling for all else, changing X does not change Y, then we have reason to believe that there is not a direct causal link between X and Y. This is why controlled experiments have become invaluable for causal inference in science.

Recently, philosophers of science have formalized this intuition and developed what has been called the manipulationist or interventionist account of causation. In this account, the necessary and sufficient conditions for X to be a cause of Y are not spatiotemporal characteristics or force relations, nor evidence of covariation per se, but the particular pattern of covariation of interventions and outcomes (see Woodward, 2003)[1]. We will discuss this interventionist account of causation in detail, first intuitively, then by defining an intervention and the formal relationship of interventions to causal relationships, and finally by suggesting how evidence from interventions and outcomes might be used in a particular case of causal learning.

Suppose you notice that drinking wine, going to parties and having insomnia all covary. When you go to a party and drink wine, you don't sleep well. This could be because wine is keeping you up, and parties make you drink wine, or it could be because parties keep you up and parties also make you drink wine. Assuming that these are the only three relevant variables, you can find out which causal possibility is correct by intervening to keep one factor constant and then intervening to vary the other factor. For example, you could try going to parties and drinking and going to parties sober. If there is no difference in how you sleep, then wine is not likely to be the direct cause of your insomnia. If there is a difference, then wine is a likely cause -- and this is true even if you don't know anything about the mechanism by which wine keeps you awake.

[1] The interventionist account stipulates that causes have an interventionist interpretation, whether or not the interventions actually take place. Although we deal with actual human interventions in this paper, "natural experiments" and counterfactual interventions are not excluded (see Pearl, 2000; Spirtes, Glymour, & Scheines, 1993, and Woodward, 2003 for discussion).

Recently, philosophers and computer scientists have formalized these intuitions systematically. Causal relations, like the relationship between drinking, partying and insomnia, can be represented by directed graphs, also called causal Bayes nets (see Figure 1). The nodes in the graph represent variables (e.g., wine consumption, party attendance, sleeplessness, etc) that can take particular values (e.g., present or absent) and arrows represent direct causal relationships between those variables. The relationships among these variables are constrained by a single assumption, called the Causal Markov assumption, which relates the structure of the graph to the patterns of probability of the variables. The Causal Markov assumption says that for a particular graph, only some patterns of probability will occur among the variables and not others. In particular, it says that all the variables in a graph are probabilistically independent of all other variables except their own effects, conditional on their own direct causes (see Gopnik et al. 2004).

The Bayes net formalism also says that for a particular graph, interventions on some variables will lead to a particular pattern of changes in the other variables. Within the formalism, interventions ~~variables~~ are treated as special additional variables with special features. 1) They must be exogenous, that is they must not be influenced by any other causal factors in the graph. 2) They must ~~also~~ fix the value (or probability distribution) of the variables of interest. After an intervention, the value (or probability distribution) of the intervened-upon variable is entirely determined by the intervention. Whether we intervene on party-going or wine, the value of that variable is determined by our intervention and independent of everything else. 3) Finally, the intervention can only influence other variables in the graph through its effect on the intervened on variable.

The result of these special features is that interventions on a Causal Bayes net break arrows *into* the variables of interest, performing what Judea Pearl vividly described as "graph surgery" (2000). If for instance, parties, wine drinking and sleeplessness are all causally related as shown in Figure 1a, an intervention I on W breaks the arrow between P and W (see Figure 1b) so that, after the intervention, the value of W depends entirely on the intervention and is independent of P. We can then look at the graph after this "graph surgery" has been performed, that is, after the intervention has taken place, and figure out what happens to the other variables in the graph.

Insert Figure 1 here

There are several different ways of formally capturing these relations between interventions, dependencies and causal arrows (see Pearl, 2000; Spirtes, Glymour, & Scheines, 1993; Woodward, 2003). One way to do this is in terms of what we have called the conditional intervention principle. The conditional intervention principle can be formally stated as follows: for a set of variables in a causal graph, X directly causes Y (that is, X→Y) if and only if: 1) there is some intervention that fixes the values of all other variables in the graph and results in Y having a particular probability distribution, pr (Y); such that 2) there is another intervention that changes the value of X and 3) changes the probability distribution of Y from pr (Y) to pr' (Y) but 4) does not influence Y other than through X and 5) does not undo the fixed value of the other variables in the graph (Gopnik et al., 2004).

Although this principle may sound complex, it is simply a formal statement of the sort of intuitions about intervention and causation that underlie experimental design, as in the wine-insomnia experiment we described above. ~~Just as i~~In an experiment, if I want to find out the causal relationship between two variables, I intervene to hold all other variables constant, and then I intervene to manipulate the value of the variable of interest. If for instance, I want to know the causal relationship between wine drinking and sleeplessness (represented by an arrow with a question mark in Figure 2a below) I can perform one intervention (I_1 in Figure 2a) to hold all other potential causes of S constant and another intervention to change the value of W (I_2 in Figure 2a). If the value (or probability distribution) of S changes, I can conclude that W causes S.

Note also that the conditional intervention principle rules out confounded interventions. Line 4 of the conditional intervention principle eliminates the graph in 2b (because the intervention on W cannot influence S except through W) and line 5 rules out the confounded graph in 2c (because interventions cannot change the fixed value of any other variable in the graph).

Insert Figure 2 here

The conditional intervention principle underlies the type of causal learning that occurs in scientific experimentation. However, this formal relationship between interventions and causal structure might support causal learning quite generally, including causal learning in children. There is abundant evidence that young children learn from the evidence of their own interventions (Rovee-Collier, 1987; Watson &

Ramey, 1987). However, the research on young children's ability to learn from interventions has (with a few exceptions noted below) largely been restricted to investigations of operant learning and trial-and-error learning. If a child intervenes to change the value of X, she can learn that her intervention caused X to change; indeed, under the right ecological conditions, rats and pigeons can make the same sort of inference.

Critically however, the conditional intervention principle enables us to learn, not just the direct outcome of interventions, but other causal relations. That is, if I intervene to fix the value of other variables in the graph, and I then intervene on X, I can learn not just that my intervention causes X, but also whether X causes Y. Moreover, in principle, such learning can occur not just in complex cases of scientific inquiry but also in everyday causal reasoning.

Imagine for instance, that you turn on a switch and two gears start to spin simultaneously. Assuming the causal relationships between the gears are generative and deterministic, the switch could: (i) activate a mechanism that moves A, which in turn moves B, or (ii) it could activate a mechanism that moves B, which in turn moves A, or, (iii) the switch could activate mechanisms that separately move gears A and B, or (iv) conceivably, the switch could activate a mechanism in which the spinning of each gear is necessary for the spinning of the other (see Figure 3).

Insert Figure 3 here

Suppose you want to know the relationship between the two gears. Note that a possible mechanism of generative transmission underlies each of these causal relationships, no cues about force, contact or time order distinguish the structures, and the association between the gears' movements is identical in all cases. You could of course spin gear A with your finger to spin gear B and vice versa but that would not tell you which causal relationship obtained in the presence of the switch.

However, you could learn the relationship between the gears by combinations of interventions. Let's assume that the switch can be set to on or off, and each gear can be removed from its spindle. Changes in these values are interventions in the system and you can use evidence from such interventions to learn the causal structure. In the simplest scenario, suppose you fixed the switch to the value "on" and then removed and replaced each gear. If, when you removed gear A, gear B stopped and when you replaced gear A, gear B began to spin -- but removing B had no effect on A -- then you could use the immediate outcome of your own intervention on A (much as in operant learning) to conclude that A→B and structure 3a was correct.

However, suppose for some reason (perhaps because your fingers might get pinched) you cannot remove a wheel when the switch is on. If you cannot remove wheel A to see what happens to wheel B, then you cannot infer the causal relation between A and B from the immediate effect of your own interventions. Nonetheless, the conditional intervention principle still applies. You could for instance, remove gear A, flip the switch on, and observe gear B. Then you could flip the switch off, replace gear A, remove gear B and flip the switch on. Now the immediate effect of your intervention to remove A is not to stop B (because the switch was off and B wasn't moving in the first place), nor is the immediate effect of your intervention to replace A, to make B spin (because B is still in both cases). Nonetheless, consistent with the conditional intervention principle, controlling for other causes of A, an intervention on A changes the value of B (that is, it changes B's relationship to the switch) whereas controlling for other causes of B, an intervention on B does not change the value of A. You should conclude that structure 3a is correct and A→B.

Critically, this type of causal learning is not easily explained by other accounts of causal inference. Several accounts of causal learning show that people can discriminate among candidate causes based on differences in associative strength (e.g., Shanks & Dickinson, 1987) and patterns of covariation (Cheng, 1997; 2000). However, these accounts only apply to distinctions *among* candidate causes, assuming that the learner knows in advance which variables are potential causes and which are effects. In the gear example this assumption does not obtain. Indeed, the task of causal learning frequently involves determining whether A causes B or B causes A.

As shown in the data tables in Figure 3, the data from interventions are unique to each structure. However, as is also evident in the data table, you need to intervene on both gears to eliminate all the possibilities. If you simply remove gear A, flip the switch and observe that B spins, you won't know whether the structure is 3b or 3c (compare lines 5 and 6 of both structures). Similarly, if you simply remove gear A, flip the switch, and see that B does not spin, you won't know whether the structure is 3a or 3d. Note also that to draw correct causal inferences you need to intervene on both the switch and the gears. If you left the switch set to "off" and an intervention on A failed to change B, you could not conclude that A did not cause B. In experimental terms, the switch is a confounder: B might not spin because of your intervention on A or because of the state of the switch. Since you are primarily interested in the relationship between the gears, A and B are the potential causal variables and the switch is the background variable that must be controlled.

You can apply the general principles of the Bayes net formalism to this particular case as follows. . For the structure S -> A -> B, for example, the causal Markov Assumption requires that for all values of A (spinning or not), of B (spinning or not), and of S (on or off) the probabilities satisfy: Pr (A, B, S) = Pr (B | A) * Pr (A | S) * Pr (S), where Pr (B | A) is the conditional probability of A given S. An intervention, such as removing wheel A, breaks the arrow between S and B, so that the value of S and the value of B become independent. In plain terms, when A is removed, turning S on and off no longer changes the value of B.

Because the gear system is deterministic, the causal Bayes net probability predictions reduce to simple deterministic equations. Each of the structures in the gear example corresponds to a particular set of Boolean equations in which each variable, X, is expressed as a function of the variables whose arrows point into X. In an intervention (such as removing a wheel from its spindle, and thereby forcing the value of the wheel to be "still"), the value of the variable is forced to equal to the value set by the intervention (Pearl, 2000). The causal graph and the equations imply predictions from interventions.

For example, suppose that A =1 means A spins and A = 0 means A does not spin, analogously for B, and S = 1 means S is on and S = 0 means S is off. If the structure is the one shown in Figure 3a, S→A→B, then S will take whatever value it is set to, A will take the same value as S and B will take the same value as A. These are the Boolean equations shown in the cell of the first row, second column of Table 1 below. If A is removed and forced not to spin (that is, A = 0), the other equations still hold. So B will take the same value as A (row 1, column 3 below). Conversely, if B is removed (B = 0), A will be unaffected because A = S (column 4). Table 1 shows the causal Bayes net predictions of how evidence will change under interventions for each of the four structures.

Insert Table 1 here

Note that because the changes in the formulas are unique to each case, the correct structure can be uniquely determined from the data that result from the interventions. In this way we can apply the general conditional intervention principle to this particular case, and correctly learn the structure from the evidence. Applied to the gears, these computational formulas produce exactly the same relation between the structure and evidence that is captured by the conditional intervention principle.

Recently causal Bayes nets have been used to model a variety of complex causal reasoning problems in both adults (Glymour, 2001; Lagnado & Sloman, 2002; Rehder & Hastie, 2001; Steyvers, Tenenbaum, Wagenmakers & Blum, 2003; Tenenbaum & Griffiths, 2001; Waldmann & Hagmayer, 2001) and children (Gopnik et al., 2004; Gopnik, Sobel, Schulz, & Glymour, 2001; Schulz & Gopnik, 2004; Sobel, Tenenbaum & Gopnik, 2004). Steyvers et al., Lagnado & Sloman, and Sobel & Kushnir additionally showed that although adults can use patterns of conditional dependence to infer complex causal structure, their performance improves greatly when they are allowed to intervene on the causal system. Gopnik et al. showed that children can also use information about interventions to make simple causal inferences (2004). For example, 4-year-olds can use information about interventions and outcomes to decide whether A causes B or whether B causes A, when A and B are the only variables. However, as yet there have been no studies looking at whether children can use patterns of intervention and dependence to infer causal structures in more complex contexts. There have also been no studies, with adults or children, looking specifically at the conditional intervention principle.

In the following experiments, we presented children with the gear machine described above and looked at whether they could use patterns of evidence to distinguish among the four causal structures. In Experiment 1, we specifically assess young children's ability to use evidence from combinations of interventions to distinguish causes from effects and determine the direction of a causal chain.

Experiment 1

Method

Participants

Eighty children (mean age was 4;6, range: 40– 64 months) were recruited from urban area preschools. Approximately equal number of boys and girls participated. Twenty-five children were randomly assigned to a test condition and eighteen children to each of three control conditions. One child in the test condition did not complete the training and one child in the salience control condition was dropped due to experimenter error. These children were replaced. While most children were from white, middle-class backgrounds, a range of ethnicities resembling the diversity of the population was represented.

Materials

Two plastic, red and yellow gears were used in the training phase. The gears could be affixed to a square base so that they interlocked and a crank could be attached to the top of either gear to make it spin.

In the test phase, a custom-built electronic gear toy was used (see Figure 3). The toy had 12 uniquely colored gears. A new pair of gears was used on each trial (particular colors chosen at random). The gears could be placed on one of two pegs, A or B. Sensors detected the presence of the gears and a control hidden in the back of the toy allowed the experimenter to implement each of the four structures shown in Figure 1 (although in Experiment 1, children only saw structures 1a and 1b). A switch on the front activated the toy. In all cases, if the switch was set to "on" and both gears were on the toy, both gears spun. If the switch was set to "on" and no gears were on the toy, the pegs did not spin.

Procedure

A female experimenter who was familiar to the children tested all the participants. Children were brought into a private game room in their school and sat facing the experimenter at a table.

Training.

Children were introduced to the plastic base and told to place the red and blue gears on the base. The experimenter attached the crank to one of the gears and said, "Look, I can use the crank to make the red wheel spin the blue wheel." The experimenter removed the crank and asked the child to position it to "Make the blue wheel spin the red wheel." The experimenter removed the crank, changed the configuration of the wheels and asked the child to “Make the red wheel spin the blue wheel.” One child in the test condition failed to complete the training and was replaced.

Test.

Children were then introduced to the electronic toy and told, “In this toy the cranks are hidden inside. You can’t see the cranks.” In all conditions, trials began with the switch in the off position. The experimenter placed two gears on the toy and asked the child to “Point to the (yellow) wheel. Point to the (green) wheel.” The experimenter then removed and replaced each gear in turn, explaining that she could “take the gears on and off the machine”. Then she flipped the switch on to make both gears spin simultaneously and flipped the switch off so that both gears were still. This provided evidence equivalent to lines 1 – 4, 5 and 7 in Figure 1.

In the Test condition, children received two trials in counterbalanced order. One trial provided evidence for the structure in Figure 1a, the other for the structure in Figure 1b. For example, for the structure represented by 1a, the experimenter removed A and, turned on the switch, and B failed to spin. She turned off the switch, replaced A, removed B, and turned on the switch, and A spun. This procedure provided the information in lines 6 and 8 of Figure 1a. A new pair of gears was used on each trial, and the toy was rotated 90° between trials to avoid any position biases. At the end of each trial, children were asked, “Which gear has the crank?”

Children might think that switches and cranks go together; that is they might assume that gears that move with switches have cranks and gears that don’t move with switches don’t have cranks. Alternatively, children might assume that if the switch fails to move one gear, the other gear must be the mechanism causing the gears’ movement and must therefore have the crank. To rule out the possibility that children were relying on such prior assumptions about causal mechanisms, eighteen children were tested in a Movement control condition and eighteen children were tested in a Still control condition. The control conditions were identical to the test condition except that on each trial, the experimenter removed only a single gear.

For example, in the Movement control condition for 1a, the experimenter removed B and flipped the switch on so that A spun. The experimenter repeated this intervention but never intervened to remove A. If children believe that switches and cranks go together, they should infer that the moving gear has the crank and the children's responses should be identical to those of children in the Test condition. However, if children rely on patterns of interventions and outcomes to infer a unique causal structure, this evidence is insufficient. The children had evidence represented by line 8 of Figure 1a, but not evidence for line 6. Given this evidence, it could be that A spins B, but B does not spin A . But it could also be that neither gear has any causal influence on the other -- that is, it could be that the switch spins each gear independently; the evidence is consistent with both structures 1a and 1c. If children are unsure of the underlying structure, then children might choose between the gears at chance.

Similarly, in the Still control condition, for structure 1a, the experimenter removed A and flipped the switch on; B failed to spin. If children are using prior knowledge about gears as causal mechanisms, they should infer that the still gear does not have the crank and respond like children in the test condition. Formally, however, the children had evidence for line 6 of Figure 1a, but not line 8. Since the evidence for Figures 1a and 1d differs only at line 8, the evidence fails to disambiguate the structures and children might choose between the gears at chance.

An additional eighteen children were tested in a Contrast control condition in which children saw a still and spinning gear, but these perceptual facts were unrelated to the switch. Thus in the control for 1a, the switch stayed in the off position, the experimenter removed A and B failed to spin; then she replaced A, removed B, and spun A with her finger. If children are relying only on the relative salience of the two gears and prefer a moving gear when contrasted with a still gear, then the children in this condition should perform like children in the test condition. Formally however, children in this condition had no information to disambiguate the causal structures.

Results

In the test condition, 84% of children correctly identified the causal gear on trial 1 and 76% of children correctly identified the causal gear on trial 2. On each trial, children identified the causal gear significantly more often than would be expected by chance (Trial 1, n = 25, $p < .001$ by binomial test; Trial 2, n = 25, $p < .025$ by binomial test). Sixty-four percent of the children correctly identified the causal gear on both trials, significantly more than would be expected by chance (n = 25, $p < .025$ by binomial test).

In the movement control condition, 56% of children chose the target gear (the moving gear) on trial 1 and 61% chose the target gear on trial 2. Children did not choose the target gear significantly more often than would be expected by chance on either trial (n = 18, $p = ns$ by binomial test for both trials). Thirty-three percent of the children chose the target gear on both trials, not significantly different from chance (n = 18, $p = ns$ by binomial test).

In the still control condition, 50% of children chose the target gear on trial 1 and 28% chose the target gear on trial 2. Children did not choose the target gear significantly more often than would be expected by chance on either trial (n = 18, $p = ns$ by binomial test for both trials). Six percent of the children chose the target gear on both trials, not significantly different from chance (n = 18, $p = ns$ by binomial test).

In the contrast control condition, 61% of children chose the target gear on trial 1 and 44% chose the target gear on trial 2. Children did not choose the target gear significantly more often than would be expected by chance on either trial (n = 18, $p = ns$ by binomial test for both trials). Twenty-two percent of the children chose the target gear on both trials, not significantly different from chance ($p = ns$ by binomial test).

Comparing across conditions, children were significantly more likely to prefer the causal gear to the non-causal gear on both trials of the test condition than they were to choose the target gear over the other gear on both trials of the moving control condition, $\chi^2(1, n = 43) = 3.94$, $p < .05$, the still control condition, $\chi^2(1, n = 43) = 14.95$, $p < .001$, or the contrast control condition, $\chi^2(1, n = 43) = 7.34$, $p < .01$. Table 1 shows the number of children choosing the target gear at ceiling across conditions.

Insert Table 2 here

Discussion

These results suggest that preschool children can learn causal structure from conditional interventions. Children were able to use each gear's relationship with the switch to determine the causal relationship between the gears. Children's chance performance in the movement and still control conditions suggest that the children were not simply relying on prior knowledge about causal mechanisms. That is, even though children were given a forced choice between the gears, children did not assume that if a gear moved with the switch it must have the crank or must make the other gear go, nor did children assume that if a gear failed to move with the switch it must not have a crank or that the other gear must make it go

Experiment 2

In Experiment 1, children were asked only to distinguish between causal chains: does the switch influence A which influences B or does the switch influence B which influences A. However, as noted, there are two other causal possibilities: neither gear might causally influence the other (because the switch independently spins both gears) or both gears might causally influence each other (because the gears only spin when both gears are present). Can children use information about interventions and outcomes to distinguish the causal chains, not only from one another but also from the common effects structure in Figure 1c and the conjunction structure in Figure 1d?

Method

Participants

Fifty-six children (mean age: 56 months; range: 49– 66 months) were tested. The children were randomly assigned to three conditions: twenty children were assigned to each of two test conditions, A and B and sixteen children were assigned to a control condition. Approximately equal number of boys and girls participated. While most children were from white, middle-class backgrounds, a range of ethnicities resembling the diversity of the population was represented.

Materials

The same materials used in Experiment 1 were used in Experiment 2. Additionally, for Group A, two sets of three pictures like the pictures in Figure 1a, 1b, and 1d were used; for Group B two sets of three pictures like the pictures in Figure 1a, 1b, and 1c were used. Each picture also depicted a crank in the appropriate position (i.e., for 1a the crank was on gear A; for 1b the crank was on gear B; for 1c a crank was on both gears, and for 1d a crank was in between the gears) and each set was colored to match the gears used during that trial. Two training pictures were also used, one for Figure 1a and the other for Figure 1b.

Procedure

A female experimenter who was familiar to the children tested all the participants. Children were brought into a private game room in their school and sat facing the experimenter at a table.

Training.

Children received the same training as in Experiment 1. Then the experimenter introduced the two training pictures, saying "Look here's a red wheel in the picture and here's a blue wheel. See, in this picture red is pushing blue. Red makes blue go. In this picture, blue is pushing red. Blue makes red go." The experimenter set up the gears on the base and turned one of the wheels with the crank. She asked the children to pick the picture that matched what was happening on the toy. She then collected the pictures, reconfigured the wheels and asked the child to pick a picture again. Then the experimenter said, "Now I'm going to give you a picture. Can you make it so the wheels on the toy match this picture?" After the child set up the gears and crank so that they matched the picture, the experimenter removed the wheels, switched pictures and asked the child to set up the toy again. Every child was able to complete the training.

Test Conditions.

Children in group A were asked to distinguish the chains from one another and from the conjunction structure, thus on each trial children had a choice of three pictures corresponding to pictures 1a, 1b, and 1d. Children in group B were asked to distinguish the chains from one another and from the common effects structure, thus on each trial children had a choice of three pictures corresponding to pictures 1a, 1b, and 1c.

Before each trial, the candidate pictures were set before the child in random order and children were told, "Here are some ways the toy could work". The pictures were described in terms of the colors of the gears. For example, for 1a, and red and blue gears, children were told, "Red is pushing blue. Red makes blue go." For 1b they were told, "Blue is pushing red. Blue makes red go." For 1c they were told: "Red doesn't push blue and blue doesn't push red. Each gear has its own crank." For 1d they were told: "The switch and red together push blue, and the switch and blue together push red. The crank is in the middle." Children were asked to redescribe each picture and corrected if necessary.

As in Experiment 1, children were told "In this toy the cranks are hidden inside. You can't see the cranks." For each structure, the experimenter provided evidence following the procedure described in Experiment 1. In Group A, one trial provided evidence for the chain represented by Figure 1a; the other trial provided evidence that neither gear would spin without the other, as in Figure 1d (trials presented in counterbalanced order throughout). For each trial, children in Group A were asked to pick the correct picture from among three pictures corresponding to the pictures in Figure 1a, 1b, and 1d. In Group B, one trial provided evidence for the chain represented by Figure 1b; the other trial provided evidence that the gears would spin independently, as in Figure 1c. After each trial, children were asked, "Can you give me the picture that shows how the toy is working right now?" In Group B children were asked to pick from among three pictures corresponding to the pictures in Figure 1a, 1b, and 1c.

Control Condition.

In the control condition we wanted to rule out the possibility that children were using perceptual cues or patterns of association between the gears to infer causal structure. Additionally, we wanted to verify that in the test condition, children were paying attention to both the intervention on the switch and the intervention on the gears, rather than simply attending to the gears. The procedure was identical to the procedure given to the children in Group A, except that after the children saw that the switch made both gears spin, the toy was turned 180° so children could see the gears but could not tell whether or not the experimenter flipped the switch. However, the perceptual features of the gears and their associations were identical to those seen by Group A. (We did not do this same procedure with Group B because the evidence for structure 1c reveals the state of the switch, whether the children are able to see the switch or not. That is, because gears never spin spontaneously, evidence that each gear spins separately is effectively evidence that the switch was flipped for each gear. Thus turning the toy around would not prevent children from conditioning on the state of the switch.)

Results and Discussion

Preliminary analyses revealed no effect of order of trial presentation. Given evidence for the causal chain in Figure 1a, children in group A were significantly more likely to choose the correct chain (picture 1a) than expected by chance (n = 20, $p < .001$ by binomial test). Children were also significantly more likely to choose the correct chain than to choose the incorrect chain represented by picture 1b (n = 16, $p < .001$ by binomial test) or the conjunction represented by picture 1d (n = 19, $p < .025$ by binomial test). Similarly, given evidence for the conjunction in Figure 1d, children were significantly more likely to choose the correct structure than expected by chance (n = 20, $p < .001$ by binomial test) and significantly more likely to choose the correct structure than to choose the incorrect chain represented by either picture 1a (n=19, $p < .001$ by binomial test) or picture 1b (n = 18, $p < .001$ by binomial test). Sixty percent of the children in Group A chose the correct structure on both trials, significantly more than expected by chance (n = 20, $p < .001$ by binomial test).

Given evidence for the causal chain in Figure 1b, children in Group B were significantly more likely to choose the correct chain (picture 1b) than expected by chance, (n = 20, $p < .005$ by binomial test) and significantly more likely to choose the correct chain than to choose the incorrect chain represented by picture 1a (n = 16, $p < .025$ by binomial test) or the common effects structure represented by picture 1c (n = 17, $p < .025$ by binomial test). Similarly, given evidence for the common effects structure in Figure 1c, children were significantly more likely to choose the correct structure than expected by chance (n = 20, $p < .001$ by binomial test) and significantly more likely to choose the correct structure than to choose the incorrect chain represented by either picture 1a or 1b, (n = 18, $p < .001$ by binomial test) for both. Fifty-five percent of the children in Group B chose the correct structure on both trials, significantly more than expected by chance (n = 20, $p < .001$ by binomial test).

By contrast, children in the control condition chose among the pictures at chance. Given evidence about the gears (but not the switch) comparable to structure 1a, 37% of the children chose picture 1a, no different from chance (n = 16, p = *ns* by binomial test). Given evidence about the gears comparable to structure 1d, 50% of the children chose picture 1d, no different from chance (n = 16, p = *ns* by binomial test). Twelve percent of the children chose 1a on the trial where one wheel moved without the other and also chose 1d on the trial where neither wheel moved without the other, no different than expected by chance (n = 16, p = *ns* by binomial test). The results from all three groups are presented in Table 3.

Insert Table 3 here

These results suggest that preschool children can use information about interventions and outcomes, not only to distinguish between causal chains, but to distinguish causal chains from common effects structures and causal conjunctions. Children's chance performance in the control condition suggests that children are not basing their causal judgments on perceptual cues or the associative strength between the gears (which were identical in the test and control conditions), or on the fact that a moving wheel might be more salient (or mechanistically more "causal-seeming") than a still wheel. Instead, consistent with the conditional intervention principle, children's causal inferences seem to draw on information from two separate interventions – the intervention to fix the state of the switch and the intervention to fix the state of the wheel.

Experiment 3

Experiment 2 suggested that children could use patterns of interventions and outcomes to distinguish chains from common effects and conjunctions; however, it did not demonstrate that children could distinguish all four causal structures. Specifically, children were never asked to distinguish the common effects structure from the conjunction structure. In Experiment 3, we gave children evidence for each of these two structures and asked the children to choose the correct structure from among the common effects structure, the conjunction structure and the causal chains.

Method

Participants

Fourteen children (mean age: 54 months; range: 42– 61 months) were tested. Approximately equal number of boys and girls participated. While most children were from white, middle-class backgrounds, a range of ethnicities resembling the diversity of the population was represented.

Materials

The same materials used in Experiment 2 were used in Experiment 3 except that the children did not participate in a training condition so the plastic gear toy and training pictures were not used.

Procedure

A female experimenter who was familiar to the children tested all the participants. Children were brought into a private game room in their school and sat facing the experimenter at a table.

Test Conditions.

Pilot work suggested that children could understand the pictures and the electronic gear toy without prior training, so the training was omitted for this experiment. Children were introduced to the electronic toy as in Experiments 1 and 2. The children received two trials (order counterbalanced between subjects); one trial provided evidence for the structure in Figure 1c; one trial provided evidence for the structure in Figure 1d. On one trial, children had a choice between pictures corresponding to 1a, 1c, and 1d; on the other they had a choice between 1b, 1c, and 1d.

Before each trial, the candidate pictures were set before the child in random order and children were told, "Here are some ways the toy could work". Because the children had not been exposed to cranks, the cranks were omitted from the pictures and the description of pictures 1c and 1d were changed to omit the reference to cranks. For example, for 1c and red and blue gears, children were told: "Red doesn't push blue and blue doesn't push red. They each push themselves." For 1d they were told: "Red pushes Blue and Blue pushes Red. Both wheels have to push together." At the end of each trial, children were asked, "Can you give me the picture that shows how the toy is working right now?"

Results and Discussion

Preliminary analyses revealed no effect of order of trial presentation. Given evidence for the common effects structure in Figure 1c, 71% of children chose the correct structure (picture 1c), significantly more than would be expected by chance (n = 14, p < .005 by binomial test). Children chose the correct structure significantly more often than they chose the conjunction or the chain (n = 12, p < .05 by binomial test for both). Given evidence for the conjunction structure in Figure 1d, 86% of children chose the correct structure (picture d), significantly more than would be expected by chance (n = 14, p < .001 by binomial test) and significantly more often than they chose the common effects structure (n = 14, p < .025 by binomial test); no children chose the chain. Fifty-seven percent of the children chose the correct structure on both trials, significantly more than expected by chance (n = 14, p < .001 by binomial test). Table 3 shows children's responses.

Insert Table 4 here

Experiment 3 suggests that children can use interventions and patterns of conditional dependence and independence to determine whether the gears' movement was a common effect of the switch or whether the gears influenced one another. Children were also able to make these inferences even without the training used in the other experiments.Taken together with Experiment 2, these results suggest that children can use evidence consistent with the conditional intervention principle to distinguish all four causal structures. Interestingly, no one structure appeared to be more difficult than any other; a significant majority of the children were able to identify the correct causal structure whether it was a chain, a common effects structure or a conjunction.

Experiment 4

The previous experiments suggest that children can use evidence about interventions and outcomes to learn causal structure. However, one of the interesting features of causal Bayes nets is that inferences can work both ways: you can use evidence from interventions to learn causal structure but you can also use knowledge of causal structure to predict the patterns of evidence that will result from interventions. In Experiment 4, we told children which causal relationship obtained and looked at whether the children could predict the evidence that would result from interventions.

Method

Participants

Sixteen children (mean age: 54 months; range: 42– 61 months) were tested. Approximately equal number of boys and girls participated. While most children were from white, middle-class backgrounds, a range of ethnicities resembling the diversity of the population was represented.

Materials

The same materials used in Experiment 3 were used in Experiment 4.

Procedure

A female experimenter who was familiar to the children tested all the participants. Children were brought into a private game room in their school and sat facing the experimenter at a table.

Training.

Children were introduced to the electronic gear toy. The experimenter placed two gears on the toy and asked the child to "Point to the (yellow) wheel. Point to the (green) wheel." The experimenter then removed and replaced each gear in turn, explaining that she could "take the gears on and off the machine". Then she flipped the switch on to make both gears spin simultaneously and flipped the switch off so that both gears were still. The experimenter then explained, "Some wheels spin by themselves." She removed the yellow wheel and flipped the switch on so that the green wheel spun. The experimenter then flipped the switch off, moved the green wheel to the other peg, flipped the switch on, and again the green wheel spun. The experimenter flipped the switch off, removed the green wheel and explained, "Some wheels don't spin by themselves. Some wheels stay still." The experimenter placed the yellow wheel on the toy and flipped the switch on. The yellow wheel stayed still. The experimenter flipped the switch off, moved the yellow wheel to the other peg, flipped the switch on, and again the yellow wheel stayed still. The experimenter flipped the switch off and removed the yellow wheel. The experimenter said, "So sometimes, both wheels spin, sometimes both wheels stay still, and sometimes one wheel spins and one wheel stays still. I'm going to show you some pictures that show how this toy can work and I want you to guess what each wheel will do."

Test.

Children received four trials, in counterbalanced order. Each of the four structures was presented on a single trial. The experimenter brought out two gears and set them on the toy. She flipped the switch on so that both wheels spun and then flipped the switch off. The experimenter then held up a single picture and explained the picture to the child. For 1a or 1b, children were told "This picture shows what is happening on the toy right now. See? This picture shows that Red is pushing Blue. Red makes Blue go." For 1c, children were told: "See? This picture shows that Red isn't pushing blue and blue isn't pushing red. They are each pushing themselves." For 1d they were told: "Red is pushing Blue and Blue is pushing Red. Both wheels are pushing together."

The experimenter placed the picture in front of the child, removed the left wheel and held the right wheel above its peg. She said, "If I turn put this wheel down right now and turn on the switch, will the wheel spin or the wheel stay still?" After the child answered, the experimenter removed the right wheel and held the left wheel above its peg and asked the same question of the right wheel. Children were counted as having answered correctly only if they answered correctly for both wheels (i.e., for 1a, the correct response was still/spin; for 1b, the correct response was spin/still; for 1c, the correct response was spin/spin; for 1d, the correct response was still/still.)

Results and Discussion

Preliminary analyses revealed no effect of order of trial presentation. For each of pictures, 1a, 1b, and 1d, 50% of children made the correct prediction, significantly more than expected by chance (n = 16, $p < .05$ by binomial test for each); for picture 1c, 44% of the children made the correct prediction (n = 16, $p = .08$ by binomial test). Other than the correct response, no pattern of responding approached significance for any structure. Twelve percent of the children made the correct predictions on all four trials, significantly more than expected by chance (n = 16, $p < .001$ by binomial test). Table 4 shows the distribution of children's predictions.

Insert Table 5 here

The results of Experiment 4 suggest that children can use knowledge of causal structure to predict the pattern of evidence that will result from interventions. Knowing the causal relationship between the gears, children were able to predict how an intervention on the switch would affect each gear individually. Consistent with the causal Bayes nets formalism, these results suggest that children can use knowledge of causal structure to predict the patterns of conditional dependence and independence that result from interventions.

On the whole, children seemed to have somewhat more difficulty predicting evidence from structure than inferring structure from evidence. Only 44-50% of the children succeeded on each trial in Experiment 4 (compared with 65-86% of children on the tasks in Experiments 1-3) and only 12% of the children performed at ceiling in Experiment 4 (compared with 55-64% of the children in Experiments 1-3). The prediction task could have been more challenging because it was more abstract: on the inference tasks, the evidence came in the form of interventions and outcomes and children could respond by choosing a picture, while on the prediction task children had to rely on pictures and verbal instructions to understand the causal structure and they also had to give a verbal response. Further research will have to establish whether, controlling for the level of abstraction, there is indeed any asymmetry between children's ability to infer structure from evidence and their ability to predict evidence from structure.

General Discussion

The results of these four experiments suggest that preschool children can use conditional interventions to learn causal structure and conversely, can use knowledge of causal structure to predict the outcome of novel interventions. Children were equally able to make these inferences whether the causal relationship was a chain, a common effects structure or a conjunction. When children were given only half of the conditional intervention information (as in the moving and still control conditions of Experiment 1) or given information about only one of the two interventions (as in the control condition of Experiment 2) children responded at chance. These results suggest that children are not using prior assumptions about gears as causal mechanisms, perceptual cues, or the strength of association between the gears to distinguish the structures. Rather children are making causal inferences based on the full pattern of interventions and outcomes that uniquely distinguish the structures from one another. These findings are consistent with the conditional intervention principle and support the idea that children's causal reasoning is consistent with the assumptions underlying the causal Bayes nets formalism.

However, this research also raises several questions. First, the causal Bayes net formalism was developed particularly to handle probabilistic data while in these experiments the data was fully deterministic. Although the conditional intervention principle is equally valid for both types of input, we do not know whether young children can make causal inferences from conditional interventions when the data is probabilistic. This question seems particularly critical given that, in the real world, children may more often be exposed to incomplete, noisy information than to deterministic input. Further research must look at how probabilistic evidence affects children's ability to infer causal structure using learning mechanisms like the conditional intervention principle.

Second, in the test conditions of these studies, children saw good, scientific interventions: that is, consistent with the conditional intervention principle, we intervened only on the target variables, and properly controlled the other variables. Under these conditions, children were able to make accurate causal inferences. However, we do not know to what extent evidence like this is available to children in everyday life and in particular, we do not know whether children can generate evidence like this on their own. There is considerable evidence suggesting that both children and adults have difficulty designing unconfounded experiments (Kuhn, 1989) and this might suggest that unconfounded evidence is rarely available to children outside the laboratory. However, although children might not have the meta-cognitive ability to design interventions appropriate for accurate causal learning, in simple cases, like with the gear toy, children – and the adults around them -- might nonetheless produce such evidence through spontaneous play or trial and error. If such evidence were available, however generated, children might be able to make accurate inferences. There is some very preliminary data suggesting that this might indeed be the case (Schulz, 2003) but further research is needed to determine if children both spontaneously generate evidence consistent with the conditional intervention principle and are able to learn from the evidence of their own interventions.

Finally, we have suggested that interventions may be central to how children think of causal relationships. These experiments support that idea, in that children were able to use information from interventions to make causal inferences when relevant differential mechanism information and spatiotemporal information was not available. Other research (Schulz & Gopnik, 2004) suggests that when domain-specific mechanism information is directly pitted against deterministic evidence from interventions and patterns of conditional probability, children preferentially use information from the interventions to make causal judgments. However, evidence from interventions does not necessarily account for all of our beliefs about causes. Evidence from interventions does not seem to explain why, for instance, children assign a more important causal role to the insides than the outsides of many entities (e.g., Gelman & Wellman, 1991). Neither does the interventionist account of causation seem to explain why we believe that for every cause, there is an intervening cause; that is, why we believe that causes are themselves mediated by causal mechanisms. Furthermore, many instances of causation do involve domain-specific causal mechanisms (i.e., in physical causality, the transmission of force or energy through spatially continuous processes). Further research must explore how formal inferences about causation, consistent with the interventionist account, interact with substantive, domain-specific concepts.

In their everyday life children intervene widely on the world and see a wide range of interventions performed by others. This study suggests that the evidence from such interventions may give children a powerful learning mechanism for inferring causal structure from evidence. At least in simple, generative, deterministic cases, preschool children seem to be able to use the conditional intervention principle to distinguish causes from effects, to infer more complex causal structure from patterns of evidence, and to predict patterns of evidence from causal structure. Previous accounts of causal learning cannot explain these results; children's causal inferences did not rely on differential mechanisms, spatio-temporal, or associative cues. Rather, even very young children seem to rely on some of the same formal principles of causal inference that underlie scientific discovery. In turn these mechanisms may help children to develop intuitive theories of the world around them.

References

Bartsch, K., & Wellman, H. (1995). *Children talk about the mind*. New York, NY: 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., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld, S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture; Based on a conference entitled "Cultural Knowledge and Domain Specificity," held in Ann Arbor, MI, Oct 13-16, 1990.* (pp. 169-200). New York, NY, US: Cambridge University Press.

Cheng, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive power. In F. C. Keil, R. A. Wilson (Eds.), *Explanation and cognition* (pp. 227-253). Cambridge, MA, US: The MIT Press.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*(2), 367-405.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development, 60(1)[243]*, v-96.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essence: Early understandings of the non-obvious. *Cognition, 38*(3), 213-244.

Glymour, C. (2001) *The mind's arrows: Bayes nets and causal graphical models in psychology*. Cambridge, MA: MIT Press.

Gopnik, A., Glymour, C.. Sobel, D., Schulz. L., Kushnir, T. & Danks. D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review 111* 1-31.

Gopnik, A., & Meltzoff, A.N. (1997). Words, thoughts and theories. Cambridge, MA. MIT Press.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37*(5), 620-629.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld, S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture; Based on a conference entitled "Cultural Knowledge and Domain Specificity," held in Ann Arbor, MI, Oct 13-16, 1990*. (pp. 257-293). New York, NY, US: Cambridge University Press.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61*(3), 233-259.

Kalish, C. (1996). Causes and symptoms in preschoolers' conceptions of illness. *Child Development, 67*(4), 1647-1670.

Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack (Eds.), *Causal cognition: A multidisciplinary debate. Symposia of the Fyssen Foundation; Fyssen Symposium, 6th Jan 1993, Pavillon Henri IV, St-Germain-en-Laye, France* (pp. 234-267). New York, NY, US: Clarendon Press/Oxford University Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674-689.

Lagnado, D. & Sloman, S.A. (2002). Learning causal structure. In W. Gray and C. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, Fairfax, VA, 560-565.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*(3), 265-288.

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development, 5,* 193-207.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufmann

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

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 3*, 323-360.

Rovee-Collier, C. (1980). Reactivation of infant memory. *Science 208*(4448), 1159-1161.

Schulz, L. E., (2003, April). *The play's the thing: Interventions and causal inference*. Paper presented at the biennial meeting of the Society for Research in Child Development. Tampa, FL.

Schulz, L. & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology, 40,* 2, 162–176.

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

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, US: Academic Press, Inc.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*(1), 1-51.

Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development, 67*, 2967–2988.

Slaughter, V., Jaakkola, R., & Carey, S. (1999). Constructing a coherent theory: Children's biological understanding of life and death. In M. Siegal & C. Peterson (Eds.), *Children's understanding of biology and health* (pp. 71–96). Cambridge MA: Cambridge University Press.

Sobel, D. M. (2004). Exploring the coherence of young children's explanatory abilities: Evidence from generating counterfactuals. *British Journal of Developmental Psychology, 22,* 37-58.

Sobel, D. M., & Kushnir, T. (2003). Interventions do not solely benefit causal structure learning: Being told what to do results in worse learning than doing it yourself. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, 1100-1105.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*(3).

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*(4), 605-632.

Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science, 7*(6), 337-342.

Spirtes, P., Glymour, C., & Scheines, R. (1993*). Causation, prediction, and search (Springer Lecture Notes in Statistics).* New York: Springer-Verlag.

Steyvers, M., Tenenbaum, J., Wagenmakers, E.J., Blum, B. (2003). Inferring Causal Networks from Observations and Interventions. *Cognitive Science, 27,* 453-489.

Tenenbaum, J. B., & Griffiths, T. L., (2003). Theory-based causal inference. In S. Becker, S. Thrun, and K. Obemayer (Eds.), *Advances in Neural Information Processing Systems 15*. Cambridge, MA: MIT Press. ~~XX-XX?~~

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. Cognition *1*, 27-58.

Watson, J. S. & Ramey, C. T. Reactions to response-contingent stimulation in early infancy. *Merrill-Palmer Quarterly 18*(3), 219-227.

Wellman, H. M., Hickling, A. K., & Schult, C. A. (1997). Young children's psychological, physical, and biological explanations. In H. M. Wellman, K. Inagaki (Eds.), *The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena. New directions for child development, No. 75* (pp. 7-25). San Francisco, CA, US: Jossey-Bass/Pfeiffer.

Woodward, J. (2003). *Making things happen: A theory of causal explanation.* New York: Oxford University Press.

Table 1

Equations associated with the gear toy.

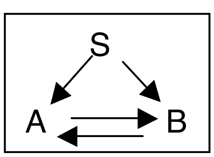
Causal Diagram	Boolean Equations	Intervention that forces A = 0	Intervention that forces B = 0
S→A→B	B = A A = S S	B = A A = 0 S	B = 0 A = S S
S→A→B	B = A A = S S	A = 0 B = S S	A = B B = 0 S
S→A→B	B = S A = S S	A = 0 B = S S	B = 0 A = S S
	A = S * B B = S * A S	A = 0 B = S * A S	B = 0 A = S * B S

Table 2

Number of children choosing the target gear on both trials (i.e., at ceiling) in Experiment 1 (n = 25 in the test condition; n = 18 in each of the three control conditions.)

Test condition	Moving control	Still control	Contrast control
16 (64)	6 (33)	1 (6)	4 (22)

Note: Percentage in parentheses

Table 3

Number of children choosing each picture in Experiment 2. Target responses are highlighted.

	Pictures		
Group A; n=20	**1a**	**1b**	**1d**
Trial giving evidence for structure 1a	15 (75)	1 (5)	4 (20)
Trial giving evidence for structure 1d	2 (10)	1 (5)	17 (85)
Group B; n=20	**1a**	**1b**	**1c**
Trial giving evidence for structure 1b	3 (15)	13 (65)	4 (20)
Trial giving evidence for structure 1c	2 (10)	2 (10)	16 (80)
Control; n=16	**1a**	**1b**	**1d**
Trial giving evidence comparable to 1a	6 (37)	2 (13)	8 (50)
Trial giving evidence comparable to 1d	4 (25)	4 (25)	8 (50)

Note: Percentage in parentheses

Table 4

Number of children choosing each picture in Experiment 3. Target responses are highlighted.

	Pictures		
N = 14	**1a/1b**	**1c**	**1d**
Trial giving evidence for structure 1c	2 (14)	10 (71)	2 (14)
Trial giving evidence for structure 1d	0 (0)	2 (14)	12 (86)

Note: Percentage in parentheses (due to rounding, percentages may not sum to 100).

Table 5

Number of children making each prediction in Experiment 4. Target responses are highlighted.

	Predictions			
N = 16	**Spin/Still**	**Still/Spin**	**Spin/Spin**	**Still/Still**
1a	8 (50)	3 (19)	1 (6)	4 (25)
1b	2 (12)	8 (50)	3 (19)	3 (19)
1c	4 (25)	4 (25)	7 (44)	1 (6)
1d	3 (19)	3 (19)	2 (12)	8 (50)

Note: Percentage in parentheses.

Figure 1

1a) A causal chain in which parties cause wine drinking which causes sleeplessness.

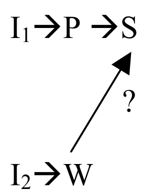
P→W→S

1b) The intervention on W breaks the arrow between P and W.

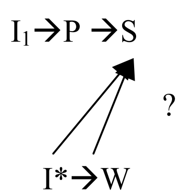
P I→W→S

Figure 2. Graphs illustrating the conditional intervention principle applied to the problem of discovering whether parties or wine is a cause of sleeplessness.

2a) I_1, fixes the value of other causes of S (clause 1 of the conditional intervention principle). I_2, changes the value of W (clause 2 of the conditional intervention principle).



2b) I* is ruled out by clause 4 of the intervention principle because the intervention affects the value of B directly.



2c) I* is ruled out by clause 5 of the intervention principle because the intervention affects other causes of B.

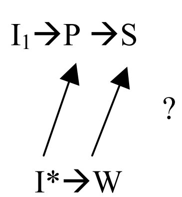
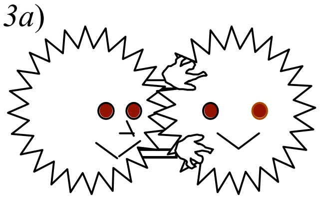


Figure 3. Pictures, graphs and evidence for the four causal structures

Pictures | Graph structures (S=Switch) | Patterns of evidence

3a)

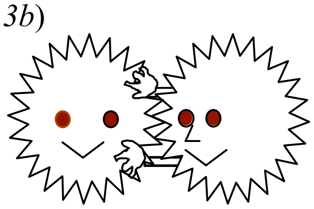


A B

S→A→B

	Interventions		Outcome
1	S Off	A on	B still
2	S Off	A off	B still
3	S Off	B on	A still
4	S Off	B off	A still
5	S On	A on	B spins
6	**S On**	**A off**	**B still**
7	S On	B on	A spins
8	**S On**	**B off**	**A spins**

3b)

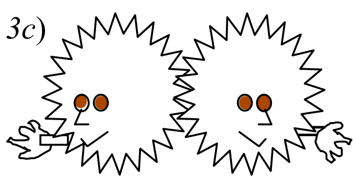


A B

S→B→A

1-4 as above			
	Interventions		**Outcome**
5	S On	A on	B spins
6	**S On**	**A off**	**B spins**
7	S On	B on	A spins
8	**S On**	**B off**	**A still**

3c)

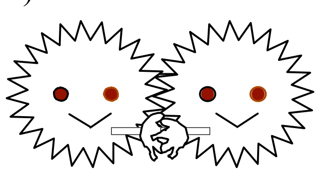


A B

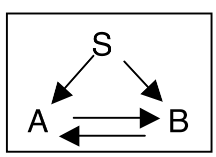
A←S→B

1-4 as above			
	Interventions		**Outcome**
5	S On	A on	B spins
6	**S On**	**A off**	**B spins**
7	S On	B on	A spins
8	**S On**	**B off**	**A spins**

3d)



A B



1-4 as above			
	Interventions		**Outcome**
5	S On	A on	B spins
6	**S On**	**A off**	**B still**
7	S On	B on	A spins
8	**S On**	**B off**	**A still**

Figure 4. Schematic of the gear toy