Inferring Hidden Causes

Tamar Kushnir (tkushnir@socrates.berkeley.edu), Alison Gopnik (gopnik@socrates.berkeley.edu), Laura Schulz (laurasch@socrates.berkeley.edu)
Department of Psychology, Tolman Hall, University of California
Berkeley, CA 94720 USA

David Danks (ddanks@ai.uwf.edu)
Institute for Human & Machine Cognition, University of West Florida
Pensacola, FL 32501 USA

Abstract

One of the important aspects of human causal reasoning is that from the time we are young children we reason about unobserved causes. How can we learn about unobserved causes from information about observed events? Causal Bayes nets provide a formal account of how causal structure is learned from a combination of associations and interventions. This formalism makes specific predictions about the conditions under which learners postulate hidden causes. In this study adult learners were shown a pattern of associations and interventions on a novel causal system. We found that they were able to infer hidden causes as predicted by the Bayes net formalism, and were able to distinguish between one hidden common cause and two hidden independent causes of the observed events.

Introduction

Causal reasoning is an important tool with which we make sense of relationships between objects and events in the world. Once we have a causal model of the world, we can make predictions, generate explanations and reason about the consequences of possible actions. How do we go about acquiring such models? Because the data available to our senses are often imperfect and incomplete, our causal learning system has to be flexible about the kind of information it requires. First of all, many of the causal relations we observe have no obvious spatio-temporal connection. We must, and indeed we do, learn about causation by observing associations, and psychological research has described this learning process in detail (Cheng 1997; Gopnik, Sobel, Schulz & Glymour, 2001; Shanks & Dickinson, 1987). In addition, our causal learning system should be able to postulate new objects/events without observing them directly. This is important both for discovering new observable causes and reasoning about phenomena that cannot be directly perceived. How we learn about hidden entities from observable ones is a topic that has not been given much attention in psychological research, and is the focus of this investigation.

There is a wealth of evidence that adults and even very young children learn and reason about unobserved causes. We appeal to unobserved mental states to explain human behavior (Gopnik & Wellman, 1994; Ross 1977; Wellman, 1990). Unobserved causes underlie our representations of basic categories (Gelman & Wellman 1991; Murphy & Medin, 1985). We also reason about physical forces that we cannot see (Shultz, 1982; Schlottmann & Surian, 1999). Scientific research is entirely devoted to explaining observed events by appeal to hidden theoretical entities (Gopnik & Melzoff, 1997). Often, as technology advances, these entities turn from hidden to observable, though it is their theoretical existence that prompts us to look for them in the first place.

A perfect example of this is a classic study in the field of epidemiology. In the 1850s, there were a series of cholera epidemics in London. In order to test a theory that cholera was a waterborne disease, a doctor named John Snow spent almost a decade meticulously recording where cholera victims lived, and which of several companies was supplying them with water. He was able to confirm his theory by using this statistical information to eliminate all other possible causes, such as those related to poverty, gender or occupation. It was not until much later that direct microscopic evidence confirmed what he was able to figure out using indirect evidence alone (Snow, 1855).

Snow's account demonstrates how powerful the combination of data and good scientific intuition is for learning about hidden causes. However, what we call "good scientific intuition" for interpreting data has traditionally had no formal account. Recently, though, a convergence of statistical models from several fields (machine learning, epidemiology, social science, statistics) has resulted in a formal account of causal learning and inference known as causal graphical models, or causal Bayes nets (Pearl, 2000; Spirtes, Glymour & Scheines, 1993). The successes of these models in aiding scientific research have prompted a recent effort in cognitive science to use causal Bayes nets to model human causal reasoning (Glymour, 2001; Gopnik, Glymour, Sobel, Schulz, Kushnir & Danks, in press; Steyvers, Tenenbaum, Wagenmakers & Blum, in press; Tenenbaum & Griffiths, 2001; Waldmann & Hagmayer, 2001).

Bayes nets represent joint probability distributions in their simplest form by exploiting the set of conditional independence relations among the variables (Jordan, 1998). Causal Bayes nets apply this theoretical framework to sets of variables that are causally related. Algorithms have been developed along these lines that use the conditional independence relations from a combination of observed associations and interventions to infer causal structure. Besides accounting for well-known findings in cognitive psychology on the role of observational data in learning causal relations (see Gopnik et al, in press), these models provide the first formal account of the role of interventions in causal learning and inference (Pearl, 2000; Spirtes et al, 1993).

So far, there is evidence that both adults and young children can learn the causal structure of a set of observed events using patterns of conditional probability in a manner consistent with the Bayes net formalism. Both children and adults can use information about conditional independence and dependence to discount (or “screen off”) spurious associations in favor of true causes (Gopnik et al, 2001; Cheng 1997; Shanks & Dickinson, 1987; Spellman, 1996). Recently, several researchers (Gopnik et al, in press; Schulz, 2001; Lagnado & Sloman , 2002, Steyvers et al, in press) have also demonstrated that adults and young children can use information from interventions to learn the causal relations between observed variables. For example, Schulz (2001) showed 4-year-olds and adults two objects (A & B) that moved simultaneously without touching (no spatio-temporal cues), and asked them to determine which object caused the movement. Participants then saw that intervening on object B did not result in the movement of object A. Both children and adults inferred that object A was the cause. The same pattern of movement (A & B together, then B alone) without an intervention resulted in chance responding.

The formal story, according to the theory of interventions on causal graphs (see Spirtes et al, 1993; Pearl, 2000) is this: Before the intervention was performed, participants had information about P(A|B) and P(B|A), namely that both were equal to 1. This, however, is very different from P(A|do(B)) (where do(*X*) notes an intervention on *X*). The intervention do(B) sets B to a fixed value determined by the intervener, thus effectively removing all other causes of B in the system (represented by removing the arrow from A to B). If A is a cause of B, then $P(A|do(B)) \neq P(A|B)$. If A is an effect of B or is independent of B, then $P(A|do(B)) = P(A|B)$. Since the former is true in this case, the learner should conclude that A causes B.

In another condition, participants (both children and adults) saw three objects (A, B & C) moving together simultaneously and were asked which was the cause of movement. An intervention on object A didn’t result in the movement of either B or C. An intervention on C left A & B unmoving. Children as young as 4 came to the (formally) correct conclusion that B was the cause. Again, the same pattern of associations without interventions resulted in chance responding.

If object B were hidden from view, the Bayes net learner would infer that a hidden common cause for A & C must exist given the same pattern of interventions as in the above example. Since the interventions on A & C are independent of each other, then only a common cause of A & C can produce the dependency between them that was initially observed. If that cause is hidden, then it must be inferred given the Bayes net modeling assumptions (see Gopnik et al, in press for a formal analysis). Moreover, in addition to simply inferring that there is an unobserved variable, learners should also be able to infer that this unobserved variable is a common cause of A and C, and to differentiate this hypothesis from the hypothesis that A and C are the result of two independent unobserved causes.

There is some preliminary evidence that children can infer an unobserved cause when the causal relations between the objects are deterministic (Gopnik et al. in press). A stronger test of the hypothesis would be to see if learners can also do this when the relations are probabilistic, and can differentiate common and independent unobserved causes. However, before asking whether children can infer a hidden cause in the above scenario, we need to investigate whether adult learners will do so -- a question that has never been investigated. In the following studies, we show that adults can infer a hidden cause from conditional probabilities without temporal or mechanistic cues, and can differentiate common and independent unobserved causes. In particular, we will show that, as predicted by Bayes net models, a combination of observations and interventions can lead to such a conclusion – even when each alone is insufficient to learn the correct causal structure.

Experiment 1

In this experiment, we showed participants two objects, colored balls on sticks, moving simultaneously up and down due to being placed in a “stick-ball machine.” The stick-ball machine could have one of several possible mechanisms operating within it on any given trial. In one trial, the evidence presented was similar to that in the above experiment (Schulz 2001, condition 2). Participants observed balls A & B move together. They then observed interventions on ball A and on ball B, neither of which resulted in the movement of the alternate ball. If the Bayes net account is correct, this should lead to the conclusion that one hidden mechanism causes both balls to move.

As one comparison, we presented participants with the identical intervention information but different initial observations – balls A and B moved independently most of the time. This observational information should lead to the conclusion that there is no association between A and B, and thus they are not caused by a common mechanism.

Because the apparatus had a hidden mechanism, we performed another control to insure that participants did not favor an unobserved causal explanation when an observed cause could account for the movement. In this condition, we constantly intervened on one of the balls, which should lead to the conclusion that it is the cause of movement.

Another possibility is that people have a preference for observed causes over unobserved ones. In order to control for this, we made the mechanism probabilistic by demonstrating to participants that the balls were only causally effective 67% of the time. This way, failed interventions could be interpreted as having failed by chance, thereby leaving open the possibility that the observed balls could still be the causes of movement.

Method

Participants: Participants were 48 undergraduates recruited from the research participation pool at an urban university.

Materials: The stick-ball machine (shown in figure 1) was a 3' x 1' x 1' wooden box with two holes at the top and an open back which faced the experimenter and was hidden from participants. Two colored rubber balls attached to wooden sticks could be placed in the holes. The mechanism in the box allowed the experimenter to move the stick-balls up and down either together or one at a time.

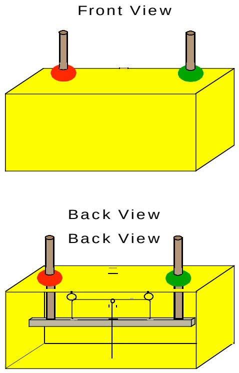

Figure 1: The stick-ball machine

Procedure: Each group of participants was seated facing the two experimenters so that they could only see the front of the stick-ball machine. One experimenter narrated the task and performed interventions while the other operated the machine. Participants were told that there was a mechanism behind the machine that could change from trial to trial, and that their job was to figure out the mechanism that made the stick-balls move on each trial. They were also told that the mechanism "almost-always" worked. This allowed for the possibility that balls could fail to move by chance. The experiment included one familiarization trial and three test trials. On each trial two new stick-balls of different colors were introduced. Each stick-ball was given a name based on its color and this name was used to refer to the stick-ball throughout (eg This is Reddy and this is Bluey). The stick balls could be moved by a hidden machine operator from behind (observations) or the experimenter could move them by pulling on the top of the stick from above (interventions). Order of trials was counterbalanced, with the familiarization trial always first. The types of movement (interventions and observations) on each trial were intermixed. The interventions were counterbalanced by side so that no ball (right or left) was always intervened on first.

Familiarization trial: On this trial alone the experimenter explicitly told participants that ball A almost always caused ball B to move. This was then demonstrated by showing both balls moving together four times and ball A moving alone twice.

1. Common unobserved cause: The stick-balls moved together four times. The narrator intervened on ball A twice and each time ball B didn't move. The narrator intervened on ball B twice and each time ball A didn't move.

2. Independent unobserved causes: The stick-balls each moved separately twice, and they moved together once. The narrator intervened on ball A twice and each time ball B didn't move. The narrator intervened on ball B twice and each time ball A didn't move.

3. One observed cause: The narrator intervened on ball A six times. Four of those times, both ball A and ball B moved. The remaining two times ball A moved and ball B didn't move.

After each trial, participants were given an answer sheet with a choice of four possible mechanisms: A causes B, B causes A, one hidden cause or two hidden causes (see figure 2) and asked to circle the one that was operating on that trial.

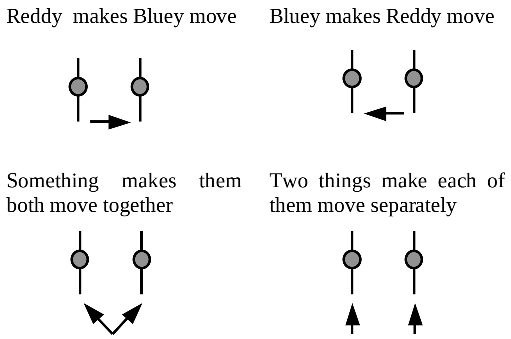
Reddy makes Bluey move
Bluey makes Reddy move
Something makes them both move together
Two things make each of them move separately

Figure 2: A sample answer sheet for one trial.

Results & Discussion

The results confirmed the predictions of the Bayes Net model. Overall, participants' responses matched the normative response for each type of trial. Table 1 shows the percentage of participants that chose each picture in the three test trials. The majority response for each trial is in boldface. In trial 1 (common unobserved cause), 63% chose the common cause picture. In trial 2 (independent unobserved causes) 96% chose the separate causes picture. In trial 3 (One observed cause), 65% chose "A makes B move," where A was the ball that the experimenter intervened on. All three response distributions are significantly different from chance (χ^2 = 26.38, 40.33, 42.50 respectively, all $p < .001$).

Participants' responses to trial 1 (common unobserved cause) were compared with their responses to the two other types of trials. Participants were more likely to pick the common cause picture in trial 1 than in trial 2 (McNemar's test, $p<.001$) or in trial 3 (McNemar's test, $p<.001$).

Table 1: Percentage of responses in each of the test trials in Experiment 1.

	1 - Common unobserved	2 - Independent unobserved	3 - One observed *
A causes B	0	0	**65**
B causes A	2	0	6
Common cause	**63**	4	8
Separate causes	35	**96**	21
χ^2 (df)	26.38 (2)**	40.33 (1)**	42.50 (3)**

*Intervention on ball A
**p < .001

The data show that adult learners inferred a hidden common cause when they observed that two events were associated with each other, but the association was not preserved when the experimenter intervened to cause either event. If the events are not associated to begin with, adults attribute their occurrence to independent hidden causes, regardless of the fact that they witness the same pattern of interventions. Also, participants clearly inferred that one observed event cause the other when it was appropriate to do so, rather than defaulting to some hidden mechanism. Interestingly, a portion of the participants seemed to default to the "separate causes" response – it was the second most frequent response in both trials 1 and 3. This may have to do with the fact that it is the safest response (could always be true) though not the most parsimonious one.

Experiment 2

In experiment 2 we explored whether adults would make similar judgments when they saw the same pattern of associations between the objects, but those patterns were not due to interventions. The Bayes net models should generate different results in these two cases. Other accounts, such as a simple associationist account, should not distinguish between observations and interventions in this way. In this experiment participants were shown the same hidden common cause task as in Experiment 1. They were also shown the same pattern of events without any interventions. Instead of intervening, the experimenter pointed at each object as it moved by itself. The pointing made each stick-ball salient in exactly the same way that the intervention did, and was a very similar perceptual event to direct intervention. However, in this case, since participants observe that the movement of A & B is associated only half of the time, they should be just as likely to infer two unobserved causes as one common unobserved cause.

Method

Participants: Participants were 24 undergraduates recruited from the research participation pool at an urban university.

Materials: The stick-ball machine and stick balls were the same as in experiment 1.

Procedure: Participants were introduced to the stick-ball machine in the same manner as in experiment 1. After the familiarization trial, there were two test trials, counterbalanced across groups of participants.

Common unobserved cause: The stick-balls moved together four times. The narrator intervened on ball A twice and each time ball B didn't move. The narrator intervened on ball B twice and each time ball A didn't move.

Pointing control: The stick-balls moved together four times. The narrator pointed at ball A twice as it moved alone. The narrator pointed at ball B twice as it moved alone. Pointing always began slightly after the movement (to rule it out as a cause).

After each trial, participants were asked to circle the mechanism behind the machine on the answer sheet (same as experiment 1).

Results & Discussion

As in Experiment 1, participants' responses matched the predictions of the Bayes net model for each trial. Table 2 shows the percentage of participants making each type of response. In trial 1, 67% of participants chose the common cause picture (replicating the findings in Experiment 1). In trial 2, 79% chose the separate causes picture. Participants were more likely to pick the common cause picture in trial 1 than in trial 2 (McNemar's test, p<.01).

This experiment again shows that, with the right combination of observations and interventions, adult learners inferred an unobserved common cause for the two events. Without interventions, adult learners were most likely to view the identical pattern of events as arising from separate hidden mechanisms.

Table 2: Percentage of responses in each of the test trials in Experiment 2.

	1 - Common unobserved	2 – Pointing Control
A causes B	0	0
B causes A	0	4
Common cause	**67**	17
Separate causes	33	**79**
χ^2 (df)	26.38 (2)**	40.33 (1)**

*Intervention on ball A
**p < .001

General Discussion

In both experiments participants were able to infer an unobserved common cause, and to distinguish unobserved common causes from unobserved independent causes. Neither identical data from observed associations without interventions (Experiment 2) nor identical interventions with different observed associations (Experiment 1) lead to the same conclusion. This investigation showed that, given certain patterns of evidence, adult learners will infer unobserved causes for observed events. In order to do this, learners relied on the crucial distinction between observed associations and interventions. Causal Bayes nets are the only formal models that currently make this distinction and that provide algorithms for how causal structure learning takes place based on both types of evidence.

However, the undergraduate participants in this experiment had extensive experience of causal inference, and often had some explicit tuition in causal reasoning. For this reason, it is important to ask whether even young children, with relatively little prior experience, would infer hidden causes under the same circumstances. Such evidence would at least suggest that a general learning mechanism is more likely than a rule based on years of experience.

Another possibility is that adults only infer hidden causes when they are explicitly presented as options. In this experiment, participants were given pictures of mechanisms with either one or two unobserved causes in them. Further research is needed to investigate other circumstances under which people will spontaneously infer a hidden cause without being given any explicit cues to do so.

Acknowledgements

We would like to thank April Davis for help with data collection, and Clark Glymour and Christine Schaefer for helpful discussions and comments.

References

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

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., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (in press). Theory formation and causal learning in children: Causal maps and Bayes nets. *Psychological Review.*

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

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 covariation. *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.

Jordan, M. (Ed.) (1998). *Learning in graphical models.* Cambridge, MA: MIT Press.

Lagnado, D. A., & Sloman, S. (2002). Learning causal structure. *Proceedings of the 24th annual meeting of the Cognitive Science Society.*

Murphy, G. L. & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review. 92(3)*, 289-316.

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

Ross, L. (1977) The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed), *Advances in experimental social psychology, Vol 10,* (pp. 174-220). New York: Academic Press.

Schaefer, C & Gopnik, A. (2003) Causal reasoning in young children: The role of unobserved variables. Poster to be presented at the biennial meeting of the Society for Research in Child Development.

Schlottmann, A. & Surian, L. (1999). Do 9-month-olds perceive causation-at-a-distance? *Perception, 28,* 1105-1113.

Schulz, L. E. (2001). *"Do-calculus": Adults and preschoolers infer causal structure from patterns of outcomes following interventions.* Paper presented at the 2001 meeting of the Cognitive Development Society, Virginia Beach, VA.

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* (Serial No. 194).

Snow, J. (1855). *On the Mode of Communication of Cholera.* London: John Churchill.

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.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (in press). Inferring causal networks from observations and interventions. *Cognitive Science.*

Tenenbaum, J, & Griffiths, T. L. (2001). *Structure learning in human causal inference.* Proceedings of the 2001 Neural Information Processing Systems Conference.

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

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