13

Data-Mining Probabilists or Experimental Determinists?

A Dialogue on the Principles Underlying Causal Learning in Children

Thomas Richardson, Laura Schulz, & Alison Gopnik

This dialogue is a distillation of a real series of conversations that took place at that most platonic of academies, the Center for Advanced Studies in the Behavioral Sciences, between the first and third \edq1\ authors.\edq1\ The second author had determinist leanings to begin with and acted as an intermediary.

Determinism, Faithfulness, and Causal Inference

Narrator: [*Meno and Laplace stand in the corridor of a daycare, observing toddlers at play through a window.*]

Meno: Do you ever wonder how it is that children manage to learn so many causal relations so successfully, so quickly? They make it all seem effortless. A 16-month-old of my acquaintance got my cordless phone to do things I didn't know it could, and very speedily I might add.

Laplace: Yes . . . not only do they manage to sidestep metaphysical questions, they also seem quite able to make do without randomized controlled experiments and with tiny sample sizes.

Meno: Leaving aside the question of how children learn for a while, can we agree on some basic principles for causal inference by anyone—child, adult, or scientist?

Laplace: I think so. I think we can both generally agree that, subject to various caveats, two variables will not be dependent in probability unless there is some kind of causal connection present.

Meno: Yes—that's what I'd call the weak causal Markov condition. I assume that the kinds of caveats you have in mind are to restrict this principle to situations in which a discussion of cause and effect might be reasonable in the first place?

Laplace: Absolutely. I don't want to get involved in discussions of whether X causes 2X or whether the monthly changes in height of my elder son are causing the monthly changes in height of my younger son.

Instead, let's consider a real, practical, and indeed, sometimes lifesaving kind of causal inference—the kind of inference we make in scientific medicine. From that perspective, a discussion of cause and effect starts from the assumption that there is a population of units (say, patients in a clinical trial) and a set of treatments (say, drug vs. placebo). For each potential assignment of a treatment to a unit, there is a well-defined outcome that we could, in principle, discover by assigning that treatment to that unit. If we could simply systematically give all the units each kind of treatment and observe the outcomes, we could discover which treatments caused particular outcomes. Causal inference is difficult because we can usually only find out what would happen to a particular unit under one treatment. We can't observe the counterfactual—what would have happened if we'd chosen another treatment. We can't observe whether for example, a particular patient would also have recovered if he had taken the placebo rather than the drug.

Narrator: *[See J. Neyman, Sur les applications de la thar des probabilities aux experiences Agaricales: Essay des principle, 1923, as excerpted in English in Dabrowska and Speed, 1990; Rubin, 1974.]*

Meno: Why do you say "usually"? Isn't it logically impossible to see the same unit under two treatments?

Laplace: Well, in some situations it may be reasonable to assume that the effect of applying the treatment to the unit is sufficiently short-lived that we can later apply another treatment to the same unit and then compare the outcomes. The effect of applying the second treatment would be assumed to be the same as if it had been applied first.

One would typically do this with a set of units and randomize the order of treatments. For instance, if you wanted to see if a particular drug had short-term side effects, you might use a within-subjects design—give each patient the drug and then, after a pause, give the patient the placebo (and vice versa).

On other occasions, we might think that distinct units were sufficiently similar in all respects that their outcomes under the same treatment might be assumed to be identical for practical purposes. We assume, for instance, that a new patient is similar enough to the patients in our sample so that the drug will affect the patient in the same way.

Meno: But, how can you make that assumption without a population of identical twins? Surely, any such assumption will be untestable; you can't escape the fundamental problem of causal inference so easily.

Laplace: Yes, of course. . . . If you like to drink your skepticism neat, then we might ask how we know that the future will conform to the past and, failing any kind of satisfactory answer, abandon the whole epistemological roadshow.

I mention these within-subjects or crossover experimental designs because I think they may be relevant to what our toddlers are doing.

Meno: Let's come back to that, but first I want to follow up on the weak causal Markov condition. How precisely do you see it as justifying causal inference in your medical example?

Laplace: If there is dependence between the treatment assignment and the outcome of the experiment, then according to the weak causal Markov condition, (a) treatment is causing outcome, (b) outcome is causing treatment, or there is a common cause of outcome and treatment . . . (or some combination thereof) (Figure 13-1).

Meno: I see, and if treatment were randomized, then that means that (b) and (c) are ruled out because the treatment a patient receives is determined by

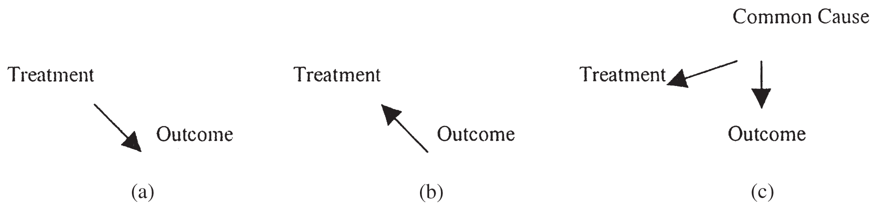

FIGURE 13-1 (a) Treatment causes outcome; (b) outcome causes treatment; (c) treatment and outcome have a common cause.

the randomizer (e.g. the flip of a coin) rather than the patient's (potential) response to the drug or any common cause (e.g., the doctor's beliefs about the patient).

Laplace: Yes—in fact, if there is a time order (i.e., outcome does not exist) prior to the treatment being assigned, then we might rule out (b) a priori, but randomization is required to eliminate (c) as a possibility.

Meno: When we say *dependence*, what do we mean? Presumably we don't mean that the outcome is always entirely determined by the treatment?

Laplace: The case in which outcome is determined by treatment is an important special case, and we return to it in a minute, but in describing the weak causal Markov condition we have in mind probabilistic dependence: The distribution of outcomes in the treatment group and in the control group are different. Naturally, this only makes sense if we have some population or hypothetical population of units.

Meno: I see, so this condition is only supposed to apply to causation between variables, not between individual events.

So we agree that, subject to certain caveats, dependence, whether probabilistic or deterministic, implies the presence of a causal connection. What about the reverse? Suppose that we do not observe any dependence between two variables, what may we then conclude? Is it reasonable to conclude that treatment does not cause the outcome?

Laplace: That assumption is an instance of the assumption known as the causal faithfulness condition. But, here the way is less straightforward. For instance, suppose that we give a group of patients a treatment to reduce the amount of insulin in the body (e.g., by changing it into some other form), but that the body responds by generating additional insulin, exactly matching the amount removed by the drug.

Meno: I see. In the situation you describe insulin level remains unchanged regardless of whether the patient has taken the drug or the placebo, so we might think that the drug had no causal effect. However, if we were able to prevent the body from generating additional insulin, then the drug would have an effect?

Laplace: Yes. For instance, it would have an effect in a population of diabetics. This is an instance of the point made about causal effects defined relative to a comparison of two (counterfactual) outcomes corresponding to different treatments. The fact that an intervention does not change any individual outcome does not mean that in the context of a second intervention (e.g., destroying the body's capability to produce insulin) the intervention will not change any outcomes.

Meno: But, is this a real problem? Within the population of nondiabetic patient treatments, the drug would, in fact, have no effect on the outcome for any individual, so for practical purposes, it is just as if it had no causal influence at all.

Laplace: Agreed. However, such cases may present problems if we want accurate representations of causal systems. These representations are useful because they allow us to say not only what has happened or even what will happen, but also what would happen if we made new interventions. For instance, suppose we want to represent a causal system with a directed graph. If the presence of a directed path is taken to indicate that there is an effect, and the graph is intended to represent the effects of simultaneous interventions on more than one variable, then one is faced with a choice between a graph that disobeys the faithfulness condition (e.g., by including an edge from the drug to the insulin level in the body) and a graph that is faithful (e.g., by omitting an edge from the drug to insulin level) but does not correctly predict the results of multiple interventions.

There are other situations for which the distribution of outcomes may not change under different treatments, so that there is no dependence that may be observed in a randomized experiment, but at the same time each person's pair of counterfactual outcomes are different under the different treatments. For instance, imagine a treatment that switches a person's gender. If the treatment and control groups initially have equal numbers of men and women, then the proportion of females in the treatment and control groups will be the same at the end of the experiment.

Meno: But the treatment would have had a noticeable effect on the individuals in the treatment group.

Laplace: That is if you are willing to assume that the people in the treatment group would not also have spontaneously switched gender had they been in the control group. (Isn't that the kind of assumption you warned me about?) Also, note that the effect would only be "evident" if you knew the gender the individual would have had if untreated. For instance, suppose there is a treatment that has this effect in the first few days after conception, before it is possible to determine the child's gender. In this case, you would not be able to observe any change; hence, there would be no way to observe the effect directly.

Meno: I see; so, every individual's outcome would be different under treatment and control, yet there would be no way to discover this from looking at the distribution of outcomes in treatment and control.

Narrator: For a human population, the ratio of males to female births is not equal; hence, given a large enough sample size, one would still be able to detect the effect.

Meno: These scenarios still seem slightly outlandish. They appear to me to be like causal illusions: Like a masterful *trompe l'oeil*, our initial impressions of the situation are incorrect, but on further inspection we can see what is really going on. Might we not agree that, absent other information, we might adopt as a working hypothesis that the absence of dependence implies the absence of a causal relationship?

Laplace: I'm fairly comfortable with that. There are technical arguments that may be advanced for such a principle: If there is independence between treatment and outcome, although treatment causes outcome, then several causal pathways must "cancel out," and this is unlikely to happen by chance. However, there is one situation that may often arise in which faithfulness routinely fails. Faithfulness often fails if the causal relationships are deterministic.

Meno: Let me see if I understand the distinction that you have in mind. In general, if a variable *X* has a causal effect on a variable *Y*, then knowing the value of *X* may inform us about the distribution of possible values of *Y*, but it will not tell us which specific value *Y* will take on. However, if the relation between *X* and *Y* is deterministic, then knowing *X*, we know the value taken on by *Y*.

Laplace: Exactly. Consider, for example, a room with an energy-saving lightbulb connected to a light sensor. The bulb only goes on when the room grows dark. Now, consider the "treatment" of opening versus closing the blinds in the room, the outcome being whether there is light in the room. It is easy to see that if the causal relationships are deterministic (i.e., the light sensor and lightbulb never fail), then pulling down the blind has no effect on the outcome variable. So, using the principle that absence of dependence indicates absence of causation, we should conclude that opening the blind has no causal effect on the light in the room.

Narrator: It is important here that the outcome, whether the light is on or off, is binary. If we had a continuous measure of the quality of light in the room, then the relationships would not be deterministic (clouds, streetlights, etc.).

Meno: Isn't this simply the scenario of the insulin-lowering drug mentioned?

Laplace: Yes and no. It is insofar as we have two mechanisms canceling one another. There is a difference, however, in that because our outcomes are determined, there is less room for detecting change by slightly perturbing the scenario. By contrast, if we allowed the relationships to be probabilistic, so that the light sensor and bulb sometimes failed, then it would be easy to detect an effect: Simply count the proportion of time there is light when the blinds are open versus the proportion of the time there is light when the blinds are closed. If the sensor or bulb ever fail, then the latter proportion must be smaller.

Narrator: This tacitly assumes that the probability of failure is unrelated to whether the blinds are open or closed.

Meno: It is unless you happen to do the experiment in an environment with permanent sunshine or darkness, such as the poles or anywhere on Mercury. I see you are arguing that causal relationships that are not deterministic are more likely to obey the causal faithfulness condition. Ironically, the "noise" in a probabilistic system may help us understand more about how the system works than we can understand in the apparently simpler deterministic case.

Laplace: Absolutely. In fact, this point becomes even clearer and more pressing if we consider contexts with more variables. So far, we have only been considering a single candidate cause (the treatment) and a single effect (the outcome). But, of course, causal structures may be a lot more complicated than this (Figure 13-2).\edq2\ \edq2\

Meno: For those contexts, don't we have to assume the strong causal Markov condition?

Laplace: Yes. Let us review this condition. We need a few more concepts first. If X causes Y, let us say that X is a (causal) *parent* of Y, and Y is a (causal) *child* of X. Similarly, let us say *U* is a (causal) *ancestor* of *V* if there is a sequence of variables starting with *U* and ending with *V* such that each variable in the chain is the parent of the next. If *U* is a causal ancestor of *V*, then we will say that *V* is a (causal) *descendant* of *U*. Finally, say that a set of variables is *causally sufficient* if any common causal ancestor of two or more variables is included in the set.\edq3\ \edq3\

Meno: So, in these terms, the strong causal Markov condition states that, in a causally sufficient set of variables, if we know the values taken by the parents of a given variable X, then learning the values taken by other variables that are not descendants of X tells us nothing about (the distribution of) X itself.

Laplace: Yes, that is exactly right. In fact, for systems in which all causal relationships between parents and children are linear, the weak causal Markov condition implies the strong condition.

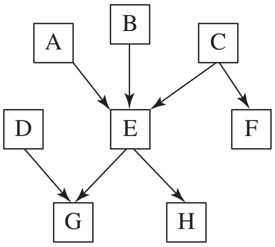

FIGURE 13-2 A, B, and C are parents of E; G and H are children of E; A, B, C, D, and E are ancestors of G; E, G, and H are descendants of B. The set {A, B, E, G} is causally sufficient; the sets {E, F} and {F, H} and {F, G} are not causally sufficient. According to the strong causal Markov condition, G is independent of A, B, C, F, and H given E and D.

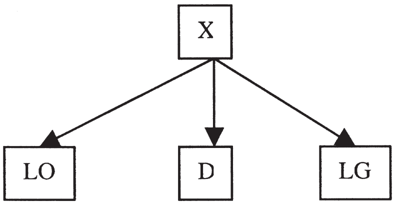

FIGURE 13-3 A simple garage door opener: X is the opener; LO is a light on the opener; D is the door opening; LG is a light in the garage.

Meno: In this context, the causal faithfulness condition now asserts that every independence relation is a consequence of the strong causal Markov condition applied to the true causal graph. Only independence relations that follow from the causal Markov condition will appear in the data. If some other independence relation appears, then the causal faithfulness assumption has been violated.

Laplace: Again, exactly right. With these concepts in hand, we are now in a position to discuss the problems brought about by deterministic relations. Consider a simple situation with a common cause: Pressing the garage door opener X leads to a light blinking on the opener LO, the door opening D, and a light going on in the garage LG (Figure 13-3).

Meno: So, the Markov condition tells us that if we know whether the opener X was pushed, then D, LO, and LG are irrelevant to one another. In technical parlance, D, LO, and LG are mutually independent conditional on X.

Laplace: Correct, but here is the problem: Suppose that the relationship between the door opener being pressed and the light on the opener LO is deterministic, so that this light goes on when and only when the opener is pressed. It is now easy to see that if I see the opener light, then I immediately know that the opener has been pressed even if I have not observed this directly. But, it then follows that the door opening D and the garage light LG are independent given only knowledge of the light on the opener LO. This independence does not follow from the causal Markov condition: If the relationship between X and LG were not deterministic, then this extra independence would not hold, yet the causal graph would be the same.

Meno: So far, I follow. Suppose I were to try to make inferences about causal structure from conditional independence, assuming the causal Markov condition and, contrary to fact, the causal faithfulness condition held? All such procedures use the fact that under these conditions, if *X* is a causal parent of *Y*, then *X* and *Y* will always be dependent regardless which other variables we know (or condition on). I do not see that causing immediate problems here because this extra independence of D and LG given LO simply tells us that there can be no edge between D and LG, which is correct.

Laplace: Yes, but there are more unfaithful independence relations here. We already know that LO and LG are independent once we know X. But, if X and LO are logically equivalent, then LG and X are also independent once we know LO because we know LO if and only if we know X. Likewise, D and X are independent once we know LO.

Meno: I see; so, in fact we will end up with no edges except the one between X and LO.

Laplace: I'm afraid so.

Meno: I see now why those proposing the causal faithfulness condition as an inferential principle for learning causal structure explicitly exclude deterministic contexts. In those contexts, the faithfulness assumption will often (in fact, usually) be false.

Narrator: For example, Spirtes, Glymour, and Scheines (1993) state: "We will not consider algorithms for constructing causal graphs when such deterministic relations obtain, nor will we consider tests for deciding whether a set of variables X determines a variable Y" (p. 57).

Determinism in Children's Causal Inferences

Meno: Can we return to children's learning?

Laplace: By all means.

Meno: Inspired by my discussion with Socrates about geometry, I also have concluded that empirical developmental psychology is the best way to answer epistemological questions. So, I have been reading the developmental literature and find that several authors have put forward the suggestion that children learn causal structure by "implementing" inference algorithms that rely on the Markov and faithfulness assumptions.

Laplace: I think I have heard of this. Can you give me an example?

Meno: In one set of experiments, children were shown a device that was called a *blicket detector*, a box with the capability of emitting a sound when blickets were placed on it.

In these experiments, objects of two different types, let us say A and B, were placed on the detector. The children observed the detector making a noise in certain configurations and were then asked various questions.

Laplace: I think I follow.

Meno: In one experiment, 3- and 4-year-old children were divided into two groups. One group, in the one-cause condition, were shown the following sequence of events:

A on detector with noise

B on detector without noise

A and B on detector with noise (repeated twice)

The second group, in the two-cause condition, were shown the following:

A on detector with noise (repeated three times)

B on detector without noise (once)

B on detector with noise (repeated twice)

In each case, the children were then asked if each object was a blicket. In the one-cause condition, children said that Object A was a blicket more than twice as often (96% vs. 41%). In the two-cause condition, they were roughly equally likely to say that A and B were blickets (97% and 81.5%, respectively). In another version of the experiment, children were asked which of the two objects was a blicket. The results were similar.

Narrator: [*See Gopnik, Sobel, Schulz, and Glymour, 2001, Experiment 1.*]

Laplace: I think I follow the logic that the children might have used, but I do not see the connection to Markov and faithfulness.

Meno: Isn't it obvious? The children took the frequencies observed in the data and observed that in the one-cause condition

$$0 = P(\text{Noise} \mid \text{not A and B}),\ P(\text{Noise} \mid \text{A and not B}) = P(\text{Noise} \mid \text{A and B}) = 1$$

Hence, the presence of A makes Noise more likely, and Noise and B are independent given A.

If the children also believe that the detector does not make a noise without a trigger, so $P(\text{Noise}\mid \text{not A and not B}) = 0$, then Noise and B are also independent given not A. Hence, by the faithfulness condition we may conclude that B is not a cause of Noise. Because A and Noise are dependent, by the Markov condition they are causally connected: It could be that the fact that A is on the detector causes the noise, that the noise causes A to be on the block, or that there is some common cause of both events. However, both the description of the blicket detector and the fact that A is placed by an investigator suggest that the placement of A is an external intervention (i.e. it is *exogenous*), hence we may conclude that A is a cause of noise.

Laplace: I see; if the placement of the block near the detector were not performed by a human (e.g., it was a consequence of some larger mechanism), then we might conclude that there was some common cause at work.

Meno: Exactly. You still look skeptical.

Laplace: I have several concerns with this argument. Broadly, I am not convinced that the formalism of probability theory needs to be invoked to explain the logic that is used here. After all the relationships between the detector and the blocks are deterministic, aren't they, at least in the one-cause task? Every time a blicket is placed on the detector, it goes off. Further, I think that if probability theory were used in the way that is suggested, then we would be less good at learning causal relationships than in fact we are. Third, I am skeptical about invoking the faithfulness assumption in this context because as an inferential principle I believe that it is incompatible with a belief that one is observing a simple deterministic system.

Meno: Please go on. What do you see as problematic about the use of probability theory.

Laplace: Were I a child, I would be hesitant about regarding the observed (relative) frequencies seen in such a small number of cases as representative of the probability that any of these events would happen in these conditions.

Meno: Why shouldn't one do so?

Laplace: Well, suppose that I first showed the following four outcomes:

Nothing on detector without noise

A on detector with noise

A on detector without noise

Nothing on detector with noise

Meno: So, from faithfulness I would conclude that A is not a cause of the noise because the probability of noise is independent of A: $P(\text{Noise} \mid \text{A}) = P(\text{Noise} \mid \text{not A}) = 1/2$.

Laplace: But, here is the problem. If you continue to apply the same logic, and I now tell you that I am going to place A on the detector, then before you see the outcome, you can conclude that you will believe that there is a causal relationship between A and the noise.

Meno: That seems like an absurd outcome. How does it follow?

Laplace: Well, if we place A on the detector and it makes a noise, then with that additional observation, according to the observed frequencies, $P(\text{Noise} \mid \text{A}) = 2/3$, while $P(\text{Noise} \mid \text{not A}) = 1/2$, so noise and A are dependent. Conversely, if we place A on the detector and it fails to make a noise, then $P(\text{Noise} \mid \text{A}) = 1/3$; $P(\text{Noise} \mid \text{not A}) = 1/2$, so again A and the detector are dependent. In fact, even before I show you any data, if you know how many trials you plan under each condition, you may be able to conclude that, if the observed frequencies are assumed to be representative, then there will have to be a causal connection. For example, if we plan an odd number of trials in some condition and assume that the observed frequencies in those trials are representative, then it will simply have to follow that the frequencies will indicate dependence.

Narrator: This assumes that the outcome is binary.

Meno: I see the problem. But, it is important to remember that we are merely observing the reasoning patterns employed by young children; there is no reason to assume that their inferences should abide by normative principles.

Laplace: Indeed. Psychologists have often documented our "irrational" belief in the "law" of small numbers.

\edq4\ Narrator: Tversky and Kahneman (1982)\edq4\ describe the law of small numbers as the belief "according to which even small samples are highly representative of the populations from which they are drawn."

Laplace: But, I think it is equally important to bear in mind that there may be more than one explanation for the observed behavior. Furthermore, the inferences made in the one- and two-cause condition experiment you described seem eminently reasonable—one would not expect an adult, even one attuned to statistical inference, to reason any differently. Surely you would agree that if we can explain children's behavior in these experiments without suggesting that they are systematically irrational, then that would be a preferable outcome?

Meno: Agreed. On reflection, I realize that when inferences about causal structure are made by machine learning algorithms employing faithfulness and Markov conditions, then these are based on databases containing hundreds, if not thousands, of cases.

Laplace: Yes—without further assumptions, any reasonable statistical procedure would be agnostic about the presence or absence of dependence from samples as small as those used by the children in the experiment you described.

Meno: Is it not possible that the children think it is safe to conclude that these small samples are representative because they are presented by a trusted adult figure in the person of the experimenter?

Laplace: One might think this, but I see two problems. First, if one really believed that one was observing a blicket detector that was not deterministic, then surely one must believe that it is outside the control of the experimenter to make it produce or fail to produce a noise on any particular occasion? In which case, there is no way for the experimenter to ensure that the data are representative: Although they might choose when to place or not to place the blocks, whether a noise is produced would not be entirely within the experimenter's control, so "representativeness" could not be guaranteed.

Hence, when the detector appears to behave indeterministically, the child would have to believe that the experimenter in fact controlled all aspects of the device and was creating the illusion of probabilistic data to (beneficently) reveal the true probabilistic properties of the device (that would pertain in the absence of the experimenter?). Although, of course, this is in fact how these experiments are conducted, I believe it would be an unusual 3-year-old who would adopt this as their working hypothesis.

Narrator: In principle, even with an indeterministic device, an experimenter might control the observed proportions by choosing to stop at an "appropriate" point. However, it would again be rather surprising if feelings of trust with respect to the experimenter were parlayed in such an elaborate manner: The sensitivity of frequentist statistical inferences to the choice of stopping rule was something that only became widely understood within the last 50 years.

Meno: I agree.

Laplace: Second, if it could be demonstrated that the children had such deep trust in the experimenter that they would consider this a plausible scenario, then one might seriously question the ecological validity of any inferences made about causal learning that took place in such a scenario.

Meno: Suppose I accept, as you appear to be arguing, that such small samples cannot be regarded as data on which one may reliably base conclusions about probabilities. You mentioned that you thought that the causal inferences made might be explained as normative without reference to probability theory. Can you expand on that?

Laplace: You read me correctly: From a statistical perspective, very little can be obtained from such small samples, other than the fact that certain combinations of events are possible (have nonzero probability). When I say *statistical perspective*, I mean if one starts out with the hypothesis that there are probabilistic causal relationships between the variables.

If this is the viewpoint with which we typically viewed the world, then it is somewhat surprising, perhaps even inexplicable, that most people would agree on the correct answer to the blicket question, at least in the one-cause scenario.

Meno: Some psychologists have made many of the same arguments and argue that therefore children's inferences must be constrained by a great deal of prior knowledge in a Bayesian way.

Narrator: [*See Tennenbaum and Griffiths, this volume.*]\edq5\

\edq5\

Laplace: If we are successful in providing a normatively rational explanation for children's inferences, then it will not be surprising if similar conclusions would be drawn by a hypothetical Bayesian agent. Some might even regard this as necessary.

However, I do not believe that this is the only explanation for such inferences, and indeed, I think that such an account leaves unresolved as many questions as it addresses.

Meno: Can you be more specific? Doesn't the Bayesian approach, in principle, provide a complete description of how to update one's beliefs?

Laplace: That it does. However, I would argue that a psychological theory should explain why people agree on the "correct" answer in the one-cause blicket experiment. The Bayesian approach does not prescribe any specific set of prior beliefs; in fact, one might expect different agents with different life experiences to have different beliefs. For example, Calvinist children might think that Divine intervention was responsible for the blicket detector making a noise at precisely the moment when the block was placed on it; Jungians might think it was just another instance of synchronicity at work.

Without an explanation regarding why we all have similar prior beliefs pertaining to such situations, the Bayesian account does not explain why we have the beliefs that we have.

Meno: I see. You contend that for any particular set of (posterior) beliefs we have after making some observations, a proponent of Bayesian inference might always concoct a hypothetical set of prior beliefs for us, which had we had them and had we been Bayesian would have resulted in the beliefs we have. But, because this could have been done for any set of posterior beliefs, the existence of such a set of prior beliefs in any given case does not constitute evidence that we arrived at our beliefs by Bayesian means.

Laplace: Indeed. Further, I believe that there is a computational issue that arises.

Meno: How so?

Laplace: It is a simple consequence of Bayes' rule that any hypothesis that is initially assigned probability 0 will continue to be assigned probability 0 regardless of the evidence that is observed.

Meno: I'm familiar with that, but how is it relevant here?

Laplace: The upshot is that if we do not wish to be unable to learn the true causal structure eventually, then we must ensure that we do not assign it probability 0 initially. Because the number of candidate causal structures increases quickly with the number of variables, an ideal Bayesian reasoner is faced with the prospect of keeping track of personal beliefs concerning hundreds, if not millions, of candidate hypotheses.

Meno: This is required if we are to be "ideal" Bayesians, but couldn't we be flawed Bayesians? For example, just entertaining seriously a few hypotheses, while regarding the remainder as having some small probability that we don't bother to update?

Laplace: We might, but again, as with the specification of prior beliefs, I believe that the "meat" of any such account lies in the details of how and why such an approximations scheme works in practice.

Narrator: Tennenbaum (personal communication) \edq6\ has proposed that a causal learner might approximate a (metropolis-Hastings) Markov chain Monte Carlo scheme for sampling from a posterior distribution. For example, a learner could keep in mind a single model but be continually switching

\edq6\

from one model to another even in the absence of any new data (but with the probability of switching determined by the data observed so far). At any given moment, the learner would have "in mind" only one model but would continually be changing this model, so that over an extended period the proportion of the time that the model is in mind would approximate the posterior probability. This is an intriguing idea, but it still requires that the learner have "access" to prior probabilities assigned to all possible models. (The issue of explaining/specifying priors also remains.)

Meno: I also see that most of us would consider it possible for us eventually to learn about a system with a structure that has features unlike anything we have ever seen, whereas an ideal Bayesian would need to have initially considered such a system at the outset. This reminds me of a discussion I once had with Socrates concerning the apparent problem of coming to learn anything new.

Narrator: [*See Plato,* Meno *80 D.*]

Laplace: Although I would not wish to rule out a Bayesian inferential approach per se,I believe that there is another, perhaps simpler, way forward: The apparent conflict between strong human agreement concerning the correct answer in the (one-cause) blicket and statistical agnosticism from small amounts of data suggests to me that most people do not adopt a statistical perspective on these problems, Bayesian or otherwise. Instead, they assume that they may simply be observing a deterministic system.

Meno: I can certainly see how that might simplify matters in the one-cause situation: The detector makes a noise if and only if Block A is placed on it; Block B is irrelevant.

Laplace: This is exactly what I had in mind, but nothing comes for free. In arriving at this conclusion, we have used the (weak) causal Markov assumption: The observation that the machine makes a noise after Block A is placed on it is interpreted as an intervention (or treatment), namely, placement of Block A then leading to an effect (or outcome), namely, the noise. The weak causal Markov condition invites us to conclude that there is a causal connection underlying the observed association. Under the hypothesis that placement of Block A is an (exogenous) intervention, this implies that A is the cause of the noise.

Meno: That tells us that A is a cause of the noise. But, how do we eliminate the possibility that B is also a cause? In the analysis, we described the investigators assumed faithfulness and assumed that the failure of B was representative, that is, that in general there was no dependence between B and the noise. How can we draw this conclusion without those assumptions?

Laplace: Rather than employ faithfulness, we simply employ another parsimony principle: Because no other causal relationships are required to explain the observed events, we assume that none are present.

Meno: Of course. Faithfulness may also be viewed as a parsimony principle in the sense that, as employed in some learning, it leads us to choose stochastic causal structures with fewer parameters. Here, in the one-cause condition we presume that there is no relationship between Block B and the noise, not because we have observed them to be statistically independent, but simply because we can explain all of the observed noises without assuming that B will lead the detector to make a noise.

Laplace: Absolutely right. From my point of view, we have not nearly enough data to say anything about the statistical independence of B and the detector.

Meno: But, wouldn't this sort of deterministic inference just collapse to good old-fashioned deductive logic? A is a blicket if and only if, if A was placed on the detector then the detector activates.

Laplace: Not exactly. Standard propositional logic does not include methods for dealing with causal interventions.

Meno: Let me see if I understand. When we have a set of propositions such as

Socrates is a man implies Socrates is mortal.

Socrates is mortal implies life insurance will not be free.

these implications are supposed to hold true always, whereas we wish to consider situations in which, via external intervention, some propositions are no longer true.

Laplace: Precisely. If there is a medical breakthrough, some (rich?) people's lives might be extended indefinitely. In such a case, the first proposition might no longer hold true—we might intervene to make Socrates immortal—but the second will no doubt continue to hold true. The propositions in a causal model are thus "modular" in the sense that an ideal intervention can override some propositions while leaving others intact.

\edq7\ Narrator: [*See Appendix; Pearl, 2000, Chapter 7; Schulz et al., submitted and this volume.*]\edq7\

Laplace: Thus, this "causal logic" has features that make it different from classical deductive logic. In particular, the difference between interventions and contingencies is inferentially crucial, but there is no such distinction in classical logic. Both children and adults seem to be appropriately sensitive to that distinction.

\edq8\ Narrator: *[See, for instance, Gopnik et al., 2004; Steyvers, Tenenbaum, Wagenmakers, and Blum, 2003; Lagnado et al., this volume.]*\edq8\

Meno: I can see that assuming that things are deterministic and applying a causal logic simplifies matters, but surely such an assumption is too stringent to be of much use in real life, when evidence is almost never deterministic. For that matter, the data presented in the two-cause condition are incompatible with a deterministic functional relationship. Remember that in the experiment children see A set off the detector 3/3 times, and B set it off 2/3 times. They conclude that both blocks are blickets. But, here on one occasion we have block B alone and a noise, and on another occasion we have block B alone and no noise. The same is true of other developmental experiments. In one of the puppet machine experiments, for example (Gopnik et al., 2004), one puppet almost always, but not always, makes the other puppet go.

Laplace: It is true that the two-cause condition is incompatible with a belief that whether a noise is produced is determined entirely by which blocks are present. However, by hypothesizing an additional unobserved variable, for example, a loose connection between the battery and the buzzer or the amount of pressure the experimenter applied when placing the object, one could easily construct a deterministic model that was compatible with the observed data. Then, you could make inferences about this deterministic model in the way I described.

Meno: But, is there any evidence to suggest that agents are willing to postulate the existence of such unobserved causes to "save" their belief in determinism. I seem to recall reading something.

Laplace: Indeed, there is. Consider the following experiment:

> Children are initially told that the experimenter likes to trick her confederate. The children then see a light, which is activated by a switch. There is also a ring on the light, which must be in place for the light to work.
>
> Children are divided into two groups. In the first group (stochastic causation condition), the confederate makes eight attempts to turn the light on by pushing the switch but is successful only on two occasions. In the second group (deterministic causation condition), the confederate is successful on all eight attempts. After seeing these eight trials, the experimenter then reveals a small key chain flashlight to the children, which has not been seen previously. Both groups of children are then asked to make it so that the switch does not work. Most of the children (15 of 16) in the stochastic causation group then reach for the flashlight even though they have never seen it do anything (one child chose the ring). By contrast, in the deterministic causation group almost all of the children (14 of 16) choose to remove the ring (two choose the flashlight).

Narrator: [*Schulz, Sommerville, and Gopnik, in press, Experiment 1.*]

Meno: Interesting. So, this indicates that the children in the stochastic causation group do not believe that "things just happen." They think that if the light is not working, there must be a (deterministic) explanation, and they are sufficiently invested in finding such an explanation that they are willing to hypothesize that an entirely new object has such powers.

But isn't it problematic that children are willing to attribute such hidden variables so easily? With enough hidden variables, we can represent any input-output function by an infinite variety of different graphs. Having too many causal answers is just as bad as having too few, and accurate causal inference will be just as difficult in these cases.

The children will be like Freudians or astrologers who can explain everything and therefore cannot really explain anything.

Laplace: But, the experiment points to more than that: Notice that most of the children in the deterministic causation group did not attribute causal powers to the flashlight. This suggests that the children do not hypothesize hidden variables in a promiscuous fashion. Rather, they do so parsimoniously and systematically. The events observed in the deterministic causation condition do not require any additional variables to be fully explained.

Meno: Still, it seems to me that there is a problem here. Let me return to your garage door example and consider the situation in which I do not directly observe whether you pressed the opener X, although we do observe the other three variables D, LO, and LG. There are then no deterministic relationships among the observed variables, yet I will still fall into error if I make inferences based on faithfulness. For example, LG and D are independent given LO; hence, I will suppose that they are not causally connected, when in fact they are.

Laplace: Absolutely right.

Meno: Well, then, here is what I do not understand. If deterministic relations, even between observed and unobserved variables, are incompatible with using faithfulness, and yet any indeterministic system may be viewed as a deterministic system with hidden variables, then how does it ever make sense to assume faithfulness? Because you seem comfortable with using faithfulness in some indeterministic contexts, doesn't your argument prove too much?

Laplace: An excellent observation. Is there no room left in this world for faithfulness? Here is the solution to your dichotomy. Suppose for a moment that we are omniscient demons, knowing the entire causal nexus.

Meno: "Laplacian" demons?

Laplace: If you insist. Given any set of observed variables, we will add variables to the set until, for any two variables in the set, if they have a common cause, then that variable is included in our set. Such a set of variables may be called *causally sufficient*. If there are no deterministic relationships among this larger set of (observed and unobserved) variables, then we may proceed to use faithfulness in our analysis of the original variables that we observed.

Meno: Of course, if we were the demon, we would not need to use faithfulness to infer the structure.

Laplace: Agreed. This is obviously a thought experiment. The point is that there is a well-defined set of variables among which we require there to be no deterministic relationships to safely base inferences on faithfulness.

Meno: I see. The scenario with the garage door opener obviously fails the test.

Laplace: Indeed. A simple way in which this condition can be satisfied is if each variable in the system is subject to at least one independent cause.

Meno: I see; so, deterministic relationships are not problematic in a system in which each variable has many causal parents.

Laplace: This is provided that we do not observe all of them, and that is usually the case in complex systems.

Meno: But this is highly problematic in deterministic systems in which variables have only a few parents.

Laplace: Whenever we make causal inferences, we are not considering all the possible variables, observed or hidden, that exist in the universe, but only a small subset of those variables.

Narrator: [*See also Glymour, chapter 18, this volume*.]\edq9\ \edq9\

Laplace: Metaphysically, we may have a hard time imagining genuinely indeterministic causal relations. But, even if we are metaphysical determinists, in complex settings we often simply ignore the unobserved variables we think are responsible for indeterministic appearances, especially in complex cases—we brush them off as "noise" that is irrelevant for causal inference. From a formal perspective, this epistemological brush-off has just the same consequences as believing in metaphysical indeterminism.

Meno: From what you say, simple deterministic systems are problematic for causal inference from conditional independence relations. Yet, many mechanical devices one can think of behave in exactly this way. After all, the blicket detector is just such a system and so are the other "machines," like the puppet or the gear-toy machine that developmentalists have used to test children's causal inferences. This brings us back to the original question of how children manage to learn such systems with so little data? If they do not use faithfulness to infer complex noisy causal systems, then how else could they manage to learn so much so quickly and accurately?

Laplace: Now, you ask me to enter the realm of conjecture. I can only guess, but I believe there are a number of factors that work to children's advantage.

Let us turn to faithfulness first. As we described, it serves to identify when variables are not causally related. This is important because, in the right context, it allows these algorithms to establish that a particular variable is unconfounded or exogenous—it is not itself affected by other variables in the system.

Meno: I see. Once we have established that variable X is exogenous, then we only require the causal Markov condition to conclude that anything dependent on X is caused by X. If X is exogenous, then we have ruled out the possibility of common causes and ruled out the possibility that anything else is causing X. But, if we cannot use faithfulness, then how else could we establish exogeneity?

Laplace: Children might establish exogeneity in other ways. For one thing, children, unlike data-mining programs, can actively intervene on the causal systems they are learning about. In fact, in their spontaneous play they perform such interventions all the time. It is what parents call "getting into everything." Children might have a background theory that allows them to attribute the property of exogeneity to actions they undertake. In particular, like adults, children might assume that their own intentional actions are the result of free will and so are intrinsically exogenous.

Meno: But in the blicket detector experiments, children do not get to intervene on the system; they just watch other people's interventions.

Laplace: This raises an interesting point. If children also assume that the actions of others are analogous to their own actions—particularly that they are also the result of free will and so are exogenous—then they could make similar inferences by just watching other people manipulate objects.

Meno: I see. In fact, interestingly, several experiments have shown that children will distinguish between actions performed by agents assumed to be like the children themselves and those performed by nonagents.

Narrator: [*See Meltzoff, Somerville, this volume* \edq*10*\. *Also see Schulz, Sommerville, et al., 2005, Experiment 4.*] edq10\

Meno: Children do not simply observe patterns of association but see goal-directed agents around them performing actions. Thus, a child might be more like a first-year graduate student in a lab or a historian of science, who may be fairly sure that if these otherwise well-adjusted adults spend a lot of time manipulating something, then it is probably causally efficacious in some way.

In fact, some developmentalists, as well as grown-up psychologists, have already argued for the significance of interventions in human causal inference.

Narrator: [*See Schulz et al., chapter* X, *this volume*; *Lagnado et al., chapter* X, *this volume*; *Hagemeyer et al., chapter* X, *this volume.*]\edq11\ \edq11\

Meno: In this respect, human inference is different from the perspective of a data-mining program, which cannot exclude the possibility that the variables are completely unrelated to one another (or completely confounded by unobserved variables). Other experiments have shown that children can use interventions on deterministic systems to make complicated inferences about the causal structure of those systems—distinguishing common causes, common effects, and causal chains.

Narrator: [*See Schulz et al. chapter* X, *this volume*\edq12\; *Schulz et al., in press.*]\edq13\ \edq12\ \edq13\

Meno: But, what you say suggests that interventions will be especially important, in fact, indispensable, if we want to understand deterministic systems.

By the way, earlier you mentioned crossover or within-individual experiments as playing a role. Can you expand on that point?

Laplace: As mentioned in our discussion, the "fundamental problem of causal inference" is that we typically do not get to view the outcomes for the same subject under two different treatments. Randomization serves to construct groups of subjects whose distributions of outcomes under the same treatment may be considered to be similar.

However, a child typically does not do experiments on a large group of blicket detectors. The child typically only has one detector, but it is often reasonable to assume that different interventions leave the device unchanged.

Meno: Yes—although it is possible to imagine that Block A is somehow "imprinted" on the blicket detector, like tweed trousers on a newly hatched Lorenzian duckling, causing it to squawk when (and only when) its first love is again placed on it—this is certainly not the first hypothesis that springs to mind. Indeed, the word *detector* seems to rule this out.

Laplace: Precisely. It would be an unusual (although perhaps not irrational) child who would say, "Block A is the blicket—it was blicketized by being the first object placed on the detector!"

Some interventions have permanent reversible effects on objects, such as dropping the glass on the tile floor or pouring ink on the Persian rug, but the fact of irreversibility is usually plain to see. Interventions that lead to undetectable, but permanent, irreversible effects are less common. Hence, children live in a world amenable to within-subject crossover designs.

Meno: Indeed; you could think of children's repetitive spontaneous play with objects as just such an experimental strategy. Grown-up psychologists often treat children's perseveration as a sign of stupidity or at least lack of executive control. But, it also might be an excellent way to get within-subject information, in particular to check that there have been no irreversible changes, so that the same intervention continues to produce the same effect.

In this way, the fundamental problem is avoided, and individual causal effects can be inferred.

Laplace: Yes, in fact, when combined with the assumption of determinism, it also makes feasible inferences about the existence of unobserved hidden causes of a single variable and the causal effect of such hidden causes (i.e., whether they are inhibitory or generative).

Narrator: [*See Schulz, Sommerville, et al., in press, Experiments 2 and 3.*]

Laplace: The Markov and faithfulness conditions sometimes make it possible to infer the existence of an unobserved common cause of two variables in indeterministic systems, and there is some evidence that adults and children make such inferences. But, inferences about the existence of single unobserved causes cannot be made purely on the basis of conditional independence and dependence. (Clustering of imputed "disturbance" terms would be one way to proceed.) Yet, children also seem to make such inferences.

Narrator: [*See Gopnik et al., 2004; Kushnir, Gopnik, Schulz, & Danks. 2003; Schulz et al., this volume.*] \edq14\ \edq14\

Meno: This may also solve another problem: One concern that I have had with the standard approach to causal inference based on directed acyclic graphs (also called Bayesian networks) is that all causal relationships are asymmetric: If X is a cause of Y, then Y is not a cause of X. In particular, under the standard account of interventions, if we were to intervene on Y, then we would produce no change in X—cyclic systems are explicitly ruled out. Yet, there are simple systems in which causal relationships appear to be reversible. For instance, I can pull the engine of a toy train, and the tender will be pulled along, but if I choose to push the tender forward, then the engine will also be moved. And, in some experiments children seem to infer such cyclic relationships. In the gear-toy experiments, for example, children hypothesized that Gear A might sometimes move Gear B while at the same time Gear B might sometimes move Gear A.

Narrator: [*See Schulz et al., chapter X, this volume;* \edq15\ *Schulz et al., in press.*]\edq16\ \edq15\ \edq16\

Laplace: Reversibility of the type you describe is simple to include in an account of intervention in which two variables are related deterministically and the relationship is one to one, so that each value of X corresponds to a unique value of Y and vice versa. This is a model of intervention corresponding to reversing edges rather than breaking edges. For example, if prior to intervention we have A → B → C and we then intervene on C, then this will lead to C → B → A.

Narrator: This intervention model may be generalized to having *p* input variables and *p* output variables, provided that each possible vector of values for the outputs corresponds to a unique vector of values for the inputs.

Meno: Like any working hypothesis, assuming determinism or near determinism (i.e., a few unobserved variables) will work well if true but may be highly misleading when false.

Laplace: But, again, we do not learn causal relationships purely out of intellectual curiosity. Considerations of utility also play a role. Deterministic causal systems are, by definition, more reliable, and thus more useful, once we have learned them. If our goal is to manipulate the world around us, then learning the subtleties of an unreliable system may not be worth the effort.

If a system is complex and indeterministic, then we have no hope of learning how to manipulate it, absent large amounts of data; hence, unless we really can gather a lot of data, from a pragmatic point of view we are losing little by ruling out such systems at the beginning.

Remaining Problems

Meno: You've convinced me that a near-deterministic experimental causal logic may serve children as well as the full apparatus of probabilistic Bayes net causal learning algorithms. But, do you see no role for indeterminism in children's learning?

Laplace: That may be going too far. I almost always would qualify anything I say. Empirically, children do seem to use observed frequency as a way of estimating causal strength, much as adults do. For instance, it has been shown that children think a block that sets off the detector 2/3 times has "more special stuff inside" than one that only sets it off 1 of 3 times. Of course, these judgments don't involve causal structure—the sort of judgments captured by causal graphs, but only the parameterization of those graphs.

Narrator: [*See Kushnir and Gopnik, 2005.*]

Meno: Don't these experiments necessitate the use of indeterministic models as cognitive constructs?

Laplace: An indeterministic model provides one explanation, but observe that it is also possible to see these experiments concerning the amount of special stuff as revealing that children are capable of using different levels of description of frequency rather than using indeterministic models per se.

Meno: How so?

Laplace: If we view the three responses resulting from placing the block on the detector three times in succession as a single response that takes four values 0, 1, 2, 3 (rings of 3), then we can build a deterministic model for the system. Certain blocks lead to a response of 1 of 3; others lead to a response of 2 of 3. For example, it might be the case that every time we place a given block on the detector a constant (deterministic) amount of special stuff is transferred to the detector. The detector accumulates special stuff until a threshold is reached, at which point the detector makes a noise, and its special stuff reservoir is depleted by some (fixed) amount. Although a given block always transfers the same amount, different blocks transfer different amounts.

Meno: I see. If we may set aside your metaphysical theory of special stuff for a moment, there appears to be a more general point here. Your reasoning seems to suggest that another route to incorporating seemingly indeterministic data into a deterministic world view is simply to provide a level of description for our variables that avoids recording the outcome in any specific case but rather just describes ensembles of outcomes. Thus, *Y* is a deterministic function of *X*, but *Y* takes values such as never, rarely, often, always, which refer to collections of individual observations.

Narrator: Note that in a deterministic system a given set of inputs either always or never produces a certain output.

Laplace: Indeed. This is an instance of the following idea, which is familiar from regression: Knowing someone's height does not allow us to predict their weight, but the average weight in a given subpopulation of people who are all of exactly the same height may be a simple deterministic function of that height. If the variable *Y* takes on values such as never or rarely, then it is basically recording the average (rate) of occurrence of an event under some condition. Thus, we may describe deterministically the way in which X (special stuff) influences the average response (frequency of the detector ringing).

Meno: This also raises a question regarding what it means to "use" an indeterministic model. Regression can be thought of as a statistical procedure derived from a probability model, or it can simply be thought of as line fitting. I could use the regression line for making predictions without explicitly assuming a probability model. In this case, am I or am I not using an indeterministic model?

Laplace: I agree that this is not so clear.

Meno: To my mind, psychological causation seems in some sense far more indeterministic than physical causation, and yet we know that children infer the structure of psychological systems as quickly and easily as they make physical inferences.

Narrator: [*See Schulz and Gopnik, 2004.*]

Laplace: Yes. Other agents are often quite unpredictable in the way in which they respond to us. Our daily interactions certainly provide plenty of time to gather data about those who are closest to us. On the other hand, such indeterministic systems may not have a fixed causal mechanism. Agents around us are changing even as we are learning about them: One of the ways in which they change is that while we learn about them they also are learning about us. This makes the learning task a bit more complicated because data are not generated by a fixed underlying distribution. You may be smiling at me because you like me, because you think that I like you, or more deviously, because you think you have made me think that you like me and so on.

Meno: Virologists and pathologists sometimes have to study systems that are constantly evolving and that change the way they function in response to interventions.

Laplace: Indeed, but viruses that evolve quickly are much harder to combat than those that do not.

Another difference is that in such circumstances the simple fact of gathering data—observing your expressions—is itself an intervention in the system. As every parent of a toddler soon finds out, often the best way to ensure that a tantrum continues is to try to find out what is wrong. So, it is puzzling that children make these inferences as easily as they do. On the other hand, the fact that children often manipulate their parents and vice versa suggests that perhaps humans are less hard to predict than we might like to believe.

Meno: Making the analogy between the way in which scientists and statisticians analyze their data and the way in which children learn from observations around them seems to me to leave two important parts of the process unexplained: hypothesis generation and concept formation. Do you agree?

Laplace: Absolutely. Statistical analysis of causation often gives no account regarding how particular variables are chosen as candidate causes or effects. Heuristics based on observing other agents may be of assistance to children in this regard. For example, Mommy seems to spend a lot of time fiddling with that little black box, so let me investigate it; someone or something turns the TV off, so let me see if I can find out what it is.

Machine learning algorithms often have a well-defined hypothesis space through which they perform some sort of search. However, children face a much less well-defined, hence larger, search space and arguably do not carry a giant list of all possible causal hypotheses. (This also causes problems for Bayesian accounts.) Choosing good candidate hypotheses in such circumstances seems like a hard problem, but one that they do well. Experiments such as those in which hidden causes (the flashlight) were hypothesized give a tantalizing glimpse of this process in action.

Meno: Finally, Laplace, as I say my association with Socrates has taught me the importance of empirical developmental findings. How could we test your ideas about determinism empirically?

Laplace: I regard the experiments relating to the key chain flashlight described as empirical evidence that children are willing to postulate the existence of hidden variables merely from observations that appear to be indeterministic in a manner not compatible with a conditional independence-based approach because such approaches only postulate common causes. Naturally, this does not rule out the use of probabilistic models in other settings.

More generally, I speculate that if we give children the same problem in a deterministic or near-deterministic way and in a way that genuinely requires them to compute conditional probabilities, then I predict that they will solve it in the deterministic case and not the probabilistic one. All the published studies have been deterministic or nearly so. I have a feeling that some unsuccessful probabilistic experiments might be lurking in wastebaskets and desk drawers.

Meno: Maybe so, but the experimental problem is harder than you might think. As you said, it may be that we use indeterministic information just when we assume that there are many uncontrolled variables, lots of noise in the system. But, a developmental psychologist's first task is to make sure that the problem is clearly posed, and there are no extra factors that might be distracting the child. You may be able to persuade undergraduates that they should only pay attention to the information about probabilities on the sheet directly in front of them. But, it will be much harder to persuade young children to do so (and even with undergraduates, the individual differences among participants suggest that they also may be considering other factors).

Narrator: [*See Lagnado et al., chapter X, this volume; Hagmeyer et al., chapter X, this volume.*]\edq17\

\edq17\

Meno: How can we be sure that children only pay attention to the variables we control while at the same time leaving them the impression that there are many other uncontrolled variables lurking in the background, and therefore that indeterminism might be appropriate?

Laplace : Perhaps we might exploit the indeterminism of psychological relations.

Meno: I see; suppose we show the child that Bunny the fussy eater will eat plain peanuts one of three times you offer them but will eat them three of four times when you add salt, although he never eats salt alone. The salt influences the probability distribution of Bunny's preferences. Will children infer that the salt has a causal effect on Bunny's actions in this indeterministic case?

Laplace: I think it would be interesting to see how children would respond. However, I believe that, as with the special stuff experiments, it would be possible for someone to describe the result of the experiment deterministically, without reference to probabilities, by saying that, "Bunny frequently eats peanuts with salt, but rarely eats them without."

As with our discussion concerning the pros and cons of a Bayesian explanation of human reasoning, I believe that although many observations may be compatible with a child entertaining an indeterministic model, I think it is unlikely to be necessary. Probability is a relatively recent addition to the set of descriptive methods used by scientists. It was also one that was fiercely resisted at first. Probability may seem to be an integral part of the metaphysical landscape in the 21st century, but it certainly was not always thus.

Meno: Oh, dear. We appear to have raised as many interesting issues as we have resolved. At least we have established the importance and primacy of experimental evidence in informing our theorizing.

As my dear friend Lavoisier\edq18\ says:

\edq18\

> In the practice of the sciences imagination, which is ever wandering beyond the bounds of truth, joined to self-love and that self-confidence we are so apt to indulge, prompt us to draw conclusions which are not immediately derived from facts; so that we become in some measure interested in deceiving ourselves.
>
> [In contrast] . . . when we begin the study of any science, we are in a situation, respecting that science, similar to that of children; and the course by which we have to advance is precisely the same which Nature follows in the formation of their ideas. . . . We ought to form no idea but what is a necessary consequence, and immediate effect, of an experiment or observation. (p. 4)

References

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., Sobel, D., Schulz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: 2-, 3-, and 4-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37, 620–629.

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

Kushnir, T., Gopnik, A., Schulz, L., & Danks, D. (2003). Inferring hidden causes. In R. Alterman & D. Kirsch (Eds.). *Proceedings of the 24th Annual Meeting of the Cognitive Science Society.* Boston:
\edq19\ Cognitive Science Society.\edq19\

Lavoisier, A. L. (1994). *Elements of Chemistry* (R. Kerr, Trans.). Chicago: Encyclopedia Britannica. (Original work published 1789)

Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly*, *2/4*, 261–264.

Dabrowska, D., & Speed, T. (Trans.). *Statistical Science*,
\edq20\ 5, 463–472.\edq20\

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*.
\edq21\ San Mateo, CA: Morgan Kaufmann.\edq21\

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

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, *66*, 688–701.

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

Schulz, L. E., Gopnik, A., & Glymour, C. (in press). Preschool children learn about causal structure from conditional interventions. *Developmental*
\edq22\ *Science*.\edq22\

Schulz, L. E., Sommerville, J., & Gopnik, A. (in press). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child*
\edq23\ *Development*.\edq23\

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction and search* (Springer Lecture
\edq24\ Notes in Statistics No. 81).\edq24\

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

Appendix

This appendix outlines simple methods for learning cause-effect relationships from small numbers of interventions in a causal system in which all nodes are binary, and all endogenous variables (i.e., with parents) are a (deterministic) disjunction of conjunctions of their parents. We present an algorithm that lays out a simple experimental procedure that enables the learner to learn first indirect and then direct causes. In contrast to other machine learning algorithms, this procedure does not say anything about what should be inferred from passive observations. This is in keeping with the contention (of Laplace) that interventions allow learners to avoid the complexities of probabilistic inference. Further, the algorithm focuses on what causes a specific outcome variable. Again, this reflects the view (again stated by Laplace) that learners are often attempting to (re)produce a particular outcome (e.g., make it go, make Mommy smile, etc.).

The method outlined here is relevant to the blicket experiments (and others like them) only insofar as a participant might view the actions of the person putting the blocks on the detector as (partially) carrying out the sequence of interventions sketched in this method.

Basic Notions and Definitions

Consider a deterministic causal model in which all variables are boolean, taking values true or false. We also refer to these states as on and off, respectively. We suppose an underlying directed causal graph in which every vertex with parents is a logical function of its parents taking the specific form of a disjunction of conjuncts:

$$x{=}(p_{11} \wedge p_{12} \cdots \wedge p_{1k1}) \vee (p_{21} \wedge \cdots \wedge p_{2k2}) \vee \cdots \vee (p_{t1} \wedge \cdots \wedge p_{tkt}). \qquad (*)$$

Here, the $\{p_{ij}\}$ are all in the set pa(v) of parents of v in the causal graph, which we will require to be acyclic (i.e., containing no directed cycles).

Intuitively, v is true if all of the p_{ij}'s inside at least one of the parentheses are true. This also includes as special cases a network in which each vertex with parents is either a conjunct or a disjunct of its parents.

This is a strong restriction that rules out many possible relationships between causes and effects (see discussion here). However, it is general enough to cover most (all?) experiments considered in the developmental literature while still being simple enough to allow a relatively direct inferential method. At a crude level, it captures the idea present in many experiments that "something" (a given conjunction) makes "it" (v) "go" (be true). There is also a close connection to Mackie's (1965) INUS\edq25\ model. \edq25\

Let **E** be the set of exogenous variables and **V** be the set of endogenous variables. We will define an *instance* of the system to consist of an assignment of truth values to all of the exogenous variables (i.e., those without parents), which we will denote by $\mathbf{E} = \mathbf{e}$. Because there is an equation of the form (*) for each of the endogenous variables, an assignment of values to the exogenous variables automatically assigns truth values to all of the endogenous variables.

Let $\Phi_X(\mathbf{e})$ be the value assigned to the endogenous variable X when the exogenous variables are assigned $\mathbf{e}$. Similarly, we define the set of true or on variables associated with an assignment as follows:

$$\Phi^T(\mathbf{e}) = \{v \mid v \in \mathbf{V}, \boldsymbol{\Phi}_{\boldsymbol{X}}(\mathbf{e}) = T\}$$

and likewise

$$\Phi^F(\mathbf{e}) = \{v \mid v \in \mathbf{V}, \boldsymbol{\Phi}_{\boldsymbol{X}}(\mathbf{e}) = F\}.$$

Note that we have not (and will not) put any distribution over the exogenous variables.

The following are some examples described in this format.

Example 1: One-cause blicket detector

Exogenous variables: Block 1 present? (B1); Block 2 present? (B2).

Endogenous variable: *Detector making a noise? (D).*

Graph: B1 → D B2

Functional relationship: D = B1

This is the trivial case of (*) where $t = 1$, and $k_1 = 1$. We have, for instance, $\Phi^T(\text{B1} = T, \text{B2} = F) = \{\text{D}\}$; $\Phi^T(\text{B1} = F, \text{B2} = T) = \{\ \}$. (Here { } indicates the empty set.)

Notational convention To simplify notation, we often simply describe an assignment via the subset of exogenous variables taking on value T, it being implicit that the remaining variables take the value F. Thus, for example, we may reexpress the statements as follows:

$$\Phi^T(\{\text{B1}\}) = \{\text{D}\}; \Phi^T(\{\text{B2}\}) = \{\ \}$$

This convention simplifies expressions, but it is also based on the intuition that the default state for exogenous variables is false or off. Thus, if we were to physically implement a particular assignment, we would only need to pay attention to those exogenous variables assigned the value true as the remaining exogenous variables would already be in the false state.

Example 2: Two-cause blicket detector

Endogenous and exogenous variables are the same as in Example 1.

Graph: B1 → D ← B2

Functional relationship: D = B1 ∨ B2

Here, $t = 2$, $k_1 = k_2 = 1$.

Example 3: Twin piston engine

See Glymour, chapter XX, this volume.\edq26\ \edq26\

Exogenous: Key present? (K)

Endogenous: Fuel Intake 1 open? (F1); Spark? (S); Fuel Intake 2 (F2)? Piston 1 moves? (P1); Piston 2 moves? (P2); Drive Shaft moves? (D) (see Figure 13-A1).

Functional relations:

$$\text{F1} = \text{C; S} = \text{C; F2} = \text{C; P1} = \text{F1} \wedge \text{C;}$$
$$\text{P2} = \text{F2} \wedge \text{C; D} = \text{P1} \wedge \text{P2.}$$

The following is an important consequence of our restriction on the functional forms of the parent-child relationships:

Lemma 1: If **e1** and **e2** are two assignments to **E** such that

$\{X \mid X \in \mathbf{E}$; X assigned T by $\mathbf{e1}\} \subseteq \{X \mid X \in \mathbf{E}$; X assigned T by $\mathbf{e2}\}$

then $\Phi^T(\mathbf{e1}) \subseteq \Phi{:}^T(\mathbf{e2})$.

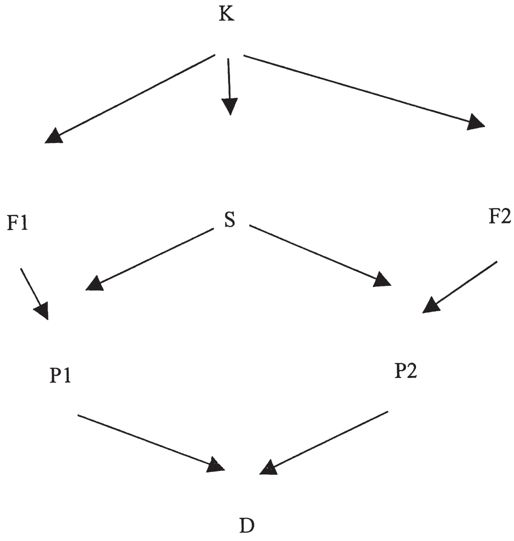

FIGURE 13-A1.

In words, if Assignment $\mathbf{e2}$ turns on every exogenous variable turned on by $\mathbf{e1}$, then at least as many endogenous variables are turned on by $\mathbf{e2}$ as by $\mathbf{e1}$.

Interventions

So far, we have not described operations for intervening in the system. An intervention turns an endogenous variable into an exogenous variable, forcing it to take a given value, and striking out the equation previously governing it. All other equations remain in place. We will simply denote this via expanding our assignment to include the intervened variables Z. By a natural extension of the previous notation, we will let $\Phi^T(\mathbf{E} = \mathbf{e}, \mathrm{Z} = \mathbf{z})$ be the value assigned to the endogenous variable X under this assignment and intervention. Likewise, the set of (remaining) endogenous variables taking the value T under this intervention is then represented via $\Phi^T(\mathbf{E} = \mathbf{e}, \mathrm{Z} = \mathbf{z})$.
For example, in the piston engine example, we have

$$\Phi^T(K = F) = \{\ \}, \text{ but } \Phi^T(K = F, \mathrm{P1} = T, \mathrm{P2} = T) = \{\mathrm{D}\}$$

expressing the fact that if the key is absent, then nothing happens, but if the key is absent and we force both pistons to move, then the drive shaft turns.

In the schemes described next, we will only ever consider interventions that force variables to take the true state. Thus, as before we can simplify notation by recording only the set of exogenous variables taking the value T and the set of endogenous variables forced to take the value T. For example, the above statements can be expressed as $\Phi^T(\{\ \}) = \{\ \}$, but $\Phi^T(\{\mathrm{P1}, \mathrm{P2}\}) = \{\mathrm{D}\}$.

More generally, under an assignment and intervention represented by the set $\mathbf{W} = \mathbf{E}^* \cup \mathbf{A}$, where $\mathbf{E}^*$ is a subset of the exogenous variables $\mathbf{E}$, and $\mathbf{A}$ is a subset of the endogenous variables $\mathbf{V}$, we mean the following:

(i) Assignment of true to the variables in $\mathbf{E}^*$,

(ii) Assignment of false to the variables in $\mathbf{E}\backslash\mathbf{E}^*$ (i.e., those not in $\mathbf{E}$), and

(iii) An intervention forcing the variables in $\mathbf{A}$ to take the value true.

This notation is not fully general in the sense that we cannot express interventions forcing endogenous variables to take the value false or off. However, for our purposes this is not a problem: As stated, the learning methods we describe next only ever require us to perform interventions forcing endogenous variables to take the value true.

We will sometimes refer to such an assignment and intervention as an intervention on $\mathbf{W}$ (= $\mathbf{E}^* \cup \mathbf{A}$). This is a slight abuse of terminology because in fact we are assigning to $\mathbf{E}$ (= $\mathbf{E}^* \cup (\mathbf{E}\backslash\mathbf{E}^*)$) and intervening on $\mathbf{A}$. However, because interventions simply make endogenous variables exogenous, assignments to exogenous variables may be viewed as trivial interventions.

Similarly, we will refer to the state that a given variable X (endogenous or exogenous) takes under $\mathbf{E}^* \cup \mathbf{A}$: If X is in $\mathbf{E}$ or in $\mathbf{A}$, then this is specified directly by the intervention; if not, then X's value is given by $\Phi_X(\mathbf{W})$. If X takes the value true under $\mathbf{W}$, then we will say that X is turned on by $\boldsymbol{W}$. The set of variables turned on under $\mathbf{W}$ consists of $\mathbf{W} \cup \Phi^T(\mathbf{W})$.
Finally, we note the following properties

> Lemma 2: $\Phi^T(\mathbf{W}) \subseteq \mathrm{de}(\mathbf{W})$, where $\mathrm{de}(\mathbf{W})$ is the set of descendants of $\mathbf{W}$.

In words, the set of endogenous variables taking the value true under $\mathbf{W}$ is a subset of the descendants of $\mathbf{W}$.

> Lemma 3: For any set $\mathbf{A} \subseteq \mathbf{V} \cup \mathbf{E}$ and any variable $X \notin \mathbf{A}$,
>
> $\Phi_X(\mathbf{A}) = \Phi_X(\mathbf{A} \cap \mathrm{an(X)}) = \Phi_X((\mathbf{A} \cap \mathrm{an(X)}) \cup \mathbf{W})$
>
> where $\mathbf{W}$ is an arbitrary subset of $(\mathbf{V} \cup \mathbf{E})\backslash(\mathrm{an(X)} \cup \{\mathrm{X}\})$.

In words, the truth value taken by an (endogenous) variable X only depends on the values assigned to variables (either exogenous or intervened on) that are ancestors of X.

Learning Indirect Causes From Interventions: How Can I Make It Go?

We can now describe the a simple method for answering the following question: For a specific variable X in a causal model, how do I get it to "go" with the least effort?

We are not necessarily trying to find the direct causes of X, we merely require a nonredundant set of minimal sufficient causes. We formalize this question as follows: For a given variable X, find a set $\mathbf{A}$ such that $\mathbf{A}$ does not contain X, turning on all the variables in $\mathbf{A}$ makes X take the value T, and no subset of $\mathbf{A}$ makes X take the value T.

Such a set $\mathbf{A}$ may be found in a simple manner: First, try turning on each variable in turn (if necessary by intervention) other than X itself. If successful, stop; otherwise, try sets of size two, and so on.

More formally, we have the following algorithm.

Input: A target variable X

Output: A set $\mathbf{A}$ such that $X \in \Phi^T(\mathbf{A})$, but for any subset $\mathbf{A}^* \subset \mathbf{A}$, $X \notin \Phi^T(\mathbf{A})$,

or failure if no such set exists.

How algorithm

```
For $k = 1$ to $| (\mathbf{V} \cup \mathbf{E}) \setminus \{X\} |$
    For each subset $\mathbf{A} \subseteq (\mathbf{V} \cup \mathbf{E}) \setminus \{X\}$, such that $|\mathbf{A}| = \mathrm{k}$
        If $X \in \Phi^T(\mathbf{A})$, return $\mathbf{A}$
    Until all subsets of size $k$ from $(\mathbf{V} \cup \mathbf{E}) \setminus \{X\}$ have been tried.
    $k = k + 1$
If $k > |(\mathbf{V} \cup \mathbf{E}) \setminus \{X\}|$, then report failure and return.
```

Failure will only occur if the target variable is in fact exogenous or if (contrary to the assumptions of the algorithm) we are not able to intervene on all variables in the system.

Example piston engine If we attempt to get the drive shaft to turn ($D = T$), then the algorithm will terminate with $k = 1$, with the set $\mathbf{A} = \{K\}$ because this is the only set of size 1 making the engine turn over. Note that K is an ancestor but not a parent of D. This will be true in general:

Lemma 4: The set $\mathbf{A}$ resulting from the how algorithm consists of ancestors of the target variable X.

Sketch of proof Suppose for a contradiction that $\mathbf{A}$ contained a variable that was not an ancestor of X. Consider the set $\mathbf{A}^* = \mathbf{A} \cap \mathrm{an}(X)$. By Lemma 3, $\Phi_X(\mathbf{A}) = \Phi_X(\mathbf{A} \cap \mathrm{an}(X)) = \Phi_X(\mathbf{A}^*)$. So, in particular, if $X \in \Phi^T(\mathbf{A})$, then $X \in \Phi^T(\mathbf{A}^*)$. But, by hypothesis because $\mathbf{A}$ contains a vertex that is not an ancestor of $\mathbf{A}$, $| \mathbf{A}^* | < | \mathbf{A} |$, so the set $\mathbf{A}^*$ would have been considered first by the algorithm, which is a contradiction.

The number of interventions required to find the set $\mathbf{A} = \{K\}$ in the piston example depends on the ordering of the variables. Under the worst ordering, we would need to perform six sets of interventions, each forcing a single variable to take the value true. In the best case, only one intervention is required.

Note that, in a system containing no conjunctions, it will only be necessary to consider sets of size 1 in the how algorithm; hence, the outer loop is unnecessary. This corresponds to the simple scheme of getting into everything by which a child simply pushes each button in turn (literally or figuratively) until the desired effect is obtained.

Because various child-proofing schemes involve conjunctions, we conjecture that such systems may be harder to learn. For example, on some dishwashers, when the child lock is activated, pressing any button causes two buttons to flash, which must then be pressed simultaneously to proceed. Similarly, some stair gates require a button to be pushed and a pedal to be pressed simultaneously.

Learning Direct Causes From Interventions: Why Does That Make It Go?

The how algorithm succeeds in finding an intervention that makes a given variable X go, that is, take the value true, but as we saw in the piston example, it does not necessarily identify the direct causes or, equivalently, the parents of X in the graph. Thus, a causal learner might ask this as a follow-up: Given that $\mathbf{A}$ makes X go, why does $\mathbf{A}$ make X go?

We reformulate this question as follows: Can we identify parents of X that are turned on by $\mathbf{A}$ and consequently turn on X? We emphasize that this is clearly a limited answer to the question, Why does $\mathbf{A}$ make X go? In particular, if $\mathbf{A}$ is a set of parents of X, then we will simply return the answer that, $\mathbf{A}$ makes X go because $\mathbf{A}$ makes X go, which, though true, is not very illuminating.

The idea behind the algorithm is that if a set $\mathbf{A}$ turns on X but does not consist solely of parents of X, then if instead we were to turn on only the parents of X that are turned on by $\mathbf{A}$, it will lead to a reduction in the number of variables turned on overall. Put more formally: For a given variable X and set $\mathbf{A}$ that turns on X, can we find a set $\mathbf{P}$ such that

(a) $X \notin \mathbf{P}$, but $X \in \Phi^T(\mathbf{P})$, that is, $\mathbf{P}$ turns on X;

(b) $\mathbf{P} \subseteq \mathbf{A} \cup \Phi^T(\mathbf{A})$, that is, every variable in $\mathbf{P}$ is turned on by $\mathbf{A}$;

(c) There is no subset $\mathbf{P}^*$ of the variables turned on by $\mathbf{P}$, that is, $\mathbf{P}^* \subset \mathbf{P} \cup \Phi^T(\mathbf{P})$, such that $X \notin \mathbf{P}^*$, but $X \in \Phi^T(\mathbf{P}^*)$.

Condition (c) states that there is no subset of the variables that take the value true under **P**, which does not contain X, and which will make X take the value true. Note that it also follows from this that no subset of **P** will make X take the value true.

> Lemma 5: A set **P** satisfying conditions (a), (b), and (c) will consist of parents of X that are either descendants of **A** or are themselves in **A**.

Proof: First suppose that **P** is not a subset of pa(X). Consider the set $\mathbf{P}^* = (\mathbf{P} \cup \Phi^T(\mathbf{P})) \cap \mathrm{pa}(X)$. Because $X \in \Phi^T(\mathbf{P})$, and by construction, the variables in pa(X) are assigned the same values under $\mathbf{P}^*$ as they take under **P**, it follows that $X \in \Phi^T(\mathbf{P}^*)$. However, $X \notin \mathbf{P}^*$. Now, $\mathbf{P}^* \subseteq \mathrm{pa}(X)$, but by hypothesis **P** is not a subset of pa(X). Thus, $\mathbf{P}^*$ is a strict subset of $\mathbf{P} \cup \Phi^T(\mathbf{P})$; hence, **P** does not satisfy Condition (c), which is a contradiction.

That the variables in **P** are descendants of **A** follows from $\mathbf{P} \subseteq \mathbf{A} \cup \Phi^T(\mathbf{A})$ and Lemma 2.

We now outline the algorithm for finding the set **P**:

Why algorithm

```
Input: A set $\mathbf{A}$ and vertex X such that $X \in \Phi^T(\mathbf{A})$;
Output: A set $\mathbf{P}$ satisfying Conditions (i), (ii), and (iii);
0. Let $\mathbf{P} = \mathbf{A}$;
1. For each vertex P ( $\mathbf{P}$
      For $k = 1$ to $|(\mathbf{P} \cup \Phi^T(\mathbf{P}))\backslash\{P,X\}|$
            For each subset $\mathbf{P}^* \subseteq (\mathbf{P} \cup \Phi^T(\mathbf{P}))\backslash\{P,X\}$ such that $|\mathbf{P}^*| = k$
                  If $X \in \Phi^T(\mathbf{P}^*)$, then let $\mathbf{P} = \mathbf{P}^*$ and return to Step 1.
            Until all subsets of size $k$ from $(\mathbf{P} \cup \Phi^T(\mathbf{P}))\backslash\{P,X\}$ have been tried.
      $k = k + 1$
   If $k > |(\mathbf{P} \cup \Phi^T(\mathbf{P}))\backslash\{P,X\}|$, output $\mathbf{P}$.
```

Step 1 attempts to remove each vertex in turn from the set **P** but at the same time intervene on additional variables that were turned on by **P**. If we are successful in removing a given vertex from **P**, then we replace **P** with **P*** and start the search all over again.

We finish by illustrating the algorithm on the piston engine example.

> After running the how algorithm, we obtained the set $\mathbf{A} = \{K\}$, which made the target variable D take the value true.
>
> Initially, $\mathbf{P} = \{K\}$, and there is only vertex P to remove.
>
> The smallest subset of $\Phi^T(\{K\})\backslash\{D\} = \{F1,F2,S,P1,P2\}$, which turns on D, is
>
> $\mathbf{P}^* = \{P1,P2\}$; thus, we set $\mathbf{P} = \mathbf{P}^* = \{P1,P2\}$ and go back to Step 1.
>
> Because **P** is now the set of parents of D, we are unable to remove any vertices from the set, and the algorithm terminates.

Exactly how many interventions are required depends on the ordering of the variables. Under the worst ordering, we would have to perform five sets of interventions on sets of size 1 and then 10 interventions on sets of size 2 before we found {P1,P2}, giving 15 sets of interventions total. Under the best ordering, we would only need 6. Because there are no vertices in $\Phi^T(\{P1,P2\})\backslash\{D\}$, there are no new experiments required to confirm that this set satisfies Conditions (a), (b), and (c).

Note that we have only uncovered some of the causal structure. In this example, we found all of the parents of X. In general, we would only find a subset of the parents corresponding to one of the conjuncts in the equation (*).

To find the whole structure of the piston engine would require us to choose each endogenous variable as the target (X) and then to run the how and why algorithms in turn. Although perhaps laborious, it is worth noting that in this way the simple interventionist procedure would allow us to recover the whole structure. In contrast, a procedure based on passive observation leaves a large set of possible structures (see Glymour, chapter XX, this volume).\edq27\ \edq27\

Relaxing the Assumption on Functional Relationships

Two questions arise from this analysis. Could the algorithms be extended to cover the case in which the relations between the variables are not restricted to disjunctions of conjunctions, for example, in which negations of variables are permitted? Conversely, Are there many causal structures that we encounter in our daily existence in which the functional relationships are not of this form?

Consider a staircase with a light and a light switch at the top and bottom of the stairs. In the usual manner in which such switches are configured, flipping one switch while leaving the other unchanged always changes the state of the light (from on to off or vice versa). A little thought reveals that such a system implements an XOR gate, for example, things might be wired so that if both switches are up or both are down, then the light is off; otherwise, it is on. This is the simplest structure that cannot be handled in the framework considered. However, it is worth noticing that we are almost never in a position to operate both switches at once. As long as we only operate one switch and regard the other as fixed in its state, then the subsystem consisting of the single switch confronting us and a lightbulb falls within our framework.

As this example illustrates, an analysis of such structures is harder because there is less clear correspondence between interventions and outcomes.

FIGURE 13-1 (a) Treatment causes outcome; (b) outcome causes treatment; (c) treatment and outcome have a common cause.

\edq1\If the conversation is supposed to be between the authors, why use Laplace and Meno? Clarify here.
\edq2\Okay to place in-text mention of Figure 13-2 here? If not, provide.
\edq3\Italics addition okay for variables? If not, indicate which variables should be italic.
\edq4\Provide Tversky and Kahneman, 1982, reference and page number of quotation.
\edq5\Cross reference to this volume. Clarify and provide specific authors for either chapter 19 or 20.
\edq6\Provide first initial and date of communication.
\edq7\Provide authors for Schulz et al., in press. Cross reference to this volume; clarify which chapter and provide all authors.
\edq8\Cross reference to this volume. Provide correct chapter and authors.
\edq9\Cross reference to chapter 18. Correct chapter?
\edq10\Cross reference to this volume. Provide correct chapter(s).
\edq11\Provide correct chapter numbers and authors per APA style.
\edq12\Provide correct chapter numbers and authors per APA style.
\edq13\Indicate which Schulz et al., in press reference.
\edq14\Cross reference to this volume. Indicate correct chapter.
\edq15\Cross reference to this volume. Indicate correct chapter.
\edq16\Indicate which Schulz et al., in press.
\edq17\Cross references to this volume. Indicate correct chapters.
\edq18\Should this be Lavoisier, 1789/1994?
\edq19\Provide inclusive page numbers for Kushnir et al., 2003.
\edq20\Provide article title.
\edq21\Please mention Pearl, 1988, in text.
\edq22\Update Schulz, Gopnik, & Glymour, in press.
\edq23\Update Schulz, Sommerville, & Gopnik, in press.
\edq24\Provide publisher's name, city, and state.
\edq25\Define INUS, if appropriate.
\edq26\Provide correct chapter number for cross reference to Glymour.
\edq27\Provide correct chapter number for cross reference to Glymour.