Causal Relevance

Glenn Shafer

Department of Accounting and Information Systems

Faculty of Management, Rutgers University

180 University Avenue, Newark, New Jersey 07102

gshafer@andromeda.rutgers.edu

Abstract

Concepts of causal relevance and irrelevance are readily formulated in the context of Bayes nets, but these formulations have significant shortcomings. Most importantly, they do not allow for the great variety that can be observed in the temporal configuration of causally related entities. For example, they deal awkwardly with progressive causation, where continued action of a cause continues to enhance an effect. This article discusses how such subtleties can be handled when we look beyond Bayes nets to a more fundamental structure: nature's probability tree.

Keywords

Bayes net, causality, probability tree, relevance, refinement, sign, simplification, tracking

1. Introduction

Relevance is relative. Any real experience is essentially infinite in detail, and almost every detail is relevant, in some way, to almost every narrative that can be drawn from the experience. When I tell you a story of a harrowing taxi ride and fail to mention the color of the hat worn by the driver, have I omitted something relevant? If I have told the story coherently, then I will contend that I did not need to mention the color—it is irrelevant at the level of detail at which I worked. Yet there may be some larger, more detailed story, perhaps one that traces the taxi driver's behavior to his mood as he left his home, in which the color of his hat is relevant. The color may be relevant at one level of detail but not at another.

This article explores relevance in the context of probabilistic representations—Bayes nets and probability trees—and it concludes that relevance is just as relative here as it is in narrative. A Bayes net or a probability tree can always be made more detailed, and variables irrelevant at one level of description may be relevant to a more refined description.

The Bayes nets and probability trees we will discuss are meant to be causal: they purport to describe events as they actually unfold in nature. Thus we will be exploring causal relevance. In the course of the discussion, we will also touch on other important issues in the probabilistic representation of causality.

In both statistics and expert systems, discussions of causal relevance usually concern variables, and it is often assumed that these variables are subsequent: they are ordered, and their values are determined in nature in this order. We begin by discussing causal sufficiency and causal relevance in this setting, and we show how Bayes nets can be interpreted in this setting. We conclude, however, that the Bayes nets used in practice often require more subtle interpretations. This leads us to probability trees, because a clear understanding of alternative causal interpretations of Bayes nets requires us to think of them as partial descriptions of more explicitly dynamic structures.

We discuss the relationship between Bayes nets and probability trees at some length. The view taken here is that probability trees provide a semantics for Bayes nets. Probability trees are more expressive in general, and Bayes nets can be interpreted in different ways in terms of probability trees. This is sometimes disputed by the most fervent proponents of Bayes nets, who are tempted to argue that Bayes nets are themselves fully as expressive as probability trees. So the issue is addressed with some care.

Since relevance is relative to degree of detail, we look at refinement carefully for both Bayes nets and probability trees.

Many of the ideas in this article are treated in more depth in a forthcoming book by the author, *The Art of Causal Conjecture* (1996). This book does not, however, deal with the concept of relevance.

2. Causal Sufficiency and Relevance for Subsequent Variables

Suppose X and Y are variables. We say that Y is *subsequent* to X, or simply that X and Y are *subsequent*, if X is always settled in nature before Y, no matter what course events take. (Let us interpret "before" loosely, so that it means "before or at the same time as.") If Y is subsequent to X, then we also say that X *precedes* Y.

This idea generalizes to sequences of variables: $X_1,X_2,...,X_n$ are *subsequent* if they are always settled in nature in the order in which they are numbered: if $i<j$, then X_i is settled before X_j.

In the context of subsequent variables, there is a very natural concept of causal sufficiency:

> **Definition 1a** A variable X is *causally sufficient* for a subsequent variable Y if the probabilities for Y at the time when X is settled depend only on the

value of *X*. (In other words, the probability distribution for *Y* when *X* is settled can be specified fully once the value of *X* is specified, without regard to how earlier events or variables may have been settled.)

Definition 1b More generally, a family of variables *A* is *causally sufficient* for a subsequent variable *Y* if the probabilities for *Y* at the time when *A* is settled depend only the configuration to which it is settled. (A family of variables is settled when all its variables are settled; if the variables are subsequent, this is when the last of its variables is settled. A *configuration* is a specification of values for all the variables in the family. The empty family has a single configuration. The empty family precedes every variable, because it is settled before time begins; it is causally sufficient for a variable only if that variable is a constant. See Section 3 of Appendix D of *The Art of Causal Conjecture*.)

Definition 1c Yet more generally, a family of variables *A* is *causally sufficient* for a subsequent family of variables *C* if the probabilities for *C* when *A* is settled depend only the configuration to which it is settled.

It should be kept in mind that causal sufficiency, as defined here, always assumes subsequence. (In *The Art of Causal Conjecture*, the concept is called "stochastic subsequence." See Section 4 of Chapter 9.)

We leave it to the reader to verify the following proposition:

Proposition 1 If *A* is causally sufficient for *B* and *AB* is causally sufficient for *C*, then *A* is causally sufficient for *BC*.

Using causal sufficiency, we can define one concept of causal relevance:

Definition 2a A variable *X* is *causally relevant* to a subsequent variable *Y* if there is a family of variables *A*, which also precede *Y*, such that *AX* is causally sufficient for *Y* but *A* alone is not. (We write *AX* for the family obtained by adding *X* to *A*.)

This concept, as it turns out, is rather weak. Practically any preceding variable is causally relevant in this sense.

The preceding definition of causal relevance can be compared with the idea of an inus condition, an idea developed abstractly by Mackie (1974) and adapted to probability by Marini and Singer (1988). An *inus condition* is an insufficient but necessary part of an unnecessary but sufficient condition. Mackie and Marini and Singer propose calling an inus condition a "cause."

It appears, from the literature on probabilistic causality (see Bollen 1989, Humphreys 1989), that many authors have a strong desire to call variables causes. This practice should, however, be avoided. Variables merely measure, usually in a relatively superficial and aggregate way, how things turn out. They may point to causes, but they are not causes themselves. Intuitively, X is causally relevant to Y if it can help point to some of Y's causes. Other variables may do the same job equally well or better.

The concept of causal relevance gains more content if we make explicit reference to the family that becomes causally sufficient when the variable is added:

> **Definition 2b** A variable X is *causally relevant* to a subsequent variable Y *relative* to a family A of variables preceding Y if AX is causally sufficient for Y but A alone is not.

A variable may, of course, be causally relevant relative to one family but not relative to another.

The concept of causal sufficiency is itself relative to the time at which variables are settled. We can make this explicit by generalizing Definition 1b in this way:

> **Definition 1d** Suppose A and C are families of variables preceding Y, and suppose $A \subseteq C$. Then A is *causally sufficient* for Y *relative to the resolution* of C if the probabilities for Y at the point where C is settled depend only the configuration of A.

When we say simply that *A* is causally sufficient for *Y*, we are saying that *A* is causally sufficient for *Y* relative to the resolution of *A*.

On the other side of the coin from this last concept of causal sufficiency is a concept of causal irrelevance:

> **Definition 2c** Suppose *B* and *C* are families of variables preceding *Y*, and suppose $B \subseteq C$. Then *B* is *causally irrelevant* for *Y relative to the resolution* of *C* if *C*-*B* is causally sufficient for *Y* relative to the resolution of *C*.

When using this concept, we must remember that the irrelevance of *B* is relative to *C* and hence, in a sense, to the other variables, *C*-*B*. It is quite possible for a family *C* to consist of two subfamilies, *A* and *B*, both of which are causally sufficient for *Y* relative to the resolution of *C*. In this case, both *A* and *B* can be called irrelevant. But this means only that one is superfluous if the other is used, not that they are together superfluous.

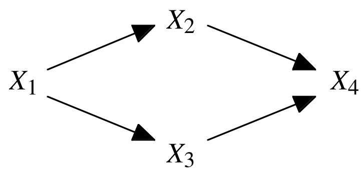
Another example of the relative nature of sufficiency and irrelevance is provided by the concept of a mediating variable. If *X*,*Y*,*Z* are subsequent variables, then we may well find that *X* is causally sufficient for *Z* relative to the resolution of *X* but causally irrelevant relative to the resolution of *XY*. In this case, we call *Y* a *mediating* variable; the causes marked by *Y* mediate the influence on *Z* of the causes marked by *X*.

3. Subsequently Causal Bayes Nets

A *Bayes net* is a directed acyclic graph whose nodes are variables, together with a representation, perhaps in tabular form, of conditional probabilities for each variable. An abstract example is given in Figure 1.

X_1 \ X_2	x_{21}	x_{22}
x_{11}	.3	.7
x_{12}	.4	.6

X_1: x_{11}	x_{12}
.5	.5



X_2,X_3 \ X_4	x_{41}	x_{42}	x_{43}
x_{21},x_{31}	.3	.3	.4
x_{21},x_{32}	.3	.5	.2
x_{22},x_{31}	.4	.1	.5
x_{22},x_{32}	.4	.3	.3

X_1 \ X_3	x_{31}	x_{32}
x_{11}	.5	.5
x_{12}	.1	.9

Figure 1 An abstract Bayes net. Each table gives conditional probabilities for the variable at the top of the table, given the values of the variables (if any) listed at the left of the table. Variables X_1, X_2, and X_3 each have two possible values, while X_4 has three.

When there is an arrow from a variable X to a variable Y in a Bayes net, we say that X is a *parent* of Y. We may assume that the variables are numbered as in Figure 1: the parents of a variable always have lower numbers. We write $X_1,X_2,...,X_n$ for the variables, and we write $\mathbf{par}_i$ for the parents of X_i. The table for X_i in the Bayes net gives conditional probabilities for X_i given $\mathbf{par}_i$. In Figure 1, $\mathbf{par}_1 = \varnothing$, $\mathbf{par}_2 = \mathbf{par}_3 = X_1$, and $\mathbf{par}_4 = X_2X_3$.

Let us call a Bayes net *subsequently causal* if the following conditions are met:

Condition 1 $X_1,X_2,...,X_n$ is a subsequent sequence.

Condition 2 $\mathbf{par}_i$ is causally sufficient for X_i relative to the resolution of $X_1X_2...X_{i-1}$.

This seems to correspond to what is usually meant when people speak abstractly of causal Bayes nets.

Not every family of variables can be arranged in a subsequently causal Bayes net. The variables must be subsequent, and once we have put them in the order in which they are settled, say $X_1X_2...X_n$, it must be true that $X_1X_2...X_{i-1}$ is causally sufficient for X_i for each i (Definition 1b). If these fundamental conditions are satisfied, then we obtain a Bayes net by drawing an arrow from X_i to X_j whenever $i<j$. If a subset C of $X_1X_2...X_{i-1}$ is causally irrelevant to X_i relative to the resolution of $X_1X_2...X_{i-1}$ (Definition 2c), we will still have a Bayes net when we omit the arrows from the variables in C to X_i.

We have just made precise the often repeated assertion that missing arrows in Bayes nets indicate that the variables from which the arrows might have pointed are irrelevant to the variables to which they might have pointed. In doing so, we have shown that this irrelevance is highly relative. The irrelevance of X_i to X_j is relative to the variables among $X_1,X_2,...,X_{j-1}$ from which we do draw arrows to X_j.

Although subsequent causality has an abstract clarity that facilitates discussion, it is elusive in practice. The Bayes nets that knowledge engineers construct for concrete problems usually are not subsequently causal. Instead of being determined sequentially, their variables tend to be related temporally in more complicated ways. Often they change together in time. Figures 2 and 3, drawn from the beginning pages of the proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, illustrate the point.

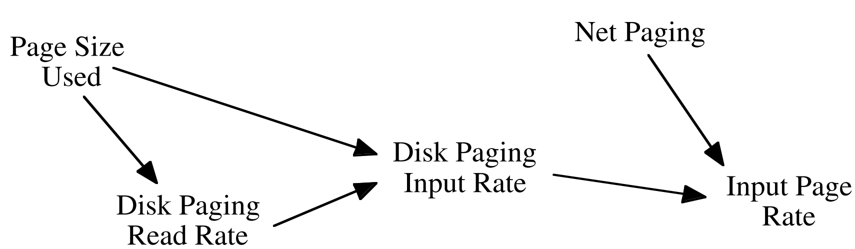


Figure 2 A fragment of a Bayes net for diagnosing bottlenecks in computer systems (Breese and Blake 1995). Most of the variables in this

net change together in time. The last variable, Input Page Rate, is a deterministic function of its parents.

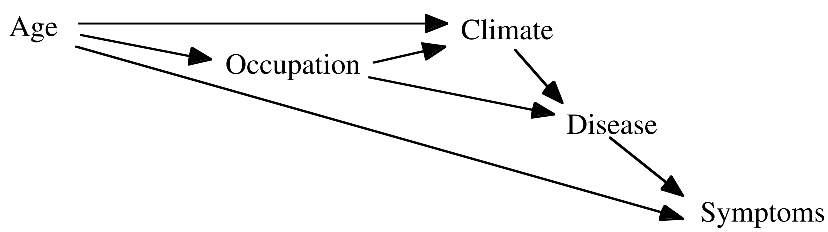


Figure 3 A simple medical problem (Buntine 1995). Many of these variables also change together.

4. Refining and Simplifying Causally Subsequent Bayes Nets

A Bayes net can be subsequently causal without being a complete account of the causal structure involved in the determination of a family of variables. By adding further variables, we can often enlarge a subsequently causal Bayes net to a more detailed one. Conversely, we can sometimes omit variables and still have a subsequently causal Bayes net. In the first case, we are *refining* the subsequently causal Bayes net. In the second case, we are *simplifying* it.

When we are not concerned with causal interpretation, there are no real constraints on the omission of variables from a Bayes net. Since any family of variables can be arranged in a Bayes net, we can always form a Bayes net from the variables that remain—possibly at the price of adding some further arrows among them (Pearl 1988). In the example at the bottom of Figure 4, for example, we need to add an arrow from X_1 to X_4 when we omit X_2 and X_3.

The desire to preserve subsequent causality does impose constraints on the omission of variables. Sometimes variables can be omitted without losing subsequent causality; sometimes they cannot be. Figure 4 shows two examples where subsequent causality is preserved, and Figure 5 shows two examples where it may be lost.

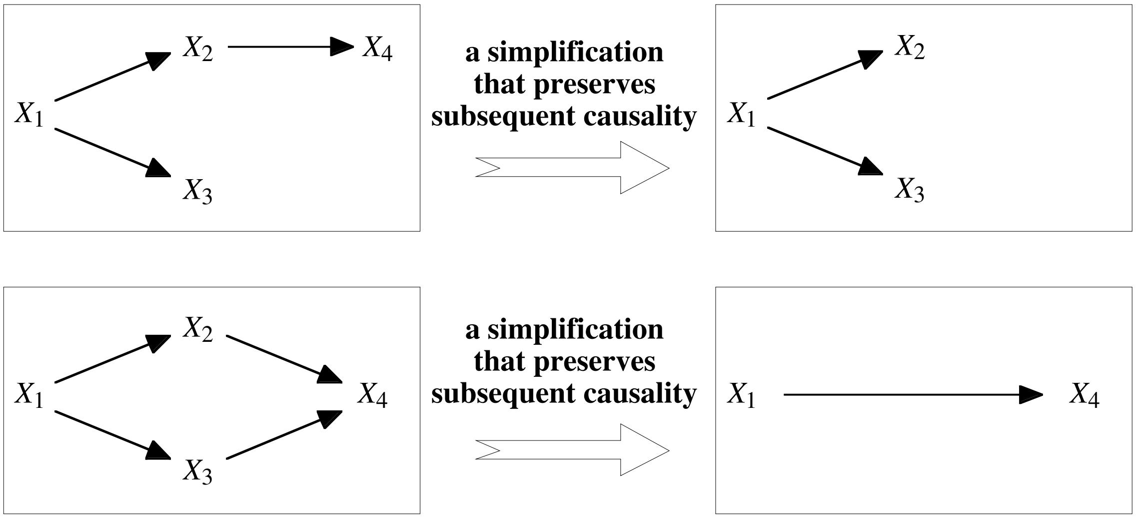


Figure 4 In the first example, the relations of causal sufficiency asserted by the simplification are already asserted in the original net. In the second example, the fact that X_1 is causally sufficient for X_4 follows, by Proposition 1, from the assertions in the original net.

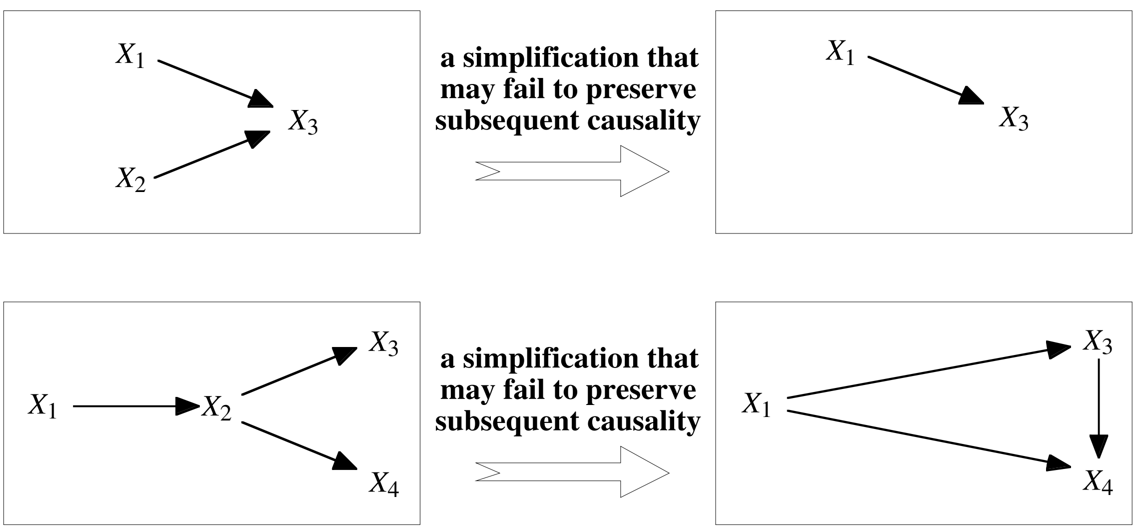


Figure 5 The nets on the right qualify as Bayes nets simply because they impose no conditional independence conditions on the joint distribution of the variables. But they are not necessarily subsequently causal. The fact that X_1X_2 is causally sufficient for X_3 does not ensure that X_1 alone is.

The fact that X_2 alone is causally sufficient for X_4 relative to the resolution of X_3 does not ensure that X_1X_3 is.

5. Probability Trees as a Semantics for Bayes Nets

The idea of subsequently causality is coherent and interesting but too narrow. As Figures 2 and 3 demonstrate, it does not encompass the whole range of causal meaning we can express using Bayes nets. In order to gain a broader understanding, we need to look beyond Bayes nets to a more fundamental level. We need to look at the semantics of probability in general and the semantics of Bayes nets in general.

In *The Art of Causal Conjecture* (especially Chapter 4), I argue that the semantics of probability goes beyond the austere framework of measure theory (Kolmogorov's axioms) that has been taken as fundamental in abstract treatments of the subject during most of this century. Probability is not static; it is inherently dynamic. Instead of interpreting probability statements in terms of a single static sample space with a probability measure, we should interpret them in terms of a probability tree. This proposal merely revives a very old idea, for probability trees were implicit and sometimes even explicit in the work of the inventors of probability theory.

Any observer can have a probability tree: a tree showing the possibilities for how the observer's knowledge may unfold, with branching probabilities indicating, at each step, odds the observer would give for what will happen next. But when we are concerned with causality, we are most interested in nature's probability tree—the probability tree that expresses nature's possibilities and probabilities.

Figure 6 shows a simple example of a probability tree. Here the question is whether a twelve-year old boy will remember to practice his saxophone on a summer afternoon, and the probability tree shows how this may be influenced by his other activities. We suppose that this probability tree is causal—e.g., that it is part of nature's probability tree. At the moment when Alex appears on Dennis's door step, even nature does not know whether the two boys will stay at Dennis's house, go to Sigmund's house, or go to Alex's house. The best nature can do is bet at odds given by the probabilities shown in the tree.

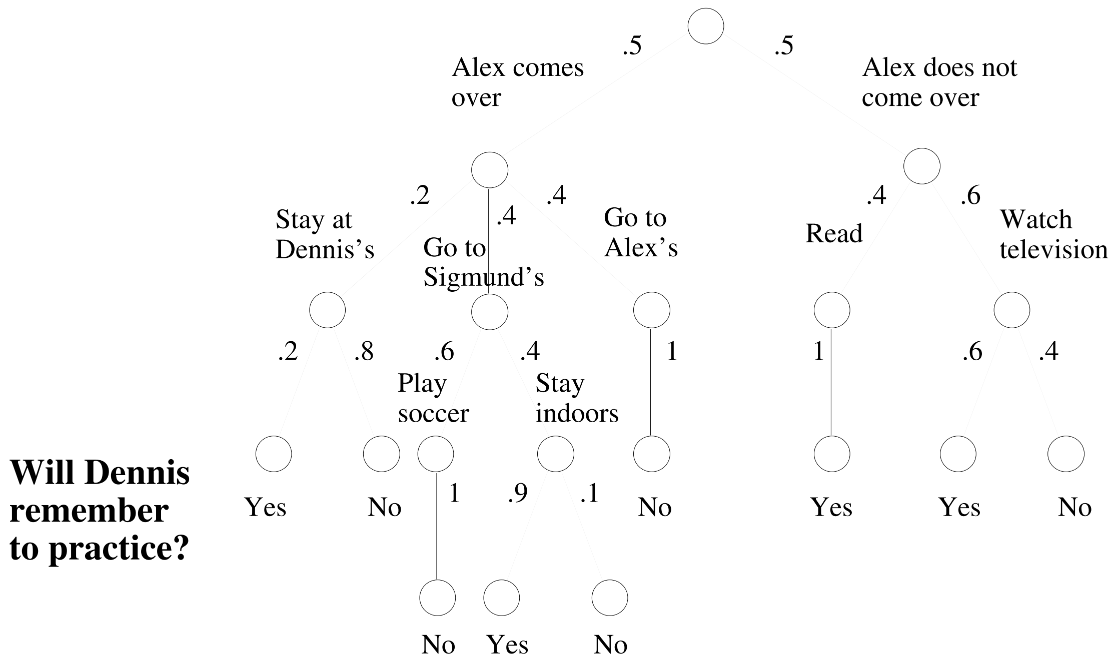


Figure 6 Nature's probability tree for whether Dennis will practice his saxophone on a summer afternoon.

A probability tree is a simple but rich structure. It combines features fundamental to the semantics of both logical and probabilistic representations in artificial intelligence. It has *situations*; these are the nodes in the tree. And it has a *sample space*, as required by standard twentieth-century probability theory; this is the set of all paths down the tree. Thus it has *events* (subsets of the sample space) and *variables* (functions on the sample space) in the usual probabilistic sense. The probability tree in Figure 6 has 19 situations and 9 paths. It has 2^9 (512) events and innumerable variables. Figure 6 itself shows the event that Dennis remembers to practice: this is the subset of the sample space consisting of the four paths ending in "Yes." Figure 7 shows the variable "Dennis's location at the end of the afternoon."

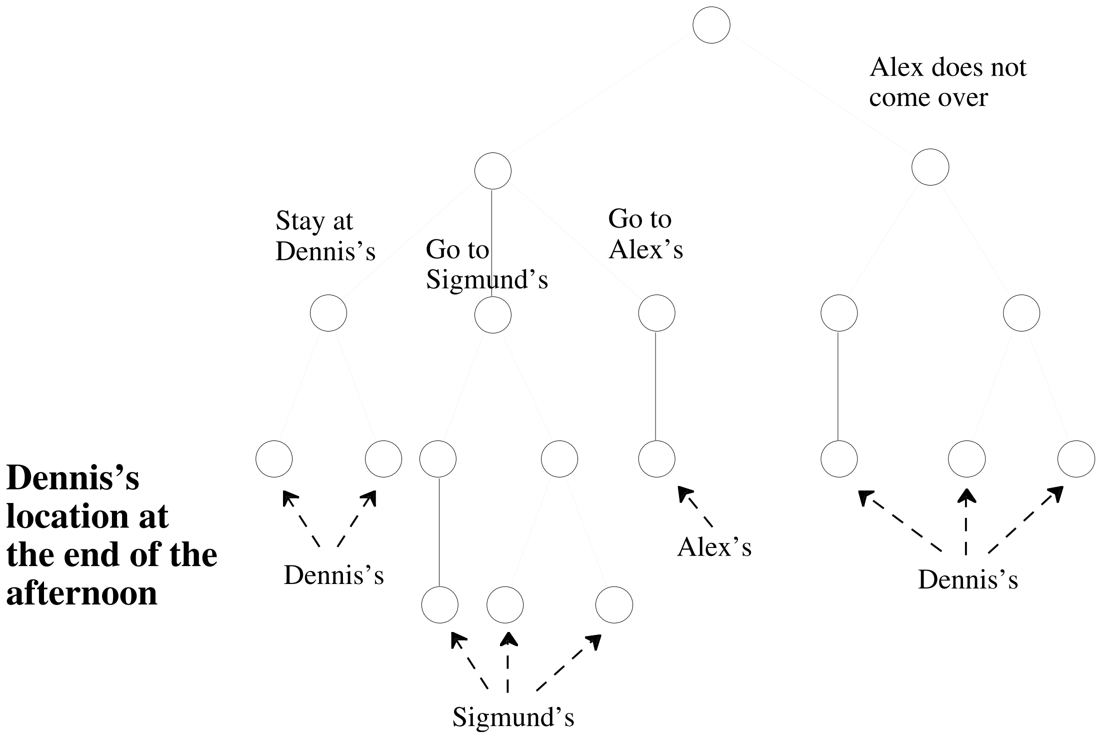


Figure 7 A variable in nature's probability tree for Dennis's afternoon.

The first thing we realize when we think about probability trees and Bayes nets is that a Bayes net can always be expanded into a probability tree: Figure 8 uses the Bayes net of Figure 1 to illustrate this point.

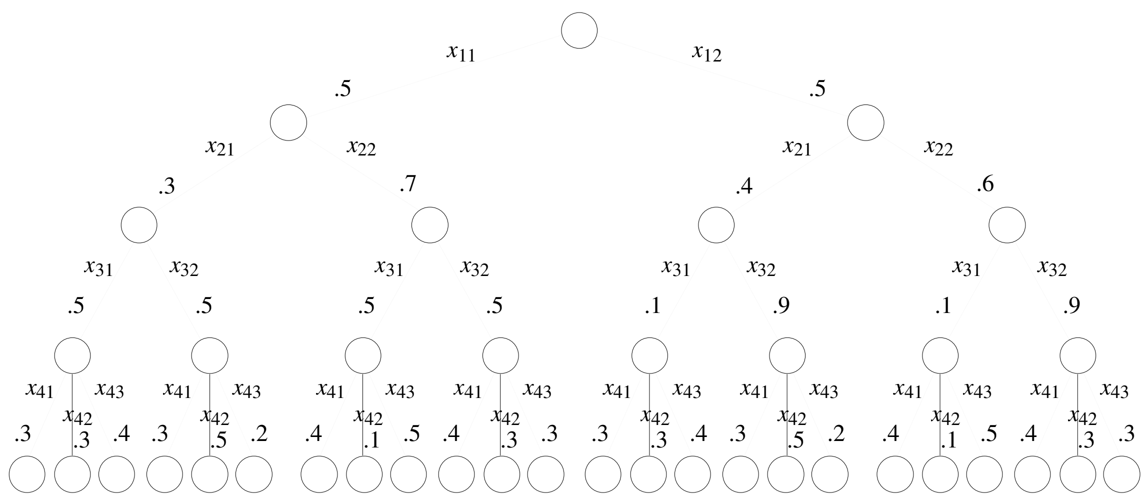


Figure 8 A probability tree for the Bayes net in Figure 1. The first branching, at the top of the tree, shows the determination of X_1, which has two possible values. The two branchings one step down show the

> determination of X_2, which again has two possible values, depending on how X_1 came out. The four branchings on the next level show the determination of X_3; the probabilities again depend on X_1. Finally, the eight branchings at the bottom show the determination of X_4, with probabilities depending on X_2 and X_3.

We may call a probability tree obtained from a Bayes net in the manner shown in Figure 8 the *subsequent expansion* of the Bayes net. The assertion that the Bayes net is subsequently causal is the same, essentially, as the assertion that its subsequent expansion is nature's probability tree. We are saying, in both cases, that the variables $X_1X_2...X_n$ are subsequent and that the probabilities for X_i at any point where $X_{i\text{-}1}$ has just been determined are those given by the table in the Bayes net. The probability tree simply makes the idea of "any point where $X_{i\text{-}1}$ has just been determined" clear, by showing the different possible points as different nodes.

It is a bit awkward, however, to call a subsequent expansion, or any other particular probability tree, nature's probability tree. We take it for granted the variables in our Bayes net are only a few of the things that nature observes, and hence we would prefer to think of "nature's probability tree" as a much larger object, in which these variables are minor details. In order to accommodate this thought, we need to step back from the Bayes net and think abstractly about subsequence, sufficiency, and irrelevance in a probability tree.

A few moments' thought reveals that the definitions laid out in Section 2 have clear and precise meanings in the context of a probability tree. When we spoke there, without a definite mathematical framework, about X being "settled in nature before" Y, the reader may have felt a bit at sea. But in a probability tree, it is quite clear whether a given variable is settled: it is settled in a given situation if it has the same value on all the paths through that situation. The situations that resolve a variable (the situations where it is just settled) form a cut across the tree, as in Figure 9; this is the variable's *resolving cut*. A variable's resolving cut has exactly one situation in common with every path down the tree. A variable *Y* is *subsequent* to a variable *X* if the situation that resolves *Y* never precedes the situation that resolves *X*, no matter what path we take down the tree.

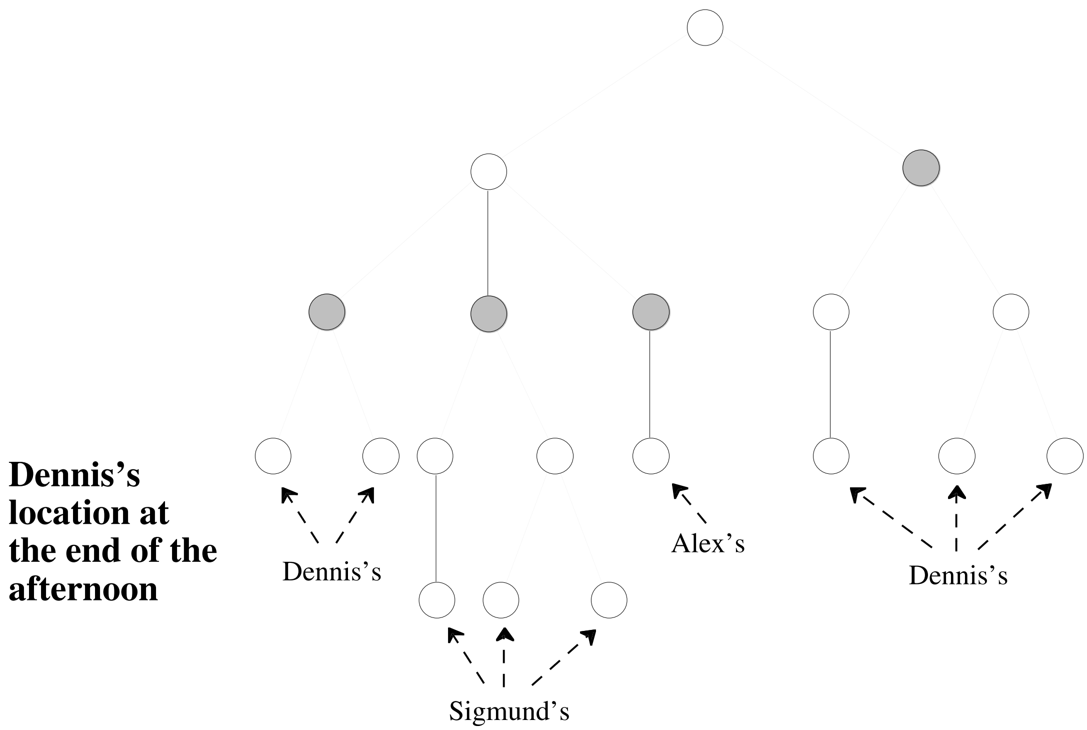


Figure 9 The resolving cut for Dennis's location: the situations where the location is resolved.

Causal sufficiency and irrelevance also have quite precise meanings in a probability tree. The branching probabilities in the tree determine probabilities and joint probabilities for all variables in all situations. Hence we can rephrase the definitions as follows:

Definition 1a *X* is *causally sufficient* for *Y* if the probabilities for *Y* in a situation that resolves *X* depend only on the value of *X* in that situation. (In other words, the probability distribution for *Y* is the same in any two situations that resolve *X* to the same value.)

Definition 1b More generally, *A* is *causally sufficient* for *Y* if the probabilities for *Y* in a situation that resolves *A* depend only on the configuration of *A* in that situation.

Definition 1c Yet more generally, *A* is *causally sufficient* for *C* if the probabilities for *C* in a situation that resolves *A* depend only on the configuration of *A* in that situation.

Definition 2a *X* is *causally relevant* to *Y* if there is a family *A* such that *AX* is causally sufficient for *Y* but *A* alone is not.

Definition 2b *X* is *causally relevant* to *Y relative* to a family *A* if *AX* is causally sufficient for *Y* but *A* alone is not.

Definition 1d Suppose $A \subseteq C$. Then *A* is *causally sufficient* for *Y relative to the resolution* of *C* if the probabilities for *Y* in a situation that resolves *C* depend only the configuration of *A*.

Definition 2c Suppose $B \subseteq C$. Then *B* is *causally irrelevant* for *Y relative to the resolution* of *C* if *C-B* is causally sufficient for *Y* relative to the resolution of *C*.

Notice that we have dropped from these definitions any requirement that the variable *Y* or family *C* be subsequent. This condition is not necessary in order for the definitions to be precise and meaningful; it was included in Section 2 in an attempt to make the concepts comprehensible without the explicit and precise framework provided by probability trees.

The fact that causal sufficiency is meaningful without any assumption of subsequence provides us immediately with a more general causal interpretation of Bayes nets. We noted in Section 3 that if a directed acyclic graph with variables as nodes satisfies two conditions,

Condition 1 $X_1,X_2,...,X_n$ is a subsequent sequence, and

Condition 2 $\mathbf{par}_i$ is causally sufficient for X_i relative to the resolution of $X_1X_2...X_{i-1}$, then it is a Bayes net, and we called such a Bayes net *subsequently causal*. It turns out (*The Art of Causal Conjecture*, Chapter 15, Section 3) that we still have a Bayes net if we weaken Condition 1 to

> **Condition 1'** The probabilities for X_i do not change on any step in the probability tree where the probabilities for $X_1X_2...X_{i-1}$ also change, unless the probabilities for $\mathbf{par}_i$ change on that step.

(This can alternatively be expressed by saying that X_i is independent, in the probability-tree sense, of $X_1X_2...X_{i-1}$ modulo $\mathbf{par}_i$.) In this case, the Bayes net still has a causal interpretation; we may call it a *causal Markov net*.

Figure 10 shows a simple Bayes net, $X_1 \rightarrow X_2$, that is causal Markov without being subsequently causal. In this example, X_2 precedes X_1. (A person chooses farm school or graduate school before settling on what kind of farmer or professor to be.) And yet X_1 is causally sufficient for X_2. (In any situation that resolves X_1, the value to which X_1 is resolved determines the probabilities for X_2; in fact it tells us what X_1 is for sure. When X_1 is resolved to Hogs; X_2 is Farmer with probability one, etc.)

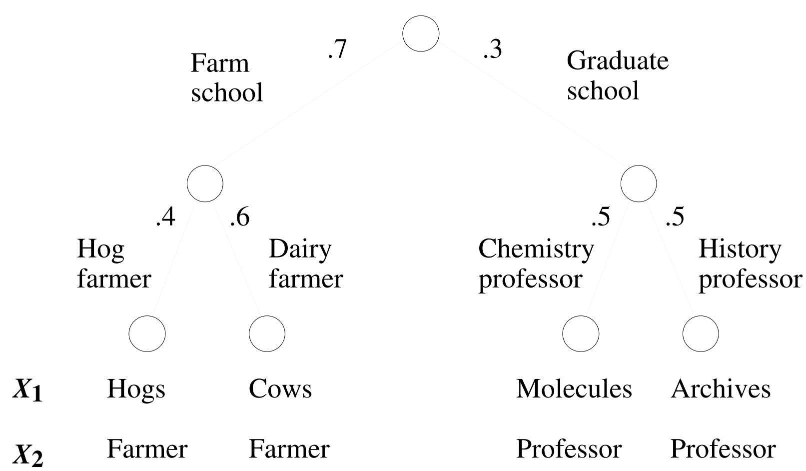


Figure 10 A probability tree in which the Bayes net $X_1 \rightarrow X_2$ is causal Markov but not subsequently causal. In the story told by this tree, a person first decides whether to go to farm school or graduate school. If she goes to farm school, she eventually becomes a hog or dairy farmer. If

> she goes to graduate school, she eventually becomes a chemistry or history professor. The variable X_1 names the principal objects with which she is professionally concerned, while X_2 is her general occupational category.

Figure 10 is a simple example of a deterministic Bayes net, one in which each variable with parents is fully determined by those parents. Such Bayes nets qualify as causal Markov whenever the variables without parents are independent in the probability-tree sense, but they may not be subsequently causal. Often, as in this example, variables are actually settled in nature before their children in the net. This illustrates the wisdom of saying "X is causally relevant to Y" instead of "X is a cause of Y" when there is an arrow from X to Y. A variable that is settled later can be causally relevant to Y, in the sense that it points to causes of Y, but we do not want to say that later things are the causes of earlier things.

Causal Markov nets are far more general than subsequently causal Bayes nets, but they themselves represent only one of many causal interpretations of Bayes nets in terms of probability trees. Another very important interpretation arises when we shift our attention from situations where the parents of a variable X are resolved to the steps in the tree where the expected values for the parents change, and we require that the expected values for X change proportionally on these steps. Bayes nets that satisfy this condition are called *causal path diagrams* (see Section 4 of Chapter 15 of *The Art of Causal Conjecture*). They often arise when we are concerned with progressive causation, as in Figures 2 and 3.

The idea of using probability trees as a semantics for causal representations can be extended beyond Bayes nets and other probabilistic representations to logical representations that have been used in discussions of causality. Indeed, since situations are integral to probability trees, these trees provide a semantics for the situation calculus and the temporal logics that elaborate it (Shoham 1988, Terenziani and Torasso 1995). In discussions that are concerned with human action and planning, however, it is necessary to generalize from probability trees to decision trees, in which some branchings are not labeled with probabilities (see Raiffa 1968 and Chapter 12 of *The Art of Causal Conjecture*,). This involves generalizing also from Bayes nets to influence diagrams, in which some variables are decision variables (Oliver and Smith 1990).

6. The Expressive Power of Bayes Nets

Historically, probability and decision trees preceded Bayes nets and influence diagrams, and many of the early users of Bayes nets and influence diagrams seem to have taken for granted that these diagrams should be interpreted in terms of probability or decision trees. In recent years, however, Bayes nets have been so popular and well understood that some research workers bridle at the idea that they should be interpreted in terms of any other representation. Indeed, these workers respond to the project of interpreting Bayes nets in terms of probability trees with the assertion that Bayes nets have just as much expressive power as probability trees.

The assertion that Bayes nets are equal to probability trees in expressive power is somewhat beside the point of this article, because we are thinking of probability trees as an underlying semantics or ontology, not as a representation that should compete with Bayes nets in practical work. But the assertion is also untrue, and it is important to understand the limitations of Bayes nets and to explore alternative causal representations that may be computationally more practical in some problems. Probability trees can provide a framework for this exploration, because the alternative causal representations, like Bayes nets, can be interpreted as partial descriptions of nature's probability tree (see Chapter 16 of *The Art of Causal Conjecture*).

It is obvious that not every probability tree can be obtained from a Bayes net in the way we obtained Figure 8 from Figure 1. Any probability tree obtained in this way must have (1) the same number of steps on every path, and (2) the same number of branches from every node at a given level. In order to obtain less symmetric probability trees from

Bayes nets, we evidently need to supply additional information, information that will direct the pruning away of certain unwanted branches. In order to generate Figure 6, for example, we can begin with the Bayes net in Figure 11 (the reader may supply the probabilities), but we will also need rather specific directions for pruning. As a first step, we can prune away branches with probability zero, but this produces only the probability tree shown in Figure 12, and we need further to be told which steps with probability one to remove.

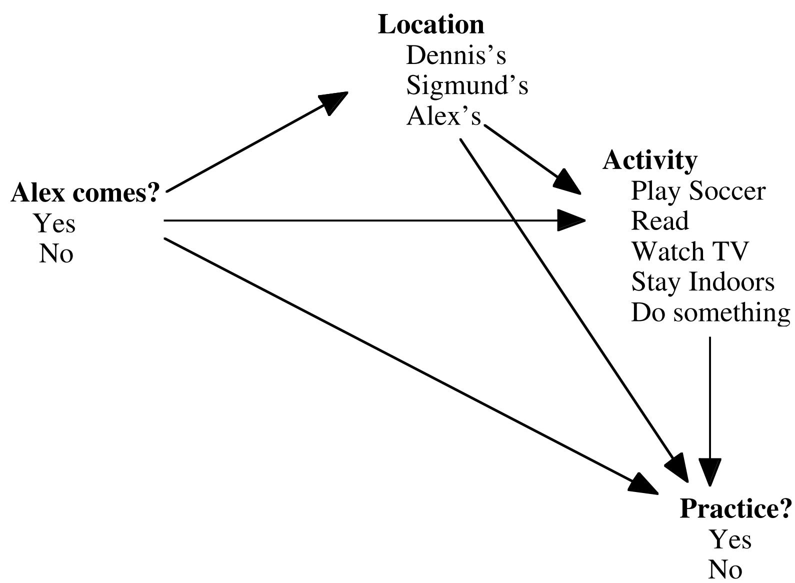


Figure 11 A Bayes net for Dennis's remembering to practice. The reader can supply the tables of probabilities. The variable "Activity" is somewhat artificial, inasmuch as "Stay indoors" and "Do something" may include some of the other activities; we call watching TV "Watch TV, "Stay indoors," or "Do something," depending on the circumstances, as indicated in Figure 12.

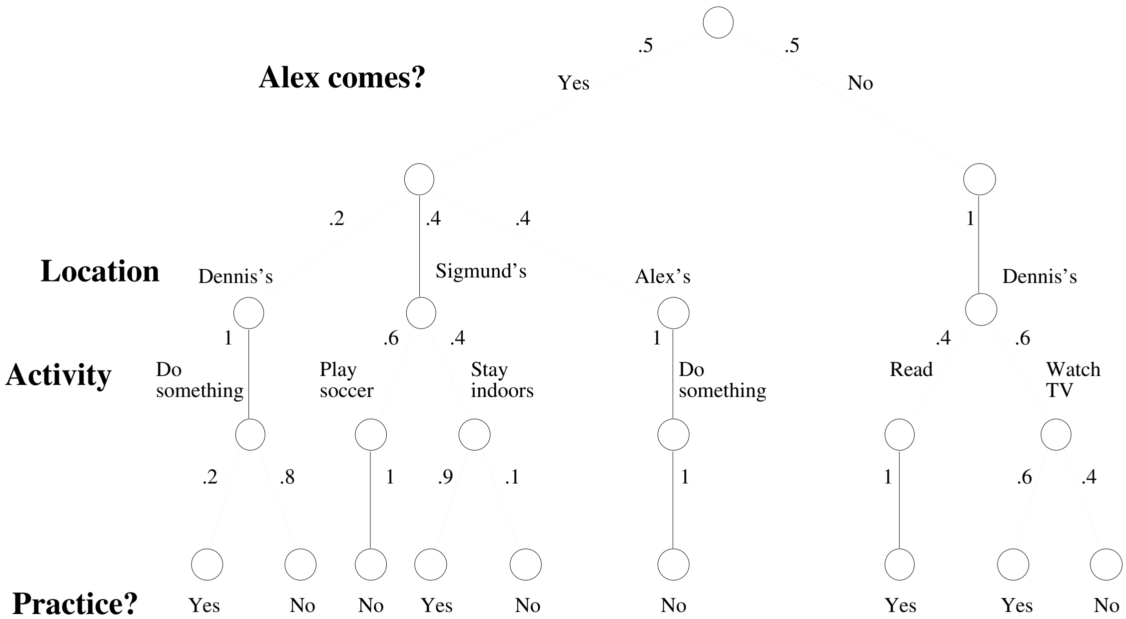


Figure 12 The probability tree initially generated by Figure 11. By selectively collapsing steps with probability one, we obtain from it the probability tree in Figure 6.

As Figures 11 and 12 reveal, we may need to define rather unnatural variables in order to represent a simple probability tree in the form of a Bayes net. The basic problem is that a Bayes net has a symmetry that a probability tree need not have. All the variables in a Bayes net must come into play no matter how things turn out, whereas some contingencies in a probability tree may enter the story down some branches. Figure 13 shows a striking example of this: only if surgery is performed does the question arise whether the patient will survive it.

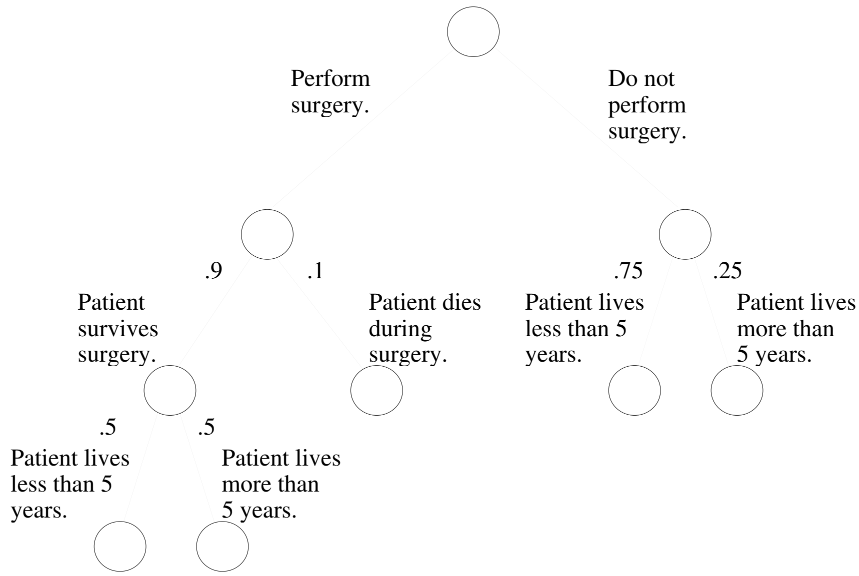


Figure 13 A probability tree that cannot be obtained naturally from a subsequent expansion of a Bayes net.

In summary, we may say that probability trees are strictly greater in expressive power than Bayes nets by themselves. Bayes nets are compact representations, and they are sufficiently useful that makes sense to expand their range of applicability by supplementing them with various kinds of collateral information. But we can best keep track of the meaning of this collateral information if we think of it as additional information about a probability tree. And we should not overlook the possibilities of radically different ways of partially describing nature's probability tree.

8. Refining and Simplifying Probability Trees

Since we think of nature as indefinitely complex, we will not be comfortable thinking of a probability tree as an accurate representation of nature unless we understand how it can be refined without being falsified. Fortunately, there is a straightforward concept of refining for probability trees, and it is consistent with the concept of refining for subsequently causal Bayes nets, which we explored in Section 4.

Figures 14 and 15 illustrate refining for probability trees. As these figures indicate, a refining matches each situation in the simplification (the less refined tree) with one or more situations in the refinement (the more refined tree). The refinement may also have situations not represented in the simplification. The probabilities must be consistent, in the sense that probabilities asserted for a situation in the simplification must remain true for all the corresponding situations in the refinement. If the trees are causal (if they are both simplifications of nature's tree), then these are nature's probabilities, and they will remain nature's probabilities no matter how much more detail is brought into the picture.

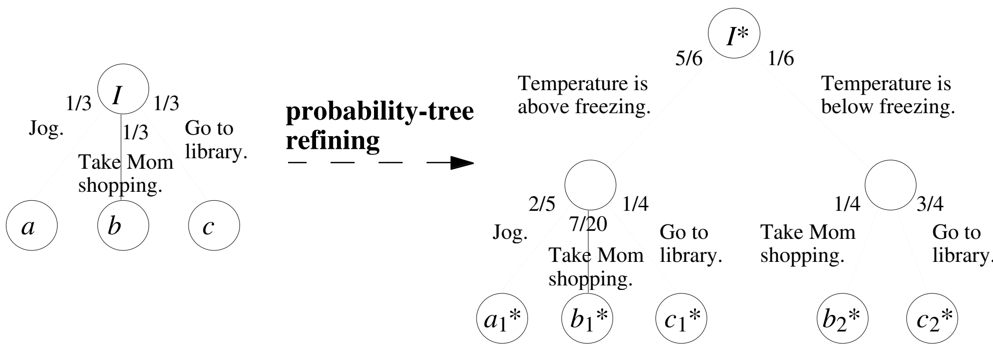


Figure 14 The temperature affects the possibility of Nell's jogging and the probabilities for whether she will take her mother shopping or go to the library.

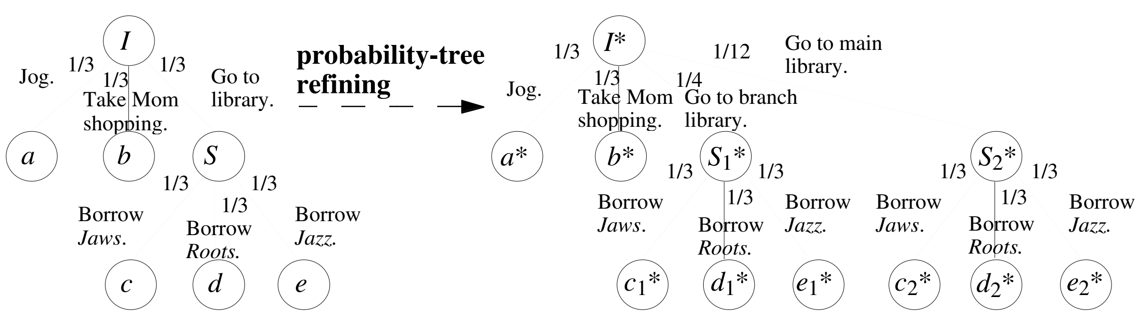


Figure 15 Nell's choice of library affects neither the possibilities nor the probabilities for what book she will borrow.

The concept of refining for probability trees is discussed at length in Chapter 13 of *The Art of Causal Conjecture*.

9. Conclusion

This article has discussed the representation of causality and causal relevance in Bayes nets and probability trees. Probability trees are the more fundamental representation; they can be thought as a semantics for Bayes nets and for other partial representations of the dynamics of nature.

Causal relevance is represented explicitly in a Bayes nets: missing arrows indicate that certain variables are irrelevant to nature's prediction of certain other variables. This irrelevance is relative, however, to the variables that are explicitly taken into account in the prediction of that variable in the Bayes net.

Variables do not play a fundamental role in a probability tree, and hence the probability tree does not explicitly represent relevance and irrelevance for variables. The assertion that a probability tree is causal contains, however, a very strong assertion about causal relevance. It says that relative to the events already represented down to a certain situation in a probability tree, all other details about what has happened in nature prior to that situation are irrelevant to nature's prediction of the later events represented in the probability tree. Within the theory of probability trees, this fact finds expression in the invariance of probabilities under refinement. It also finds expression in the judgments of irrelevance that are encoded in Bayes nets.

References

Bollen, Kenneth A. (1989). *Structural Equations with Latent Variables.* New York: Wiley.

Humphreys, Paul (1989). *The Chances of Explanation: Causal Explanation in the Social, Medical, and Physical Sciences*. Princeton, New Jersey: Princeton University Press.

Mackie, J. L. (1974). *The Cement of the Universe*. Oxford: Oxford University Press.

Marini, Margaret Mooney, and Burton Singer (1988). Causality in the Social Sciences. *Sociological Methodology* **18** 347-409.

Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems.* San Mateo, California: Morgan Kaufmann.

Raiffa, Howard (1968). *Decision Analysis.* Reading, Massachusetts: Addison-Wesley.

Shafer, Glenn (1996). *The Art of Causal Conjecture.* Cambridge, Massachusetts: The MIT Press.

Shoham, Yoav (1988). *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence.* Cambridge, Massachusetts: The MIT Press.

Terenziani, Paolo, and Pietro Torasso (1995). Time, action-types, and causation: an integrated analysis. *Computational Intelligence* **11** 529-552.