# Bayesian Networks

**Judea Pearl**
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
(310) 825-3243
(310) 825-2273 Fax
*judea@cs.ucla.edu*

## 1 INTRODUCTION

This paper surveys the historical development of Bayesian networks, summarizes their semantical basis and assesses their properties and applications vis a vis those of neural networks.

Bayesian networks are directed acyclic graphs (DAGs) in which the nodes represent variables of interest (e.g., the temperature of a device, the gender of a patient, a feature of an object, the occurrence of an event) and the links represent causal influences among the variables. The strength of an influence is represented by conditional probabilities that are attached to each cluster of parents-child nodes in the network.

Figure 1 illustrates a simple yet typical Bayesian network. It describes the causal relationships among the season of the year ($X_1$), whether rain falls ($X_2$) during the season, whether the sprinkler is on ($X_3$) during that season, whether the pavement would get wet ($X_4$), and whether the pavement would be slippery ($X_5$). All variables in this figure are binary, taking a value of either true or false, except the root variable $X_1$ which can take one of four values: Spring, Summer, Fall, or Winter. Here, the absence of a direct link between $X_1$ and $X_5$, for example, captures our understanding that the influence of seasonal variations on the slipperiness of the pavement is mediated by other conditions (e.g., the wetness of the pavement).

As this example illustrates, a Bayesian network constitutes a model of the environment rather than, as in many other knowledge representation schemes (e.g., rule-based systems and neural networks), a model of the reasoning process. It simulates, in fact, the causal mechanisms that operate in the environment, and thus allows the investigator to answer a variety of queries, including: associational queries, such as "Having observed $A$, what can we expect of $B$?"; abductive queries, such as "What is the most plausible explanation for a given set of observations?"; and control queries; such as "What will happen if we intervene and act on the environment?". Answers to the first type of query depend only on probabilistic knowledge of the domain, while answers to the second and third types rely on the causal knowledge embedded in the network. Both types of knowledge, associative and causal, can effectively be represented and processed in Bayesian networks.

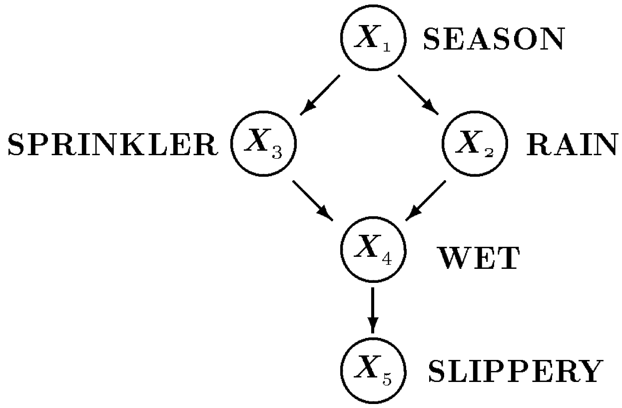


Figure 1: A Bayesian network representing causal influences among five variables.

The associative facility of Bayesian networks may be used to model cognitive tasks such as object recognition, reading comprehension, and temporal projections. For such tasks, the probabilistic basis of Bayesian networks offers a coherent semantics for coordinating top-down and bottom-up inferences, thus bridging information from high-level concepts and low-level percepts. This capability is important for achieving selective attention, that is, selecting the most informative next observation before actually making the observation. In certain structures, the coordination of these two modes of inference can be accomplished by parallel and distributed processes that communicate through the links in the network.

However, the most distinctive feature of Bayesian networks, stemming largely from their causal organization, is their ability to represent and respond to changing configurations. Any local reconfiguration of the mechanisms in the environment can be translated, with only minor modification, into an isomorphic reconfiguration of the network topology. For example, to represent a disabled sprinkler, we simply delete from the network all links incident to the node "Sprinkler". To represent a pavement covered by a tent, we simply delete the link between "Rain" and "Wet". This flexibility is often cited as the ingredient that marks the division between deliberative and reactive agents, and that enables the former to manage novel situations instantaneously, without requiring retaining or adaptation. Thus, Bayesian networks can model a wide spectrum of cognitive activities, ranging from low-level perception (reaction) to planning and explaining (deliberation).

# 2 HISTORICAL BACKGROUND

Networks employing directed acyclic graphs (DAGs) have a long and rich tradition, which began with the geneticist Sewall Wright [Wright 1921]. He developed a method called *path analysis*, which later became an established representation of causal models in economics, sociology, and psychology. *Recursive models* is the name given to such networks by statisticians seeking meaningful and effective decompositions of contingency tables. *Influence diagrams* represent another application of DAG representation developed for decision analysis. The primary role of a DAG in these applications is to provide an efficient description of the probability functions; once the network is configured, all subsequent computations are pursued by symbolic manipulation of probability expressions.

The potential for the network to work as a computational architecture, and hence as a model of cognitive activities, was noted in [Pearl 1982], where a distributed scheme was demonstrated for probabilistic updating on tree-structured networks. The motivation behind this particular development was the modeling of distributed processing in reading comprehension, where both top-down and bottom-up inferences are combined to form a coherent interpretation. This dual mode of reasoning is at the heart of Bayesian updating, and in fact motivated Reverend Bayes's original 1763 calculations of posterior probabilities (representing explanations), given prior probabilities (representing causes), and likelihood functions (representing evidence).

Bayesian networks have not attracted much attention in cognitive modeling circles, but they did in expert systems. The ability to coordinate bi-directional inferences filled a void in expert systems technology of the late 1970s, and it is in this area that Bayesian networks truly flourished. Over the past ten years, Bayesian networks have become a tool of great versatility and power, and they are now the most common representation scheme for probabilistic knowledge [Shafer 1990, Shachter 1990]. They have been used to aid in the diagnosis of medical patients and malfunctioning systems, to understand stories, to filter documents, to interpret pictures, to perform filtering, smoothing, and prediction, to facilitate planning in uncertain environments, and to study causation, nonmonotonicity, action, change, and attention. Some of these applications are described in a tutorial article by [Charniak 1991]; others can be found in [Pearl 1988] and [Shafer 1990].

# 3 FORMAL SEMANTICS

## 3.1 Bayesian Networks as Carriers of Probabilistic Information

Given a DAG $\Gamma$ and a joint distribution $P$ over a set $X = \{X_1, ..., X_n\}$ of discrete variables, we say that $\Gamma$ *represents* $P$ if there is a one-to-one correspondence between the variables in $X$ and the nodes of $\Gamma$, such that $P$ admits the recursive product decomposition

$$P(x_1, ..., x_n) = \prod_i P(x_i \mid \mathbf{pa}_i) \tag{1}$$

where $\mathbf{pa}_i$ are the direct predecessors (called *parents*) of $X_i$ in $\Gamma$. For example, the DAG in Figure 1 induces the decomposition

$$P(x_1, x_2, x_3, x_4, x_5) = P(x_1)\ P(x_2|x_1)\ P(x_3|x_1)\ P(x_4|x_2, x_3)\ P(x_5|x_4) \tag{2}$$

The recursive decomposition in Eq. (1) implies that, given its parent set $\mathbf{pa}_i$, each variable $X_i$ is conditionally independent of all its other predecessors $\{X_1, X_2, ..., X_{i-1}\} \backslash \mathbf{pa}_i$. Using Dawid's notation [Dawid 1979], we can state this set of independencies as

$$X_i \perp\!\!\!\perp \{X_1, X_2, ..., X_{i-1}\} \backslash \mathbf{pa}_i \mid \mathbf{pa}_i \quad i = 2, ..., n \tag{3}$$

Such a set of independencies is called *Markovian*, since it reflects the Markovian condition for state transitions: each state is rendered independent of the past, given its immediately preceding state. For example, the DAG of Figure 1 implies the following Markovian independencies:

$$X_2 \perp\!\!\!\perp \{0\} \mid X_1, \quad X_3 \perp\!\!\!\perp X_2 \mid X_1, \quad X_4 \perp\!\!\!\perp X_1 \mid \{X_2, X_3\}, \quad X_5 \perp\!\!\!\perp \{X_1, X_2, X_3\} \mid X_4 \tag{4}$$

In addition to these, the decomposition of Eq. (1) implies many more independencies, the sum total of which can be identified from the DAG using the graphical criterion of $d$-separation [Pearl 1988]:

**Definition 3.1** (d-*separation*) If $X, Y$, and $Z$ are three disjoint subsets of nodes in a DAG $\Gamma$, then $Z$ is said to $d$-separate $X$ from $Y$, denoted $d(X, Z, Y)_\Gamma$, if and only if there is no path from a node in $X$ to a node in $Y$ along which the following two conditions hold: (1) every node with converging arrows either is or has a descendant in $Z$, and (2) every other node is outside $Z$. A path satisfying these two conditions is said to be *active*; otherwise, it is said to be *blocked* (by $Z$). By *path* we mean a sequence of consecutive edges (of any directionality) in the DAG.

In Figure 1, for example, $X = \{X_2\}$ and $Y = \{X_3\}$ are $d$-separated by $Z = \{X_1\}$; the path $X_2 \leftarrow X_1 \rightarrow X_3$ is blocked by $X_1 \in Z$, while the path $X_2 \rightarrow X_4 \leftarrow X_3$ is blocked because $X_4$ and all its descendants are outside $Z$. Thus $d(X_2, X_1, X_3)$ holds in $\Gamma$. However, $X$ and $Y$ are not $d$-separated by $Z' = \{X_1, X_5\}$, because the path $X_2 \rightarrow X_4 \leftarrow X_3$ is rendered active by virtue of $X_5$, a descendant of $X_4$, being in $Z'$. Consequently, $d(X_2, \{X_1, X_5\}, X_3)$ does not hold in $\Gamma$; in words, learning the value of the consequence $X_5$ renders its causes $X_2$ and $X_3$ dependent, as if a pathway were opened along the arrows converging at $X_4$.

The $d$-separation criterion has been shown to be both necessary and sufficient relative to the set of distributions that are represented by a DAG $\Gamma$ (see Geiger et al. in [Shachter 1990]). In other words, there is a one-to-one correspondence between the set of independencies implied by the recursive decomposition of Eq. (1) and the set of triples $(X, Z, Y)$ that satisfy the $d$-separation criterion in $\Gamma$. Furthermore, the $d$-separation criterion can be tested in time linear in the number of edges in $\Gamma$. Thus, a DAG can be viewed as an efficient scheme for representing Markovian independence assumptions and for deducing and displaying all the logical consequences of such assumptions. Additional properties of DAGs and their applications to evidential reasoning in expert systems are discussed in [Pearl 1988, Shachter 1990, Spiegelhalter et al. 1993].

## 3.2 Bayesian Networks as Carriers of Causal Information

The interpretation of DAGs as carriers of independence assumptions does not necessarily imply causation and will in fact be valid for any set of Markovian independencies along any ordering (not necessarily causal or chronological) of the variables. However, the patterns of independencies portrayed in a DAG are typical of causal organizations and some of these patterns can only be given meaningful interpretation in terms of causation. Consider, for example, two independent events, $E_1$ and $E_2$, that have a common effect $E_3$. This triple represents an intransitive pattern of dependencies: $E_1$ and $E_3$ are dependent, $E_3$ and $E_2$ are dependent, yet $E_1$ and $E_2$ are independent. Such a pattern cannot be represented in undirected graphs because connectivity in undirected graphs is transitive. Likewise, it is not easily represented in neural networks, because $E_1$ and $E_2$ should turn dependent once $E_3$ is known. The DAG representation provides a perfect language for intransitive dependencies via the converging pattern $E_1 \rightarrow E_3 \leftarrow E_2$, which implies the independence of $E_1$ and $E_2$ as well as the dependence of $E_1$ and $E_3$ and of $E_2$ and $E_3$. The distinction between transitive and intransitive dependencies is the basis for the causal discovery systems of [Pearl & Verma 1991] and [Spirtes et al. 1993] (see Section 5).

However, the Markovian account still leaves open the question of how such intricate patterns of independencies relate to the more basic notions associated with causation, such as influence, manipulation, and control, which reside outside the province of probability theory. The connection is made in the mechanism-based account of causation.

The basic idea behind this account goes back to H. Simon and it was adapted in [Pearl & Verma 1991] for defining probabilistic causal theories, as follows. Each child-parents family in a DAG $\Gamma$ represents a deterministic function

$$X_i = f_i(\mathbf{pa}_i, \epsilon_i) \tag{5}$$

where $\mathbf{pa}_i$ are the parents of variable $X_i$ in $\Gamma$, and $\epsilon_i$, $0 < i < n$, are mutually independent, arbitrarily distributed random disturbances. Characterizing each child-parent relationship as a deterministic function, instead of the usual conditional probability $P(x_i \mid \mathbf{pa}_i)$, imposes equivalent independence constraints on the resulting distributions and leads to the same recursive decomposition that characterizes DAG models (see Eq. (1)). However, the functional characterization $X_i = f_i(\mathbf{pa}_i, \epsilon_i)$ also specifies how the resulting distributions would change in response to external interventions, since each function is presumed to represent a stable mechanism in the domain and therefore remains constant unless specifically altered. Thus, once we know the identity of the mechanisms altered by the intervention and the nature of the alteration, the overall effect of an intervention can be predicted by modifying the appropriate equations in the model of Eq. (5) and using the modified model to compute a new probability function of the observables.

The simplest type of external intervention is one in which a single variable, say $X_i$, is forced to take on some fixed value $x_i'$. Such *atomic* intervention amounts to replacing the old functional mechanism $X_i = f_i(\mathbf{pa}_i, \epsilon_i)$ with a new mechanism $X_i = x_i'$ governed by some external force that sets the value $x_i'$. If we imagine that each variable $X_i$ could potentially be subject to the influence of such an external force, then we can view each Bayesian network as an efficient code for predicting the effects of atomic interventions and of various combinations of such interventions, without representing these interventions explicitly.

This function-replacement operation yields a simple and direct transformation between the pre-intervention and the post-intervention distributions:

$$P_{x_i'}(x_1, ..., x_n) = \begin{cases} \frac{P(x_1, ..., x_n)}{P(x_i \mid \mathbf{pa}_i)} & \text{if } x_i = x_i' \\ 0 & \text{if } x_i \neq x_i' \end{cases} \tag{6}$$

which reflects the removal of the term $P(x_i \mid \mathbf{pa}_i)$ from the product decomposition of Eq. (1), since $\mathbf{pa}_i$ no longer influence $X_i$ [Goldszmidt & Pearl 1992]. Graphically, the removal of this term is equivalent to removing the links between $\mathbf{pa}_i$ and $X_i$ while keeping the rest of the network intact [Spirtes et al. 1993]. Transformations involving conjunctive actions can be obtained by straightforward generalization of Eq. (6).

The transformation in Eq. (6) exhibits all the properties we normally associate with actions, and it was therefore proposed as a solution to the frame problem and its two satellites, the ramification problem and the concurrency problem [Darwiche & Pearl 1994, Pearl 1994a]. For example, to represent the intervention "turning the sprinkler ON" in the network of Figure 1, we delete the link $X_1 \rightarrow X_3$ and fix the value of $X_3$ to ON. The resulting joint distribution on the remaining variables will be

$$P(x_1, x_2, x_4, x_5) = P(x_1)\ P(x_2|x_1)\ P(x_4|x_2, X_3 = \text{ON})\ P(x_5|x_4) \tag{7}$$

Note the difference between the action $do(X_3 = \text{ON})$ and the observation $X_3 = \text{ON}$. The latter is encoded by ordinary Bayesian conditioning, while the former by conditioning a mutilated graph, with the link $X_1 \rightarrow X_3$ removed. This mirrors indeed the difference between seeing and doing: after observing that the sprinkler is ON, we wish to infer that the season is dry, that it probably did not rain, and so on; no such inferences should be drawn in evaluating the effects the contemplated action "turning the sprinkler ON".

# 4 PROPERTIES AND ALGORITHMS

By providing graphical means for representing and manipulating probabilistic knowledge, Bayesian networks overcome many of the conceptual and computational difficulties of rule-based systems [Pearl 1988]. Their basic properties and capabilities can be summarized as follows:

1. Graphical methods make it easy to maintain consistency and completeness in probabilistic knowledge bases. They also prescribe modular procedures of knowledge acquisition which significantly reduce the number of assessments required.

2. Independencies can be dealt with explicitly. They can be articulated by an expert, encoded graphically, read off the network, and reasoned about, yet they forever remain robust to numerical imprecision.

3. Graphical representations uncover opportunities for efficient computation. Distributed updating is feasible in knowledge structures that are rich enough to exhibit intercausal interactions (e.g., "explaining away"). And, when extended by clustering or conditioning, tree-propagation algorithms are capable of updating networks of arbitrary topology [Pearl 1988, Shafer 1990].

4. The combination of predictive and abductive inferences resolves many problems encountered by first-generation expert systems and renders belief networks a viable model for cognitive functions requiring both top-down and bottom-up inferences.

5. The causal information encoded in Bayesian networks facilitates the analysis of action sequences, their consequences, their interaction with observations, and their expected utilities, and hence the synthesis of plans and strategies under uncertainty [Dean & Wellman 1991, Pearl 1994a].

6. The isomorphism between the topology of Bayesian networks and the stable mechanisms that operate in the environment facilitates modular reconfiguration of the network in response to changing conditions, and permits deliberative reasoning about novel situations.

The first algorithms proposed for probability updating in Bayesian networks used message-passing architecture and were limited to trees [Pearl 1982] and singly connected networks [Kim 1983]. The idea was to assign each variable a simple processor, forced to communicate only with its neighbors, and to permit asynchronous back-and-forth message-passing until equilibrium was achieved. Coherent equilibrium can indeed be achieved this way, but only in singly connected networks, where an equilibrium state occurs in time proportional to the diameter of the network.

Many techniques have been developed and refined to extend the tree-propagation method to general, multiply connected networks. Among the most popular are Shachter's method of node elimination, Lauritzen and Spiegelhalter's method of clique-tree propagation, and the method of loop-cut conditioning (see [Pearl 1988, Shafer 1990]).

Clique-tree propagation, the most popular of the three methods, works as follows. Starting with a directed network representation, the network is transformed into an undirected graph that retains all of its original dependencies. This graph, sometimes called a Markov network, is then triangulated to form local clusters of nodes (cliques) that are tree-structured. Evidence propagates from clique to clique by ensuring that the marginal probability of their intersection set is the same, regardless of which of the two cliques is marginalized. Finally, when the propagation process subsides, the posterior probability of an individual variable is computed by projecting (marginalizing) the distribution of the hosting clique onto this variable.

While the task of updating probabilities in general networks is NP-hard, the complexity for each of the three methods cited above is exponential in the size of the largest clique found in some triangulation of the network. It is fortunate that these complexities can be estimated prior to actual processing; when the estimates exceed reasonable bounds, an approximation method such as stochastic simulation [Pearl 1988] can be used instead. Learning techniques have also been developed for systematic updating of the conditional probabilities $P(x_i|\mathbf{pa}_i)$ so as to match empirical data (see Spiegelhalter and Lauritzen in [Shachter 1990]).

# 5 RECENT DEVELOPMENTS

**Causal Discovery**. One of the most exciting prospects in recent years has been the possibility of using Bayesian networks to discover causal structures in raw statistical data. Several systems have been developed for this purpose [Pearl & Verma 1991, Spirtes et al. 1993]. Technically, such discovery is feasible only if one is willing to accept weaker forms of guarantees weaker than those obtained through controlled randomized experiments: minimality and stability [Pearl & Verma 1991]. Minimality guarantees that any other structure compatible with the data is necessarily less specific, and hence less falsifiable and less trustworthy, than the one(s) inferred. Stability ensures that any alternative structure compatible with the data must be less stable than the one(s) inferred; namely, slight fluctuations in experimental conditions will render that structure no longer compatible with the data. With these forms of guarantees, the theory provides criteria for identifying genuine and spurious causes, with or without temporal information, and yields algorithms for recovering causal structures with hidden variables from empirical data.

**Plain Beliefs.** In mundane decision making, beliefs are revised not by adjusting numerical probabilities but by tentatively accepting some sentences as "true for all practical purposes". Such sentences, often named *plain beliefs*, exhibit both logical and probabilistic character. As in classical logic, they are propositional and deductively closed; as in probability, they are subject to retraction and to varying degrees of entrenchment [Goldszmidt & Pearl 1992].

Bayesian networks can be adopted to model the dynamics of plain beliefs by replacing ordinary probabilities with non-standard probabilities, that is, probabilities that are infinitesimally close to either zero or one. This amounts to taking an "order of magnitude" approximation of empirical frequencies, and adopting new combination rules tailored to reflect this approximation. The result is an integer-addition calculus, very similar to probability calculus, with summation replacing multiplication and minimization replacing addition. A plain belief is then identified as a proposition whose negation obtains an infinitesimal probability (i.e., an integer greater than zero).

This combination of infinitesimal probabilities with the causal information encoded by the structure of Bayesian networks facilitates linguistic communication of belief commitments, explanations, actions, goals, and preferences, and serves as the basis for current research on qualitative planning under uncertainty [Darwiche & Pearl 1994, Goldszmidt & Pearl 1992, Pearl 1994b].

# 6 DISCUSSION

The most distinctive characteristics of Bayesian networks are their ability to faithfully represent causal relationships, to combine top-down and bottom-up inferences, and to adapt to changing conditions by updating the probability measures attached to the links. Although Bayesian networks can model a wide spectrum of cognitive activity, their greatest strength is in causal reasoning, which, in turn, facilitates reasoning about actions, explanations, counterfactuals, and preferences. Such capabilities are not easily implemented in neural networks, whose strengths lie in quick adaptation of simple motor-visual functions. Except for the common ability to perform distributed inferencing, the relation between Bayesian networks and neural networks is rather tenuous. For example, there are very few neural features in Bayesian networks: weights, sums, and sigmoids play no significant role; all computational units represent familiar linguistic notions; and deployment of bi-directional messages in acyclic structures has no known biological bias.

Some questions arise: Does an architecture resembling that of Bayesian networks exist anywhere in the human brain? If not, how does the brain perform those cognitive functions in which Bayesian networks excel? The answer is, I speculate, that nothing resembling Bayesian networks actually resides permanently in the brain. Instead, fragmented structures of causal organizations are constantly being assembled on the fly, as needed, from a stock of functional building blocks. Each such building block is specialized to handle a narrow context of experience and is probably embodied in an architecture of a neural network. For example, the network of Figure 1 may be assembled from several neural networks, one specializing in the experience surrounding seasons and rains, another in the properties of wet pavements, and so forth. Such specialized networks are probably stored permanently in some mental library, from which they are drawn and assembled into the structure shown in Figure 1 only when a specific problem presents itself, for example, to determine whether an operating sprinkler could explain why a certain person slipped and broke a leg in the middle of a dry season.

I believe the properties of Bayesian networks will be useful to scientists studying higher cognitive functions, where the problem of organizing and supervising large assemblies of specialized neural networks becomes important.

## Acknowledgments

The research was partially supported by Air Force grant #F49620-94-1-0173, NSF grant #IRI-9200918, and Northrop-Rockwell Micro grant #93-124.

# References

[Charniak 1991] * Charniak, E., "Bayesian networks without tears," *AI Magazine*, 12 (4), 50-63, 1991.

[Darwiche & Pearl 1994] Darwiche, A., and Pearl, J., "Symbolic causal networks for planning under uncertainty," in *Symposium Notes of the 1994 AAAI Spring Symposium on Decision-Theoretic Planning*, Stanford, CA, 41-47, 1994.

[Dawid 1979] Dawid, A.P., "Conditional independence in statistical theory," *Journal of the Royal Statistical Society, Series A*, 41, 1-31, 1979.

[Dean & Wellman 1991] * Dean, T.L., and Wellman, M.P., *Planning and Control*, Morgan Kaufmann, San Mateo, CA, 1991.

[Goldszmidt & Pearl 1992] Goldszmidt, M., and Pearl, J., "Default ranking: A practical framework for evidential reasoning, belief revision and update," in *Proceedings of the Third International Conference on Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, CA, 661-672, 1992.

[Kim 1983] Kim, J.H., and Pearl, J., "A computational model for combined causal and diagnostic reasoning in inference systems," in *Proceedings IJCAI-83*, Karlsruhe, Germany, 190-193, 1983.

[Pearl 1982] Pearl, J., "Reverend Bayes on inference engines: A distributed hierarchical approach," in *Proceedings AAAI National Conference on AI*, Pittsburgh, PA, 133-136, 1982.

[Pearl 1988] * Pearl, J., *Probabilistic Reasoning in Intelligence Systems*, Morgan Kaufmann, San Mateo, CA, 1988. (Revised 2nd printing, 1991).

[Pearl 1994a] Pearl, J., "A probabilistic calculus of actions," in *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence (UAI-94)*, Morgan Kaufmann, San Mateo, CA, 1994.

[Pearl 1994b] * Pearl, J., "From Adams' conditionals to default expressions, causal conditionals, and counterfactuals," to appear in E. Eells and B. Skyrms (Eds.), *Probability and Conditionals*, Cambridge University Press, New York, NY, 47-74, 1994.

[Pearl & Verma 1991] Pearl, J., and Verma, T., "A theory of inferred causation," in J.A. Allen, R. Fikes, and E. Sandewall (Eds.), *Principles of Knowledge Representation and Reasoning: Proceeding of the Second International Conference*, Morgan Kaufmann, San Mateo, CA, 441-452, 1991.

[Shachter 1990] * Shachter, R.D. (Ed.), Special Issue on Influence Diagrams, *Networks: An International Journal*, 20 (5), 1990.

[Shafer 1990] * Shafer, G., and Pearl, J. (Eds.), *Readings in Uncertain Reasoning*, Morgan Kaufmann, San Mateo, CA, 1990.

[Spiegelhalter et al. 1993] Spiegelhalter, D.J., Lauritzen, S.L., Dawid, P.A., and Cowell, R.G., "Bayesian analysis in expert systems," *Statistical Science*, 8 (3), 219-247, 1993.

[Spirtes et al. 1993] Spirtes, P., Glymour, C., and Schienes, R., *Causation, Prediction, and Search*, Springer-Verlag, New York, 1993.

[Wright 1921] Wright, S., "Correlated and causation," *Journal of Agricultural Research*, 20, 557-585, 1921.