# Bayesian Artificial Intelligence

## SECOND EDITION

## Kevin B. Korb
## Ann E. Nicholson

# Bayesian Artificial Intelligence

**SECOND EDITION**

Chapman & Hall/CRC
# Computer Science and Data Analysis Series

The interface between the computer and statistical sciences is increasing, as each discipline seeks to harness the power and resources of the other. This series aims to foster the integration between the computer sciences and statistical, numerical, and probabilistic methods by publishing a broad range of reference works, textbooks, and handbooks.

**SERIES EDITORS**
David Blei, Princeton University
David Madigan, Rutgers University
Marina Meila, University of Washington
Fionn Murtagh, Royal Holloway, University of London

Proposals for the series should be sent directly to one of the series editors above, or submitted to:

**Chapman & Hall/CRC**
4th Floor, Albert House
1-4 Singer Street
London EC2A 4BQ
UK

---

## Published Titles

Bayesian Artificial Intelligence, Second Edition
*Kevin B. Korb and Ann E. Nicholson*

Clustering for Data Mining:
   A Data Recovery Approach
*Boris Mirkin*

Computational Statistics Handbook with
   MATLAB®, Second Edition
*Wendy L. Martinez and Angel R. Martinez*

Correspondence Analysis and Data
   Coding with Java and R
*Fionn Murtagh*

Design and Modeling for Computer
   Experiments
*Kai-Tai Fang, Runze Li, and Agus Sudjianto*

Exploratory Data Analysis with MATLAB®
*Wendy L. Martinez and Angel R. Martinez*

Exploratory Multivariate Analysis by
   Example Using R
*François Husson, Sébastien Lê, and
   Jérôme Pagès*

Introduction to Data Technologies
*Paul Murrell*

Introduction to Machine Learning
   and Bioinformatics
*Sushmita Mitra, Sujay Datta,
   Theodore Perkins, and George Michailidis*

Microarray Image Analysis:
   An Algorithmic Approach
*Karl Fraser, Zidong Wang, and Xiaohui Liu*

Pattern Recognition Algorithms for
   Data Mining
*Sankar K. Pal and Pabitra Mitra*

R Graphics
*Paul Murrell*

R Programming for Bioinformatics
*Robert Gentleman*

Semisupervised Learning for
   Computational Linguistics
*Steven Abney*

Statistical Computing with R
*Maria L. Rizzo*

# Bayesian Artificial Intelligence

### SECOND EDITION

## Kevin B. Korb
## Ann E. Nicholson

*To Judea Pearl and Chris Wallace*

# *Contents*

# List of Figures

# List of Tables

# *Preface*

Bayesian Artificial Intelligence, in our understanding, is the incorporation of Bayesian inferential methods in the development of a software architecture for an Artificial Intelligence (AI). We believe that important ingredients of such an architecture will be Bayesian networks and the Bayesian learning of Bayesian networks (Bayesian causal discovery) from observation and experiment. In this book we present the elements of Bayesian network technology, automated causal discovery, learning probabilities from data, and examples and ideas about how to employ these technologies in developing probabilistic expert systems, which we call Knowledge Engineering with Bayesian Networks.

This is a very practical project, because data mining with Bayesian networks (applied causal discovery) and the deployment of Bayesian networks in industry and government are two of the most promising areas in applied AI today. But it is also a very theoretical project, because the achievement of a Bayesian AI would be a major theoretical achievement.

With our title there are a number of subjects we could naturally include, but have not. Thus, another necessary aspect of an effective Bayesian AI will be the learning of concepts, and hierarchies of concepts. Bayesian methods for concept formation exist (e.g., Chris Wallace's Snob, Wallace and Boulton, 1968), but we do not treat them here. We could also have discussed function discovery, polynomial curve fitting, time series modeling, etc. We have chosen to hew close to the theme of using and discovering Bayesian networks both because this is our own main research area and because, important as the other Bayesian learning methods are, we believe the Bayesian network technology is central to the overall project. We have added for this edition a treatment of the use of Bayesian networks for prediction (classification).

Our text differs from others available on Bayesian networks in a number of ways. We aim at a practical and accessible introduction to the main concepts in the technology, while paying attention to foundational issues. Most texts in this area require somewhat more mathematical sophistication than ours; we presuppose only a basic understanding of algebra and calculus. Also, we give roughly equal weight to the causal discovery of networks and to the Bayesian inference procedures using a network once found. Most texts either ignore causal discovery or treat it lightly. Richard Neapolitan's book, *Learning Bayesian Networks* (2003), is an exception, but it is more technically demanding than ours. Another distinguishing feature of our text is that we advocate a causal interpretation of Bayesian networks, and we discuss the use of Bayesian networks for causal modeling. We also illustrate various applications of the technology at length, drawing upon our own applied research. We hope that these illustrations will be of some interest and indicate some of the possibilities for the

technology. Our text is aimed at advanced undergraduates in computer science who have some background in artificial intelligence and at those who wish to engage in applied or pure research in Bayesian network technology.

A few remarks about notation before we begin. The notation special to Bayesian networks (or to our treatment of them) will be introduced as we proceed; first introductions of notation (including general mathematical notation) and acronyms will be recorded, with page numbers, in Appendix A. When first introducing new terminology we shall employ boldface to point it out; thus, for example, the first appearance (after this) of "Bayesianism" will be in boldface.

Here we describe the simplest aspects of the notation we adopt. First, variables (nodes in the network) will be named, with the names being capitalized and usually italicized (e.g., *Y*, *Alarm*, *Cancer*). Sets of variables will be set in boldface (e.g., **X**, **Y**). The values that variables take will not be capitalized, but will be italicized; thus, to assert that the alarm is on, we might write *Alarm = on*. Values abbreviated to single letters, such as *True (T)* and *False (F)*, however, will be capitalized. Where no confusion is likely to arise, variables and values may be abbreviated.

The book Web site is `http://www.csse.monash.edu.au/bai` and contains a variety of aids for study, including example Bayesian networks and data sets. Instructors can email us for sample solutions to many of the problems in the text.

---

## New in the second edition

We have found and corrected numerous mistakes in the first edition, most of the credit for which belongs to readers and students (see below). We claim full credit for all the errors remaining or newly introduced. Other new introductions include: §4.7, object-oriented Bayesian networks, Chapter 7, Bayesian Network Classifiers, which describes naive Bayes models and other classifiers; §9.7, which addresses two foundational problems with causal discovery as well as Markov blanket discovery; §9.8, which treats methods of evaluating causal discovery programs; substantial new material in Chapter 10, including discussions of many common modeling errors; and new applications and case studies in Chapters 5 and 11. The uses of causal interventions to understand and reason with causal Bayesian networks receive fuller coverage in a number of places. The Evaluation Chapter of the first edition has been divided, enhanced and scattered: evaluation of prediction goes into Chapter 7, evaluation of causal discovery into Chapter 9 and sensitivity analysis and related issues into Chapter 10. The result is a somewhat thicker book and one we hope is more useful for students and practitioners.

# Acknowledgments

# *About the Authors*

**Kevin B. Korb**, Ph.D., earned his doctorate in the philosophy of science at Indiana University, Bloomington (1992), working on the philosophical foundations for the automation of Bayesian reasoning. Since then he has lectured at Monash University, Melbourne, Australia, in computer science, combining his interests in philosophy of science and artificial intelligence in work on understanding and automating inductive inference, the use of MML in learning causal theories, artificial evolution of cognitive and social behavior and modeling Bayesian and human reasoning in the automation of argumentation.

**Ann E. Nicholson**, D.Phil., did her undergraduate computer science studies at the University of Melbourne, Australia, and her doctorate in the robotics research group at Oxford University, UK (1992), working on dynamic Bayesian networks for discrete monitoring. She then spent two years at Brown University, Providence, Rhode Island, as a post-doctoral research fellow before taking up a lecturing position at Monash University, Melbourne, Australia, in computer science. Her general research area is AI methods for reasoning under uncertainty, while her current research focus is on knowledge engineering with Bayesian networks and applications of Bayesian networks.

# Part I

# PROBABILISTIC REASONING

# 1

## *Bayesian Reasoning*

## 1.1   Reasoning under uncertainty

Artificial intelligence (AI), should it ever exist, will be an intelligence developed by humans, implemented as an artifact. The level of intelligence demanded by Alan Turing's famous test (1950) — the ability to fool ordinary (unfoolish) humans about whether the other end of a dialogue is being carried on by a human or by a computer — is some indication of what AI researchers are aiming for. Such an AI would surely transform our technology and economy. We would be able to automate a great deal of human drudgery and paperwork. Since computers are universal, programs can be effortlessly copied from one system to another (to the consternation of those worried about intellectual property rights!), and the labor savings of AI support for bureaucratic applications of rules, medical diagnosis, research assistance, manufacturing control, etc. promises to be enormous. If a serious AI is ever developed.

There is little doubt that an AI will need to be able to reason logically. An inability to discover, for example, that a system's conclusions have reached inconsistency is more likely to be debilitating than the discovery of an inconsistency itself. For a long time there has also been widespread recognition that practical AI systems shall have to cope with **uncertainty** — that is, they shall have to deal with incomplete evidence leading to beliefs that fall short of knowledge, with fallible conclusions and the need to recover from error, called **non-monotonic reasoning**. Nevertheless, the AI community has been slow to recognize that any serious, general-purpose AI will need to be able to reason probabilistically, what we call here **Bayesian reasoning**.

There are at least three distinct forms of uncertainty which an intelligent system operating in anything like our world shall need to cope with:

1. **Ignorance.** The limits of our knowledge lead us to be uncertain about many things. Does our poker opponent have a flush or is she bluffing?

2. **Physical randomness or indeterminism.** Even if we know everything that we might care to investigate about a coin and how we impart spin to it when we toss it, there will remain an inescapable degree of uncertainty about whether it will land heads or tails when we toss it. A die-hard determinist might claim otherwise, that some unimagined amount of detailed investigation might someday reveal which way the coin will fall; but such a view is for the foreseeable future a mere act of scientistic faith. We are all practical indeterminists.

3. **Vagueness.** Many of the predicates we employ appear to be vague. It is often

unclear whether to classify a dog as a spaniel or not, a human as brave or not, a thought as knowledge or opinion.

**Bayesianism** is the philosophy that asserts that in order to understand human opinion as it ought to be, constrained by ignorance and uncertainty, the probability calculus is the single most important tool for representing appropriate strengths of belief. In this text we shall present Bayesian computational tools for reasoning with and about strengths of belief as probabilities; we shall also present a Bayesian view of physical randomness. In particular we shall consider a probabilistic account of causality and its implications for an intelligent agent's reasoning about its physical environment. We will not address the third source of uncertainty above, vagueness, which is fundamentally a problem about semantics and one which has no good analysis so far as we are aware.

## 1.2   Uncertainty in AI

The successes of formal logic have been considerable over the past century and have been received by many as an indication that logic should be the primary vehicle for **knowledge representation** and reasoning within AI. **Logicism** in AI, as this has been called, dominated AI research in the 1960s and 1970s, only losing its grip in the 1980s when artificial neural networks came of age. Nevertheless, even during the heyday of logicism, any number of practical problems were encountered where logic would not suffice, because uncertain reasoning was a key feature of the problem. In the 1960s, medical diagnosis problems became one of the first attempted application areas of AI programming. But there is no symptom or prognosis in medicine which is strictly logically implied by the existence of any particular disease or syndrome; so the researchers involved quickly developed a set of "probabilistic" relations. Because probability calculations are hard — in fact, NP hard in the number of variables (Cooper, 1990) (i.e., computationally intractable; see §1.11) — they resorted to implementing what has subsequently been called "naive Bayes" (or, "Idiot Bayes"), that is, probabilistic updating rules which assume that symptoms are independent of each other given diseases.[1]

The independence constraints required for these systems were so extreme that the systems were received with no wide interest. On the other hand, a very popular set of expert systems in the 1970s and 1980s were based upon Buchanan and Shortliffe's MYCIN, or the uncertainty representation within MYCIN which they called **certainty factors** (Buchanan and Shortliffe, 1984). Certainty factors (CFs) were obtained by first eliciting from experts a "degree of increased belief" which some evidence $e$ should imply for a hypothesis $h$, $MB(h,e) \in [0,1]$, and also a corresponding

---

[1]We will look at naive Bayes models for prediction in Chapter 7.

"degree of increased disbelief," $MD(h,e) \in [0,1]$. These were then combined:

$$CF(h,e) = MB(h,e) - MD(h,e) \in [-1,1]$$

This division of changes in "certainty" into changes in belief and disbelief reflects the curious notion that belief and disbelief are not necessarily related to one another (cf. Buchanan and Shortliffe, 1984, section 11.4). A popular AI text, for example, sympathetically reports that "it is often the case that an expert might have confidence 0.7 (say) that some relationship is true and have no feeling about it being not true" (Luger and Stubblefield, 1993, p. 329). The same point can be put more simply: experts are often inconsistent. Our goal in Bayesian modeling is, at least largely, to find the most accurate representation of a real system about which we may be receiving inconsistent expert advice, rather than finding ways of modeling the inconsistency itself.

Regardless of how we may react to this interpretation of certainty factors, no operational semantics for CFs were provided by Buchanan and Shortliffe. This meant that no real guidance could be given to experts whose opinions were being solicited. Most likely, they simply assumed that they were being asked for conditional probabilities of $h$ given $e$ and of $\neg h$ given $e$. And, indeed, there finally was a probabilistic semantics given for certainty factors: David Heckerman (1986) proved that a consistent probabilistic interpretation of certainty factors[2] would once again require strong independence assumptions: in particular that, when combining multiple pieces of evidence, the different pieces of evidence must always be independent of each other. Whereas this appears to be a desirable simplification of **rule-based** systems, allowing rules to be "modular," with the combined impact of diverse evidence being a compositional function of their separate impacts it is easy to demonstrate that the required independencies are frequently unavailable. The price of rule-based simplicity is irrelevance.

Bayesian networks provide a natural representation of probabilities which allow for (and take advantage of, as we shall see in Chapter 2) any independencies that may hold, while not being limited to problems satisfying strong independence requirements. The combination of substantial increases in computer power with the Bayesian network's ability to use any existing independencies to computational advantage make the approximations and restrictive assumptions of earlier uncertainty formalisms pointless. So we now turn to the main game: understanding and representing uncertainty with probabilities.

## 1.3 Probability calculus

The probability calculus allows us to represent the independencies which other systems require, but also allows us to represent any dependencies which we may need.

---

[2]In particular, a mapping of certainty factors into likelihood ratios.
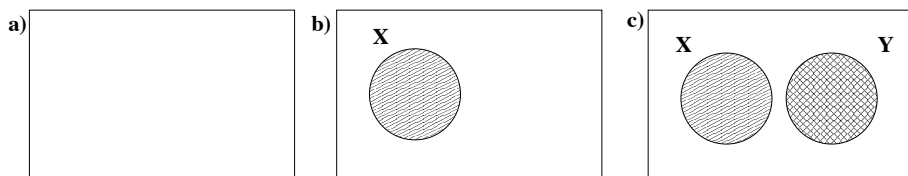
FIGURE 1.1: (a) The event space U; (b) $P(X)$; (c) $P(X \cup Y)$.

The probability calculus was specifically invented in the 17th century by Fermat and Pascal in order to deal with the problems of physical uncertainty introduced by gambling. But it did not take long before it was noticed that the concept of probability could be used in dealing also with the uncertainties introduced by ignorance, leading Bishop Butler to declare in the 18th century that "probability is the very guide to life." So now we introduce this formal language of probability, in a very simple way using Venn diagrams.

Let $U$ be the universe of possible events; that is, if we are uncertain about which of a number of possibilities is true, we shall let $U$ represent all of them collectively (see Figure 1.1(a)). Then the maximum probability must apply to the true event lying within $U$. By convention we set the maximum probability to 1, giving us Kolmogorov's first axiom for probability theory (Kolmogorov, 1933):

**Axiom 1.1** $P(U) = 1$

This probability mass, summing or integrating to 1, is distributed over $U$, perhaps evenly or perhaps unevenly. For simplicity we shall assume that it is spread evenly, so that the probability of any region is strictly proportional to its area. For any such region $X$ its area cannot be negative, even if $X$ is empty; hence we have the second axiom (Figure 1.1(b)):

**Axiom 1.2** *For all* $X \subseteq U, P(X) \geq 0$

We need to be able to compute the probability of combined events, $X$ and $Y$. This is trivial if the two events are mutually exclusive, giving us the third and last axiom (Figure 1.1(c)), known as **additivity**:

**Axiom 1.3** *For all* $X, Y \subseteq U$, *if* $X \cap Y = \emptyset$, *then* $P(X \cup Y) = P(X) + P(Y)$

Any function over a field of subsets of $U$ satisfying the above axioms will be a probability function.[3]

A simple theorem extends addition to events which overlap (i.e., sets which intersect):

**Theorem 1.1** *For all* $X, Y \subseteq U, P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$.

---

[3]A set-theoretic field is a set of sets containing $U$ and $\emptyset$ and is closed under union, intersection and complementation.
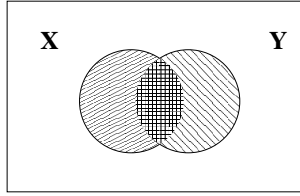
FIGURE 1.2: Conditional probability: $P(X|Y) = P(X \cap Y)/P(Y)$.

This can be intuitively grasped from Figure 1.2: the area of $X \cup Y$ is less than area of $X$ plus the area of $Y$ because when adding the area of intersection $X \cap Y$ has been counted twice; hence, we simply remove the excess to find $P(X \cup Y)$ for any two events $X$ and $Y$.

The concept of **conditional probability** is crucial for the useful application of the probability calculus. It is usually introduced by definition:

**Definition 1.1 Conditional probability**

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$

That is, given that the event $Y$ has occurred, or will occur, the probability that $X$ will also occur is $P(X|Y)$. Clearly, if $Y$ is an event with zero probability, then this conditional probability is undefined. This is not an issue for probability distributions which are **positive**, since, by definition, they are non-zero over every event. A simple way to think about probabilities conditional upon $Y$ is to imagine that the universe of events $U$ has shrunk to $Y$. The conditional probability of $X$ on $Y$ is just the measure of what is left of $X$ relative to what is left of $Y$; in Figure 1.2 this is just the ratio of the darker area (representing $X \cap Y$) to the area of $Y$. This way of understanding conditional probability is justified by the fact that the conditional function $P(\cdot|Y)$ is itself a probability function[4] — that is, it provably satisfies the three axioms of probability.

Another final probability concept we need to introduce is that of **independence** (or, marginal independence). Two events $X$ and $Y$ are probabilistically independent (in notation, $X \perp\!\!\!\perp Y$) whenever conditioning on one leaves the probability of the other unchanged:

**Definition 1.2 Independence** $X \perp\!\!\!\perp Y \equiv P(X|Y) = P(X)$

This is provably symmetrical: $X \perp\!\!\!\perp Y \equiv Y \perp\!\!\!\perp X$. The simplest examples of independence come from gambling. For example, two rolls of dice are normally independent. Getting a one with the first roll will neither raise nor lower the probability of getting a one the second time. If two events are **dependent**, then one coming true will *alter* the probability of the other. Thus, the probability of getting a diamond flush in poker

---

[4]$P(\cdot|Y)$ is just the function equal to $P(X|Y)$ for all $X \subset U$.

(five diamonds in five cards drawn) is *not* simply $(1/4)^5 = 1/1024$: the probability that the first card drawn being a diamond is $1/4$, but the probability of subsequent cards being diamonds is influenced by the fact that there are then fewer diamonds left in the deck.

**Conditional independence** generalizes this concept to $X$ and $Y$ being independent given some additional event $Z$:

**Definition 1.3  Conditional independence** $X \perp\!\!\!\perp Y | Z \equiv P(X|Y,Z) = P(X|Z)$

This is a true generalization because, of course, $Z$ can be the empty set $\emptyset$, when it reduces to marginal independence. Conditional independence holds when the event $Z$ tells us everything that $Y$ does about $X$ and possibly more; once you know $Z$, learning $Y$ is uninformative. For example, suppose we have two diagnostic tests for cancer $X$, an inexpensive but less accurate one, $Y$, and a more expensive and more accurate one, $Z$. If $Z$ is more accurate partly because it effectively incorporates all of the diagnostic information available from $Y$, then knowing the outcome of $Z$ will render an additional test of $Y$ irrelevant — $Y$ will be "screened off" from $X$ by $Z$.

### 1.3.1   Conditional probability theorems

We introduce without proof two theorems on conditional probability which will be of frequent use:

**Theorem 1.2  Total Probability** *Assume the set of events $\{A_i\}$ is a partition of $U$; i.e., $\bigcup_i A_i = U$ and for any distinct $i$ and $j$ $A_i \cap A_j = \emptyset$. Then*

$$P(U) = \sum_i P(A_i)$$

We can equally well partition the probability of any particular event $B$ instead of the whole event space. In other words, under the above conditions (and if $\forall i A_i \neq \emptyset$),

$$P(B) = \sum_i P(B \cap A_i)$$

We shall refer to either formulation under the title "Total Probability".

**Theorem 1.3  The Chain Rule** *Given three events $A, B, C$ in a chain of influence (i.e., A and C independent given B),*

$$P(C|A) = P(C|B)P(B|A) + P(C|\neg B)P(\neg B|A)$$

assuming the conditional probabilities are defined. This allows us to divide the probabilistic influence of $C$ on $A$ across the different states of a third variable. (Here, the third variable is binary, but the theorem is easily generalized to variables of arbitrary arity.)

### 1.3.2 Variables

Although we have introduced probabilities over events, in most of our discussion we shall be concerned with probabilities over **random variables**. A random variable is a variable which reports the outcome of some measurement process. It can be related to events, to be sure. For example, instead of talking about which event in a partition $\{A_i\}$ turns out to be the case, we can equivalently talk about which state $x_i$ the random variable $X$ takes, which we write $X = x_i$. The set of states a variable $X$ can take form its state space, written $\Omega_X$, and its size (or arity) is $|\Omega_X|$.

The discussion thus far has been implicitly of discrete variables, those with a finite state space. However, we need also to introduce the concept of probability distributions over **continuous variables**, that is, variables which range over real numbers, like *Temperature*. For the most part in this text we shall be using probability distributions over discrete variables (events), for two reasons. First, the Bayesian network technology is primarily oriented towards handling discrete state variables, for example the inference algorithms of Chapter 3. Second, for most purposes continuous variables can be **discretized**. For example, temperatures can be divided into ranges of $\pm 5$ degrees for many purposes; and if that is too crude, then they can be divided into ranges of $\pm 1$ degree, etc.

Despite our ability to evade probabilities over continuous variables much of the time, we shall occasionally need to discuss them. We introduce these probabilities by first starting with a **density function** $f(X)$ defined over the continuous variable $X$. Intuitively, the density assigns a weight or measure to each possible value of $X$ and can be approximated by a finely partitioned histogram reporting samples from $X$. Although the density is not itself a probability function, it can be used to generate one so long as $f(\cdot)$ satisfies the conditions:

$$f(x) \geq 0 \tag{1.1}$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \tag{1.2}$$

In words: each point value is positive or zero and all values integrate to 1. In that case we can define the **cumulative probability distribution** $F(\cdot)$ by

$$F(x) = P(X \leq x) = \int_{x' \leq x} f(x')dx' \tag{1.3}$$

This function assigns probabilities to ranges from each possible value of $x$ down to negative infinity. Note that we can analogously define probabilities over any continuous interval of values for $X$, so long as the interval is not degenerate (equal to a point). In effect, we obtain a probability distribution by discretizing the continuous variable — i.e., by looking at the mass of the density function over intervals.

## 1.4    Interpretations of probability

There have been two main contending views about how to understand probability. One asserts that probabilities are fundamentally dispositional properties of non-deterministic physical systems, the classical such systems being gambling devices, such as dice. This view is particularly associated with **frequentism**, advocated in the 19th century by John Venn (1866), identifying probabilities with long-run frequencies of events. The obvious complaint that short-run frequencies clearly do not match probabilities (e.g., if we toss a coin only once, we would hardly conclude that its probability of heads is either one or zero) does not actually get anywhere, since no claim is made identifying short-run frequencies with probabilities. A different complaint does bite, however, namely that the distinction between short-run and long-run is vague, leaving the commitments of this frequentist interpretation unclear. Richard von Mises in the early 20th century fixed this problem by formalizing the frequency interpretation (von Mises, 1919), identifying probabilities with frequency limits in infinite sequences satisfying certain assumptions about randomness. Some version of this frequency interpretation is commonly endorsed by statisticians.

A more satisfactory theoretical account of physical probability arises from Karl Popper's observation (1959) that the frequency interpretation, precise though it was, fails to accommodate our intuition that probabilities of singular events exist and are meaningful. If, in fact, we do toss a coin once and once only, and if this toss should *not* participate in some infinitude (or even large number) of appropriately similar tosses, it would not for that reason fail to have some probability of landing heads. Popper identified physical probabilities with the **propensities** (dispositions) of physical systems ("chance setups") to produce particular outcomes, whether or not those dispositions were manifested repeatedly. An alternative that amounts to much the same thing is to identify probabilities with counterfactual frequencies generated by hypothetically infinite repetitions of an experiment (van Fraassen, 1989).

Whether physical probability is relativized to infinite random sequences, infinite counterfactual sequences or chance setups, these accounts all have in common that the assertion of a probability is relativized to *some* definite physical process or the outcomes it generates.

The traditional alternative to the concept of physical probability is to think of probabilities as reporting our subjective degrees of belief. This view was expressed by Thomas Bayes (1958) (Figure 1.3) and Pierre Simon de Laplace (1951) two hundred years ago. This is a more general account of probability in that we have subjective belief in a huge variety of propositions, many of which are not at all clearly tied to a physical process capable even in principle of generating an infinite sequence of outcomes. For example, most of us have a pretty strong belief in the Copernican hypothesis that the earth orbits the sun, but this is based on evidence not obviously the same as the outcome of a sampling process. We are not in any position to generate solar systems repeatedly and observe the frequency with which their planets revolve around the sun, for example. Bayesians nevertheless are prepared to talk

about the probability of the truth of the Copernican thesis and can give an account of the relation between that probability and the evidence for and against it. Since these probabilities are typically subjective, not clearly tied to physical models, most frequentists (hence, most statisticians) deny their meaningfulness. It is not insignificant that this leaves their (usual) belief in Copernicanism unexplained.

The first thing to make clear about this dispute between physicalists and Bayesians is that Bayesianism can be viewed as *generalizing* physicalist accounts of probability. That is, it is perfectly compatible with the Bayesian view of probability as measuring degrees of subjective belief to adopt what David Lewis (1980) dubbed the **Principal Principle** whenever you learn that the physical probability of an outcome is $r$, set your subjective probability for that outcome to $r$. This is really just common sense: you may think that the probability of a friend shaving his head is 0.01, but if you learn that he will do so if and only if a fair coin yet to be flipped lands heads, you'll revise your opinion accordingly.

So, the Bayesian and physical interpretations of probability are compatible, with the Bayesian interpretation *extending* the application of probability beyond what is directly justifiable in physical terms. That is the view we adopt here. But what justifies this extension?



FIGURE 1.3: Reverend Thomas Bayes (1702–1761).

## 1.5   Bayesian philosophy

### 1.5.1   Bayes' theorem

The origin of Bayesian philosophy lies in an interpretation of **Bayes' Theorem:**

**Theorem 1.4  Bayes' Theorem**

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)}$$

This is a non-controversial (and simple) theorem of the probability calculus. Under its usual Bayesian interpretation, it asserts that the probability of a hypothesis $h$ conditioned upon some evidence $e$ is equal to its **likelihood** $P(e|h)$ times its probability prior to any evidence $P(h)$, **normalized** by dividing by $P(e)$ (so that the conditional probabilities of all hypotheses sum to 1). Proof is trivial.

The further claim that this is a right and proper way of adjusting our beliefs in our hypotheses given new evidence is called **conditionalization**, and it is controversial.

**Definition 1.4  Conditionalization** *After applying Bayes' theorem to obtain $P(h|e)$ adopt that as your posterior degree of belief in $h$ — or, $Bel(h) = P(h|e)$.*

Conditionalization, in other words, advocates belief updating via probabilities conditional upon the available evidence. It identifies **posterior probability** (the probability function after incorporating the evidence, which we are writing $Bel(\cdot)$) with **conditional probability** (the prior probability function conditional upon the evidence, which is $P(\cdot|e)$). Put thus, conditionalization may also seem non-controvertible. But there are certainly situations where conditionalization very clearly does not work. The two most basic such situations simply violate what are frequently explicitly stated as assumptions of conditionalization: (1) There must exist joint priors over the hypothesis and evidence spaces. Without a joint prior, Bayes' theorem cannot be used, so conditionalization is a non-starter. (2) The evidence conditioned upon, $e$, is all and only the evidence learned. This is called the **total evidence condition**. It is a significant restriction, since in many settings it cannot be guaranteed.

The first assumption is also significant. Many take it as the single biggest objection to Bayesianism to raise the question "Where do the numbers come from?" For example, the famous anti-Bayesian Clark Glymour (1980) doesn't complain about Bayesian reasoning involving gambling devices, when the outcomes are engineered to start out equiprobable, but doubts that numbers can be found for more interesting cases. To this kind of objection Bayesians react in a variety of ways. In fact, the different varieties of response pretty much identify the different schools of Bayesianism. Objectivists, such as Rudolf Carnap (1962) and Ed Jaynes (1968), attempt to define prior probabilities based upon the structure of language. Extreme subjectivists, such as de Finetti (1964), assert that it makes no difference what source your priors have:

given that de Finetti's representation theorem shows that non-extreme priors converge in the limit (under reasonable constraints), it just doesn't matter what priors you adopt.

The practical application of Bayesian reasoning does not appear to depend upon settling this kind of philosophical problem. A great deal of useful application can be done simply by refusing to adopt a dogmatic position and accepting common-sense prior probabilities. For example, if there are ten possible suspects in a murder mystery, a fair starting point for any one of them is a 1 in 10 chance of guilt; or, again, if burglaries occur in your neighborhood of 10,000 homes about once a day, then the probability of your having been burglarized within the last 24 hours might reasonably be given a prior probability of 1/10000.

Colin Howson points out that conditionalization is a valid rule of inference if and only if $Bel(e|h) = P(e|h)$, that is, if and only if your prior and posterior probability functions share the relevant conditional probabilities (cf. Howson, 2001). This is certainly a pertinent observation, since encountering some possible evidence may well inform us more about defects in our own conditional probability structure than about the hypothesis at issue. Since Bayes' theorem has $P(h|e)$ being proportional to $P(e|h)$, if the evidence leads us to revise $P(e|h)$, we will be in no position to conditionalize.

How to generate prior probabilities or new conditional probability structure is not dictated by Bayesian principles. Bayesian principles advise how to update probabilities once such a conditional probability structure has been adopted, given appropriate priors. Expecting Bayesian principles to answer all questions about reasoning is expecting too much. Nevertheless, we shall show that Bayesian principles implemented in computer programs can deliver a great deal more than the nay-sayers have ever delivered.

**Definition 1.5 Jeffrey conditionalization** *Suppose your observational evidence does not correspond specifically to proposition e, but can be represented as a posterior shift in belief about e. In other words, posterior belief in e is not full but partial, having shifted from $P(e)$ to $Bel(e)$. Then, instead of Bayesian conditionalization, apply Jeffrey's update rule for* **probability kinematics**: *$Bel(h) = P(h|e)Bel(e) + P(h|\neg e)Bel(\neg e)$ (Jeffrey, 1983).*

Jeffrey's own example is one where your hypothesis is about the color of a cloth, the evidence proposition *e* describes the precise quality of your visual experience under good light, but you are afforded a view of the cloth only under candlelight, in such a way that you cannot exactly articulate what you have observed. Nevertheless, you have learned *something*, and this is reflected in a shift in belief about the quality of your visual experience. Jeffrey conditionalization is very intuitive, but again is not strictly valid. As a practical matter, the need for such partial updating is common in Bayesian modeling.

## 1.5.2 Betting and odds

Odds are the ratio between the cost of a bet in favor of a proposition and the reward should the bet be won. Thus, assuming a stake of $1 (and otherwise simply rescaling the terms of the bet), a bet at 1:19 odds costs $1 and returns $20 should the proposition come true (with the reward being $20 minus the cost of the bet).[5] The odds may be set at any ratio and may, or may not, have something to do with one's probabilities. Bookies typically set odds for and against events at a slight discrepancy with their best estimate of the probabilities, for their profit lies in the difference between the odds for and against.

While odds and probabilities may deviate, probabilities and **fair odds** $O(\cdot)$ are strictly interchangeable concepts. The fair odds in favor of $h$ are defined simply as the ratio of the probability that $h$ is true to the probability that it is not:

**Definition 1.6 Fair odds**

$$O(h) = \frac{P(h)}{1 - P(h)}$$

Given this, it is an elementary matter of algebraic manipulation to find $P(h)$ in terms of odds:

$$P(h) = \frac{O(h)}{1 + O(h)} \tag{1.4}$$

Thus, if a coin is fair, the probability of heads is 1/2, so the odds in favor of heads are 1:1 (usually described as "50:50"). Or, if the odds of getting "snake eyes" (two 1's) on the roll of two dice are 1:35, then the probability of this is:

$$\frac{1/35}{1 + 1/35} = \frac{1/35}{36/35} = 1/36$$

as will always be the case with fair dice. Or, finally, suppose that the probability an agent ascribes to the Copernican hypothesis ($CH$) is zero; then the odds that agent is giving to Copernicus having been wrong ($\neg CH$) are *infinite*:

$$O(\neg CH) = \frac{1}{0} = \infty$$

At these odds, incidentally, it is trivial that the agent can never reach a degree of belief in $CH$ above zero on any finite amount of evidence, if relying upon conditionalization for updating belief.

With the concept of fair odds in hand, we can reformulate Bayes' theorem in terms of (fair) odds, which is often useful:

**Theorem 1.5 Odds-Likelihood Bayes' Theorem**

$$O(h|e) = \frac{P(e|h)}{P(e|\neg h)} O(h)$$

---

[5]It is common in sports betting to invert the odds, quoting the odds *against* a team winning, for example. This makes no difference; the ratio is simply reversed.

This is readily proven to be equivalent to Theorem 1.4. In English it asserts that the odds on $h$ conditional upon the evidence $e$ are equal to the prior odds on $h$ times the **likelihood ratio** $P(e|h) : P(e|\neg h)$. Clearly, the fair odds in favor of $h$ will rise if and only if the likelihood ratio is greater than one.

### 1.5.3   Expected utility

Generally, agents are able to assign utility (or, value) to the situations in which they find themselves. We know what we like, we know what we dislike, and we also know when we are experiencing neither of these. Given a general ability to order situations, and bets with definite probabilities of yielding particular situations, Frank Ramsey (1931) demonstrated that we can identify particular utilities with each possible situation, yielding a **utility function**.

If we have a utility function $U(O_i|A)$ over every possible outcome of a particular action $A$ we are contemplating, and if we have a probability for each such outcome $P(O_i|A)$, then we can compute the probability-weighted average utility for that action — otherwise known as the **expected utility** of the action:

**Definition 1.7  Expected utility**

$$EU(A) = \sum_i U(O_i|A) \times P(O_i|A)$$

It is commonly taken as axiomatic by Bayesians that agents ought to *maximize their expected utility*. That is, when contemplating a number of alternative actions, agents ought to decide to take that action which has the maximum expected utility. If you are contemplating eating strawberry ice cream or else eating chocolate ice cream, presumably you will choose that flavor which you prefer, other things being equal. Indeed, if you chose the flavor you liked *less*, we should be inclined to think that other things are *not* equal — for example, you are under some kind of external compulsion — or perhaps that you are not being honest about your preferences. Utilities have behavioral consequences *essentially*: any agent who consistently ignores the putative utility of an action or situation arguably does not have that utility.

Regardless of such foundational issues, we now have the conceptual tools necessary to understand what is fair about fair betting. **Fair bets** are fair because their expected utility is zero. Suppose we are contemplating taking the fair bet $B$ on proposition $h$ for which we assign probability $P(h)$. Then the expected utility of the bet is:

$$EU(B) = U(h|B)P(h|B) + U(\neg h|B)P(\neg h|B)$$

Typically, betting on a proposition has no effect on the probability that it is true (although this is not necessarily the case!), so $P(h|B) = P(h)$. Hence,

$$EU(B) = U(h|B)P(h) + U(\neg h|B)(1 - P(h))$$

Assuming a stake of 1 unit for simplicity, then by definition $U(h|B) = 1 - P(h)$ (i.e., this is the utility of $h$ being true given the bet for $h$) while $U(\neg h|B) = -P(h)$, so,

$$EU(B) = (1 - P(h))P(h) - P(h)(1 - P(h)) = 0$$

Given that the bet has zero expected utility, the agent should be no more inclined to take the bet in favor of *h* than to take the opposite bet against *h*.

### 1.5.4   Dutch books

The original Dutch book argument of Ramsey  (1931) (see also de Finetti, 1964) claims to show that subjective degrees of belief, if they are to be rational, *must* obey the probability calculus. It has the form of a *reductio ad absurdum* argument:

1. A rational agent should be willing to take either side of any combination of fair bets.
2. A rational agent should never be willing to take a combination of bets which guarantees a loss.
3. Suppose a rational agent's degrees of belief violate one or more of the axioms of probability.
4. Then it is provable that some combination of fair bets will lead to a guaranteed loss.
5. Therefore, the agent is both willing and not willing to take this combination of bets.

Now, the inferences to (4) and (5) in this argument are not in dispute (see §1.11 for a simple demonstration of (4) for one case). A *reductio* argument needs to be resolved by finding a prior assumption to blame, and concluding that it is false. Ramsey, and most Bayesians to date, supposed that the most plausible way of relieving the contradiction of (5) is by refusing to suppose that a rational agent's degrees of belief may violate the axioms of probability. This result can then be generalized beyond settings of explicit betting by taking "bets with nature" as a metaphor for decision-making generally. For example, walking across the street is in some sense a bet about our chances of reaching the other side.

Some anti-Bayesians have preferred to deny (1), insisting for example that it would be uneconomic to invest in bets with zero expected value (e.g., Chihara and Kennedy, 1979). But the ascription of the radical incoherence in (5) simply to the willingness of, say, bored aristocrats to place bets that will net them nothing clearly will not do: the effect of incoherence is entirely out of proportion with the proposed cause of effeteness.

Alan Hájek (2008)  has pointed out a more plausible objection to (2). In the scenarios presented in Dutch books there is always some combination of bets which guarantees a net loss whatever the outcomes on the individual bets. But equally there is always some combination of bets which guarantees a net gain — a "Good Book." So, one agent's half-empty glass is another's half-full glass! Rather than dismiss the Dutch-bookable agent as irrational, we might commend it for being open to a guaranteed win! So, Hájek's point seems to be that there is a fundamental symmetry in Dutch book arguments which leaves open the question whether violating probability axioms is rational or not. Certainly, when metaphorically extending betting to a "struggle" with Nature, it becomes rather implausible that She is really out to Dutch book us!

Hájek's own solution to the problem posed by his argument is to point out that whenever an agent violates the probability axioms there will be some variation of its system of beliefs which is guaranteed to win money whenever the original system is guaranteed to win, and which is also capable of winning in some situations when the original system is not. So the variant system of belief in some sense dominates the original: it is everywhere at least as good as the original and in some places better. In order to guarantee that your system of beliefs cannot be dominated, you must be probabilistically coherent (see §1.11). This, we believe, successfully rehabilitates the Dutch book in a new form.

Rather than rehabilitate, a more obviously Bayesian response is to consider the probability of a bookie hanging around who has the smarts to pump our agent of its money and, again, of a simpleton hanging around who will sign up the agent for guaranteed winnings. In other words, for rational choice surely what matters is the relative expected utility of the choice. Suppose, for example, that we are offered a set of bets which has a guaranteed loss of $10. Should we take it? The Dutch book assumes that accepting the bet is irrational. But, if the one and only alternative available is another bet with an expected loss of $1,000, then it no longer seems so irrational. An implicit assumption of the Dutch book has always been that betting is voluntary and when all offered bets are turned down the expected utility is zero. The further implicit assumption pointed out by Hájek's argument is that there is always a shifty bookie hanging around ready to take advantage of us. No doubt that is not always the case, and instead there is only some probability of it. Yet referring the whole matter of justifying the use of Bayesian probability to expected utility smacks of circularity, since expectation is understood in terms of Bayesian probability.

Aside from invoking the rehabilitated Dutch book, there is a more pragmatic approach to justifying Bayesianism, by looking at its importance for dealing with cases of practical problem solving. We take Bayesian principles to be normative, and especially to be a proper guide, under some range of circumstances, to evaluating hypotheses in the light of evidence. The form of justification that we think is ultimately most compelling is the "method of reflective equilibrium," generally attributed to Goodman (1973) and Rawls (1971), but adumbrated by Aristotle in his *Nicomachian Ethics*. In a nutshell, it asserts that the normative principles to accept are those which best accommodate our basic, unshakable intuitions about what is good and bad (e.g., paradigmatic judgments of correct inference in simple domains, such as gambling) and which best integrate with relevant theory and practice. We now present some cases which Bayesian principle handles readily, and better than any alternative normative theory.

### 1.5.5 Bayesian reasoning examples

#### 1.5.5.1 Breast cancer

*Suppose the women attending a particular clinic show a long-term chance of 1 in 100 of having breast cancer. Suppose also that the initial screening test used at the clinic has a false positive rate of 0.2 (that is, 20% of women without cancer will test*

*positive for cancer) and that it has a false negative rate of 0.1 (that is, 10% of women with cancer will test negative). The laws of probability dictate from this last fact that the probability of a positive test given cancer is 90%. Now suppose that you are such a woman who has just tested positive. What is the probability that you have cancer?*

This problem is one of a class of probability problems which has become notorious in the cognitive psychology literature (cf. Tversky and Kahneman, 1974). It seems that very few people confronted with such problems bother to pull out pen and paper and compute the right answer via Bayes' theorem; even fewer can get the right answer without pen and paper. It appears that for many the probability of a positive test (which is observed) given cancer (i.e., 90%) dominates things, so they figure that they have quite a high chance of having cancer. But substituting into Theorem 1.4 gives us:

$$P(Cancer|Pos) = \frac{P(Pos|Cancer)P(Cancer)}{P(Pos)}$$

Note that the probability of *Pos* given *Cancer* — which is the likelihood 0.9 — is only *one* term on the right hand side; the other crucial term is the prior probability of cancer. Cognitive psychologists studying such reasoning have dubbed the dominance of likelihoods in such scenarios "base-rate neglect," since the base rate (prior probability) is being suppressed (Kahneman and Tversky, 1973). Filling in the formula and computing the conditional probability of *Cancer* given *Pos* gives us quite a different story:

$$
\begin{aligned}
P(Cancer|Pos) &= \frac{P(Pos|Cancer)P(Cancer)}{P(Pos)} \\
&= \frac{P(Pos|Cancer)P(Cancer)}{P(Pos|Cancer)P(Cancer) + P(Pos|\neg Cancer)P(\neg Cancer)} \\
&= \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.2 \times 0.99} \\
&= \frac{0.009}{0.009 + 0.198} \\
&\approx 0.043
\end{aligned}
$$

Now the discrepancy between 4% and 80 or 90% is no small matter, particularly if the consequence of an error involves either unnecessary surgery or (in the reverse case) leaving a cancer untreated. But decisions similar to these are constantly being made based upon "intuitive feel" — i.e., without the benefit of paper and pen, let alone Bayesian networks (which are simpler to use than paper and pen!).

### 1.5.5.2  People v. Collins

The legal system is replete with misapplications of probability and with incorrect claims of the irrelevance of probabilistic reasoning as well.

In 1964 an interracial couple was convicted of robbery in Los Angeles, largely on the grounds that they matched a highly improbable profile, a profile which fit witness reports (Sullivan, Sullivan). In particular, the two robbers were reported to be

- A man with a mustache

- Who was black and had a beard
- And a woman with a ponytail
- Who was blonde
- The couple was interracial
- And were driving a yellow car

The prosecution suggested that these characteristics had the following probabilities of being observed at random in the LA area:

1. A man with a mustache 1/4
2. Who was black and had a beard 1/10
3. And a woman with a ponytail 1/10
4. Who was blonde 1/3
5. The couple was interracial 1/1000
6. And were driving a yellow car 1/10

The prosecution called an instructor of mathematics from a state university who apparently testified that the "product rule" applies to this case: where mutually independent events are being considered jointly, the joint probability is the product of the individual probabilities.[6] This last claim is, in fact, correct (see Problem 2 below); what is false is the idea that the product rule is relevant to this case. If we label the individual items of evidence $e_i$ $(i = 1, \ldots, 6)$, the joint evidence $e$, and the hypothesis that the couple was guilty $h$, then what is claimed is

$$P(e|\neg h) = \prod_i P(e_i|\neg h) = 1/12000000$$

The prosecution, having made this inference, went on to assert that the probability the couple were innocent was no more than 1/12000000. The jury convicted.

As we have already suggested, the product rule does *not* apply in this case. Why not? Well, because the individual pieces of evidence are obviously *not* independent. If, for example, we know of the occupants of a car that one is black and the other has blonde hair, what then is the probability that the occupants are an interracial couple? Clearly not 1/1000! If we know of a man that he has a mustache, is the probability of having a beard unchanged? These claims are preposterous, and it is simply shameful that a judge, prosecutor and defence attorney could not recognize how preposterous they are — let alone the mathematics "expert" who testified to them. Since $e_2$ implies $e_1$, while $e_2, e_3, e_4$ jointly imply $e_5$ (to a fair approximation), a far better estimate for $P(e|\neg h)$ is $P(e_2|\neg h)P(e_3|\neg h)P(e_4|\neg h)P(e_6|\neg h) = 1/3000$.

To be sure, if we accepted that the probability of innocence were a mere 1/3000 we might well accept the verdict. But there is a more fundamental error in the prosecution reasoning than neglecting the conditional dependencies in the evidence. If,

---

[6]Coincidentally, this is just the kind of independence required for certainty factors to apply.

unlike the judge, prosecution and jury, we take a peek at Bayes' theorem, we discover that the probability of guilt $P(h|e)$ is *not* equal to $1 - P(e|\neg h)$; instead

$$P(h|e) = \frac{P(e|h)P(h)}{P(e|h)P(h) + P(e|\neg h)P(\neg h)}$$

Now if the couple in question *were* guilty, what are the chances the evidence accumulated would have been observed? That's a rather hard question to answer, but feeling generous towards the prosecution, let us simplify and say 1. That is, let us accept that $P(e|h) = 1$. Plugging in our assumptions we have thus far:

$$P(h|e) = \frac{P(h)}{P(h) + P(\neg h)/3000}$$

We are missing the crucial prior probability of a random couple being guilty of the robbery. Note that we cannot here use the prior probability of, for example, an interracial couple being guilty, since the fact that they are interracial is a piece of the evidence. The most plausible approach to generating a prior of the needed type is to count the number of couples in the LA area and give them an equal prior probability. In other words, if N is the number of possible couples in the LA area, $P(h) = 1/N$. So, what is $N$? The population at the time was about 6.5 million people (Demographia, Demographia). If we conservatively take half of them as being eligible to be counted (e.g., being adult humans), this gives us 1,625,000 eligible males and as many females. If we simplify by supposing that they are all in heterosexual partnerships, that will introduce a slight bias in favor of innocence; if we also simplify by ignoring the possibility of people traveling in cars with friends, this will introduce a larger bias in favor of guilt. The two together give us 1,625,000 available couples, suggesting a prior probability of guilt of 1/1625000. Plugging this in we get:

$$P(h|e) = \frac{1/1625000}{1/1625000 + (1 - 1/1625000)/3000} \approx 0.002$$

In other words, even ignoring the huge number of trips with friends rather than partners, we obtain a 99.8% chance of innocence and so a very large probability of a nasty error in judgment. The good news is that the conviction (of the man only!) was subsequently overturned, partly on the basis that the independence assumptions are false. The bad news is that the appellate court finding also suggested that probabilistic reasoning is just irrelevant to the task of establishing guilt, which is a nonsense. One right conclusion about this case is that, assuming the likelihood has been *properly* worked out, a sensible prior probability must also be taken into account. In some cases judges have specifically ruled out all consideration of prior probabilities, while allowing testimony about likelihoods! Probabilistic reasoning which simply ignores half of Bayes' theorem is dangerous indeed!

Note that we do not claim that 99.8% is the best probability of innocence that can be arrived at for the case of People v. Collins. What we *do* claim is that, for the particular facts represented as having a particular probabilistic interpretation, this is far closer to a reasonable probability than that offered by the prosecution, namely

1/12000000. We also claim that the forms of reasoning we have here illustrated are *crucial* for interpreting evidence in general: namely, whether the offered items of evidence are conditionally independent and what the prior probability of guilt may be.

## 1.6  The goal of Bayesian AI

The most commonly stated goal for artificial intelligence is that of producing an artifact which performs difficult intellectual tasks at or beyond a human level of performance. Of course, machine chess programs have satisfied this criterion for some time now. Although some AI researchers have claimed that therefore an AI has been produced — that denying this is an unfair shifting of the goal line — it is absurd to think that we ought to be satisfied with programs which are strictly special-purpose and which achieve their performance using techniques that deliver nothing when applied to most areas of human intellectual endeavor.

Turing's test for intelligence appears to be closer to satisfactory: fooling ordinary humans with verbal behavior not restricted to any domain would surely demonstrate some important *general* reasoning ability. Many have pointed out that the conditions for Turing's test, strictly verbal behavior without any afferent or efferent nervous activity, yield at best some kind of disembodied, ungrounded intelligence. John Searle's Chinese Room argument (Searle, 1980) for example, can be interpreted as making such a case; for this kind of interpretation of Searle see Harnad (1989) and Korb (1991). A more convincing criterion for human-like intelligence is to require of an artificial intelligence that it be capable of powering a robot-in-the-world in such a way that the robot's performance cannot be distinguished from human performance in terms of behavior (disregarding, for example, whether the skin can be so distinguished). The program that can achieve this would surely satisfy any sensible AI researcher, or critic, that an AI had been achieved.

We are not, however, actually motivated by the idea of behaviorally cloning humans. If all we wish to do is reproduce humans, we would be better advised to employ the tried and true methods we have always had available. Our motive is to understand *how* such performance can be achieved. We are interested in knowing how humans perform the many interesting and difficult cognitive tasks encompassed by AI — such as, natural language understanding and generation, planning, learning, decision making — but we are also interested in knowing how they might be performed otherwise, and in knowing how they might be performed optimally. By building artifacts which model our best understanding of how humans do these things (which can be called **descriptive artificial intelligence**) and also building artifacts which model our best understanding of what is optimal in these activities (**normative artificial intelligence**), we can further our understanding of the nature of intelligence and also produce some very useful tools for science, government and industry.

As we have indicated through example, medical, legal, scientific, political and

most other varieties of human reasoning either consider the relevant probabilistic factors and accommodate them or run the risk of introducing egregious and damaging errors. The goal of a Bayesian artificial intelligence is to produce a thinking agent which does as well or better than humans in such tasks, which can adapt to stochastic and changing environments, recognize its own limited knowledge and cope sensibly with these varied sources of uncertainty.

## 1.7    Achieving Bayesian AI

Given that we have this goal, how can we achieve it? The first step is to develop algorithms for doing Bayesian conditionalization properly and, insofar as possible, efficiently. This step has already been achieved, and the relevant algorithms are described in Chapters 2 and 3. The next step is to incorporate methods for computing expected utilities and develop methods for maximizing utility in decision making. We describe algorithms for this in Chapter 4. We would like to test these ideas in application: we describe some Bayesian network applications in Chapter 5.

These methods for probability computation are fairly well developed and their improvement remains an active area of research in AI today. The biggest obstacles to Bayesian AI having a broad and deep impact outside of the research community are the difficulties in developing applications, difficulties with eliciting knowledge from experts, and integrating and validating the results. One issue is that there is no clear methodology for developing, testing and deploying Bayesian network technology in industry and government — there is no recognized discipline of "software engineering" for Bayesian networks. We make a preliminary effort at describing one — Knowledge Engineering with Bayesian Networks (KEBN) in Part III, including its illustration in case studies of Bayesian network development in Chapter 11.

Another important response to the difficulty of building Bayesian networks by hand is the development of methods for their automated learning — the machine learning of Bayesian networks (aka "data mining"). In Part II we introduce and develop the main methods for learning Bayesian networks with reference to the theory of causality underlying them. These techniques logically come before the knowledge engineering methodology, since that draws upon and integrates machine learning with expert elicitation.

## 1.8    Are Bayesian networks Bayesian?

Many AI researchers like to point out that Bayesian networks are not inherently Bayesian at all; some have even claimed that the label is a misnomer. At the 2002 Australasian Data Mining Workshop, for example, Geoff Webb made the former

claim. Under questioning it turned out he had two points in mind: (1) Bayesian networks are frequently "data mined" (i.e., learned by some computer program) via non-Bayesian methods. (2) Bayesian networks at bottom represent probabilities; but probabilities can be interpreted in any number of ways, including as some form of frequency; hence, the networks are not intrinsically either Bayesian or non-Bayesian, they simply represent values needing further interpretation.

These two points are entirely correct. We shall ourselves present non-Bayesian methods for automating the learning of Bayesian networks from statistical data. We shall also present Bayesian methods for the same, together with some evidence of their superiority. The interpretation of the probabilities represented by Bayesian networks is open so long as the philosophy of probability is considered an open question. Indeed, much of the work presented here ultimately depends upon the probabilities being understood as *physical probabilities*, and in particular as propensities or probabilities determined by propensities. Nevertheless, we happily invoke the Principal Principle: where we are convinced that the probabilities at issue reflect the true propensities in a physical system we are certainly going to use them in assessing our own degrees of belief.

The advantages of the Bayesian network representations are largely in simplifying conditionalization, planning decisions under uncertainty and explaining the outcome of stochastic processes. These purposes all come within the purview of a clearly Bayesian interpretation of what the probabilities mean, and so, we claim, the Bayesian network technology which we here introduce is aptly named: it provides the technical foundation for a truly Bayesian artificial intelligence.

## 1.9   Summary

How best to reason about uncertain situations has always been of concern. From the 17th century we have had available the basic formalism of probability calculus, which is far and away the most promising formalism for coping with uncertainty. Probability theory has been used widely, but not deeply, since then. That is, the elementary ideas have been applied to a great variety of problems — e.g., actuarial calculations for life insurance, coping with noise in measurement, business decision making, testing scientific theories, gambling — but the problems have typically been of highly constrained size, because of the computational infeasibility of conditionalization when dealing with large problems. Even in dealing with simplified problems, humans have had difficulty handling the probability computations. The development of Bayesian network technology automates the process and so promises to free us from such difficulties. At the same time, improvements in computer capacity, together with the ability of Bayesian networks to take computational advantage of any available independencies between variables, promise to both widen and deepen the domain of probabilistic reasoning.

## 1.10  Bibliographic notes

An excellent source of information about different attempts to formalize reasoning about uncertainty — including certainty factors, non-monotonic logics, Dempster-Shafer calculus, as well as probability — is the anthology *Readings in Uncertain Reasoning* edited by Shafer and Pearl (1990). Three polemics against non-Bayesian approaches to uncertainty are those by Drew McDermott (1987), Peter Cheeseman (1988) and Kevin Korb (1995). For understanding Bayesian philosophy, Ramsey's original paper "Truth and Probability" is beautifully written, original and compelling (1931); for a more comprehensive and recent presentation of Bayesianism see Howson and Urbach's *Scientific Reasoning* (2007). For Bayesian decision analysis see Richard Jeffrey's *The Logic of Decision* (1983). DeGroot and Schervish (2002) provide an accessible introduction to both the probability calculus and statistics.

Karl Popper's original presentation of the propensity interpretation of probability is (Popper, 1959). This view is related to the elaboration of a probabilistic account of causality in recent decades. Wesley Salmon (1984) provides an overview of probabilistic causality.

Naive Bayes models, despite their simplicity, have done surprisingly well as predictive classifiers for data mining problems; see Chapter 7.

## 1.11  Technical notes

### A Dutch book

Here is a simple Dutch book. Suppose someone assigns $P(A) = -0.1$, violating probability Axiom 2. Then $O(A) = -0.1/(1 - (-0.1)) = -0.1/1.1$. The reward for a bet on $A$ with a \$1 stake is $\$(1 - P(U)) = \$1.1$ if $A$ comes true and $\$ - P(U) = \$0.1$ if $A$ is false. That's everywhere positive and so is a "Good Book." The Dutch book simply requires this agent to take the fair bet *against $A$*, which has the payoffs $-\$1.1$ if $A$ is true and $-\$0.1$ otherwise.

### The rehabilitated Dutch book

Following Hájek, we can show that incoherence (violating the probability axioms) leads to being "dominated" by someone who is coherent — that is, the coherent bettor can take advantage of offered bets that the incoherent bettor cannot and otherwise will do as well.

Suppose Ms. Incoherent assigns $P_I(U) < 1$ (where $U$ is the universal event that *must* occur), for example. Then Ms. Incoherent will take any bet for $U$ at odds of $P_I(U)/(1 - P_I(U))$ or greater. But Ms. Coherent has assigned $P_C(U) = 1$, of course, and so can take any bet for $U$ at any odds offered greater than zero. So for the odds within the range $[0, \frac{P_I(U)}{1 - P_I(U)}]$ Ms. Coherent is guaranteed a profit whereas Ms. Inco-

herent is sitting on her hands.

### NP hardness

A problem is Non-deterministic Polynomial-time (NP) if it is solvable in polynomial time on a non-deterministic Turing machine. A problem is Non-deterministic Polynomial time hard (NP hard) if every problem that is NP can be translated into this NP hard problem in polynomial time. If there is a polynomial time solution to any NP hard problem, then because of polynomial time translatability for all other NP problems, there must be a polynomial time solution to all NP problems. No one knows of a polynomial time solution to any NP hard problem; the best known solutions are exponentially explosive. Thus, "NP hard" problems are generally regarded as computationally intractable. (The classic introduction to computational complexity is Garey and Johnson (1979).)

## 1.12   Problems

### Probability Theory

### Problem 1

Prove that the conditional probability function $P(\cdot|e)$, if well defined, is a probability function (i.e., satisfies the three axioms of Kolmogorov).

### Problem 2

Given that two pieces of evidence $e_1$ and $e_2$ are conditionally independent given the hypothesis — i.e., $P(e_1|e_2,h) = P(e_1|h)$ — prove the "product rule": $P(e_1,e_2|h) = P(e_1|h) \times P(e_2|h)$.

### Problem 3

Prove the theorems of §1.3.1, namely the Total Probability theorem and the Chain Rule.

### Problem 4

There are five containers of milk on a shelf; unbeknownst to you, two of them have passed their use-by date. You grab two at random. What's the probability that neither have passed their use-by date? Suppose someone else has got in just ahead of you, taking one container, after examining the dates. What's the probability that the two you take at random after that are ahead of their use-by dates?

## Problem 5

The probability of a child being a boy (or a girl) is 0.5 (let us suppose). Consider all the families with exactly two children. What is the probability that such a family has two girls given that it has at least one girl?

## Problem 6

The frequency of male births at the Royal Women's Hospital is about 51 in 100. On a particular day, the last eight births have been female. The probability that the next birth will be male is:

1. About 51%
2. Clearly greater than 51%
3. Clearly less than 51%
4. Almost certain
5. Nearly zero

## Bayes' Theorem

### Problem 7

After winning a race, an Olympic runner is tested for the presence of steroids. The test comes up positive, and the athlete is accused of doping. Suppose it is known that 5% of all victorious Olympic runners do use performance-enhancing drugs. For this particular test, the probability of a positive finding given that drugs are used is 95%. The probability of a false positive is 2%. What is the (posterior) probability that the athlete did in fact use steroids, given the positive outcome of the test?

### Problem 8

You consider the probability that a coin is double-headed to be 0.01 (call this option $h'$); if it isn't double-headed, then it's a fair coin (call this option $h$). For whatever reason, you can only test the coin by flipping it and examining the coin (i.e., you can't simply examine both sides of the coin). In the worst case, how many tosses do you need before having a posterior probability for either $h$ or $h'$ that is greater than 0.99, i.e., what's the maximum number of tosses until that happens?

### Problem 9

(Adapted from Fischoff and Bar-Hillel (1984).) Two cab companies, the Blue and the Green, operate in a given city. Eighty-five percent of the cabs in the city are Blue; the remaining 15% are Green. A cab was involved in a hit-and-run accident at night. A witness identified the cab as a Green cab. The court tested the witness' ability to distinguish between Blue and Green cabs under night-time visibility conditions. It found that the witness was able to identify each color correctly about 80% of the time, but confused it with the other color about 20% of the time.

What are the chances that the errant cab was indeed Green, as the witness claimed?

## Odds and Expected Value

### Problem 10

Construct a Dutch book against someone who violates the Axiom of Additivity. That is, suppose a Mr. Fuzzy declares about the weather tomorrow that $P(Sunny) = 0.5, P(Inclement) = 0.5,$ and $P(Sunny \text{ or } inclement) = 0.5$. Mr. Fuzzy and you agree about what will count as sunny and as inclement weather and you both agree that they are incompatible states. How can you construct a Dutch book against Fuzzy, using only fair bets?

### Problem 11

A bookie offers you a ticket for \$5.00 which pays \$6.00 if Manchester United beats Arsenal and nothing otherwise. What are the odds being offered? To what probability of Manchester United winning does that correspond?

### Problem 12

You are offered a Keno ticket in a casino which will pay you \$1 million if you win! It only costs you \$1 to buy the ticket. You choose 4 numbers out of a 9x9 grid of distinct numbers. You win if all of your 4 numbers come up in a random draw of four from the 81 numbers. What is the expected dollar value of this gamble?

## Applications

### Problem 13

(Note: this is the case of Sally Clark, convicted in the UK in 1999, found innocent on appeal in 2003, and tragically died in 2007 of alcohol poisoning. See Innocent, 2002.) A mother was arrested after her second baby died a few months old, apparently of sudden infant death syndrome (SIDS), exactly as her first child had died a year earlier. According to prosecution testimony, about 2 in 17200 babies die of SIDS. So, according to their argument, there is only a probability of $(2/17200)^2 \approx 1/72000000$ that two such deaths would happen in the same family by chance alone. In other words, according to the prosecution, the woman was guilty beyond a reasonable doubt. The jury returned a guilty verdict, even though there was no significant evidence of guilt presented beyond this argument. Which of the following is the truth of the matter? Why?

1. Given the facts presented, the probability that the woman is guilty is greater than 99%, so the jury decided correctly.
2. The argument presented by the prosecution is irrelevant to the mother's guilt or innocence.
3. The prosecution argument is relevant but inconclusive.

4. The prosecution argument only establishes a probability of guilt of about 16%.

5. Given the facts presented, guilt and innocence are equally likely.

## Problem 14

A DNA match between the defendant and a crime scene blood sample has a probability of 1/100000 if the defendant is innocent. There is no other significant evidence.

1. What is the probability of guilt?

2. Suppose we agree that the prior probability of guilt under the (unspecified) circumstances is 10%. What then is the probability of guilt?

3. The suspect has been picked up through a universal screening program applied to all Australians seeking a Medicare card. So far, 10 million people have been screened. What then is the probability of guilt?

# *Bibliography*

Abdelbar, A. M. and S. M. Hedetniemi (1998). The complexity of approximating MAP explanation. *Artificial Intelligence 102*, 21–38.

Abramson, B., J. Brown, W. Edwards, A. Murphy, and R.B. Winkler (1996). Hailfinder: A Bayesian system for forecasting severe weather. *International Journal of Forecasting* 20(1), 57–71. http://www.lis.pitt.edu/~dsl/hailfinder/.

Abramson, B. and A. Finizza (1991, November). Using belief networks to forecast oil prices. *International Journal of Forecasting 7*(3), 299–315.

Acid, S., L. de Campos, J. Fernández-Luna, S. Rodríguez, J. María Rodríguez, and J. Luis Salcedo (2004). A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. *Artificial Intelligence in Medicine 30*(3), 215–232.

Adusei-Poku, K. (2005). *Operational risk management: Implementing a Bayesian network for foreign exchange and money market settlement*. Ph. D. thesis, University of Gottingen, German.

Aitkenhead, M. and I. Aalders (2009). Predicting land cover using GIS, Bayesian and evolutionary algorithm methods. *Journal of Environmental Management 90*(1), 236–250.

Albrecht, D. W., I. Zukerman, and A. E. Nicholson (1998). Bayesian models for keyhole plan recognition in an adventure game. *User Modeling and User-Adapted Interaction 8*(1-2), 5–47.

Aliferis, C., A. Statnikov, I. Tsamardinos, S. Mani, and X. Koutsoukos (2010a). Local causal and Markov blanket induction for causal discovery and feature selection for classification Part I: Algorithms and empirical evaluation. *Journal of Machine Learning Research 11*, 171–234.

Aliferis, C., A. Statnikov, I. Tsamardinos, S. Mani, and X. Koutsoukos (2010b). Local causal and Markov blanket induction for causal discovery and feature selection for classification Part II: Analysis and extensions. *Journal of Machine Learning Research 11*, 235–284.

Allen, D. and A. Darwiche (2008). Rc_link: Genetic linkage analysis using Bayesian networks. *International Journal of Approximate Reasoning 48*(2), 499–525.

Andersen, S., K. Olesen, F. Jensen, and F. Jensen (1989). HUGIN — a shell for building Bayesian belief universes for expert systems. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI)*, Detroit, MI.

Anderson, J. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Anderson, K. M., P. M. Odell, P. W. Wilson, and W. B. Kannel (1991). Cardiovascular disease risk profiles. *American Heart Journal 121*, 293–298.

Anderson, S., D. Madigan, and M. Perlman (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Annals of Statistics 25*, 505–541.

Andreassen, S., J. Benn, R. Hovorks, K. Olesen, and R. Carson (1994). A probabilistic approach to glucose prediction and insulin dose adjustment. *Computer Methods and Programs in Biomedicine 41*, 153–165.

Andreassen, S., F. V. Jensen, S. K. Andersen, B. Falck, U. Kjærulff, M. Woldbye, A. R. Sørensen, A. Rosenfalck, and F. Jensen (1989). MUNIN — an expert EMG assistant. In J. Desmedt (Ed.), *Computer-Aided Electromyography and Expert Systems*. Elsevier, Amsterdam.

Aristotle. *The Nicomachean Ethics*.

Asher, H. B. (1983). *Causal Modeling* (Second ed.). Beverly Hills, CA: Sage.

Assmann, G., P. Cullen, and H. Schulte (2002). Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the Prospective Cardiovascular Münster (PROCAM) Study. *Circulation 105*(3), 310–315. http://circ.ahajournals.org/cgi/content/abstract/105/3/310.

Asuncion, A. and D. Newman (2007). UCI machine learning repository. http://www.ics.uci.edu/∼mlearn/MLRepository.html.

Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press, London.

Bacon, P., J. Cain, and D. Howard (2002). Belief network models of land manager decisions and land use change. *Journal of Environmental Management 65*(1), 1–23.

Baesens, B., T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society 54*, 627–635.

Balakrishnan, N. and V. Nevzorov (2003). *A Primer on Statistical Distributions*. New York: Wiley.

Banerjee, J., H.-T. Chou, J. Garza, W. Kim, D. woelk, N. Ballou, and H.-J. Kim (1987). Data model issues for object-oriented applications. *ACM Transactions on Office Informations Systems 5*(1), 3–27.

Bangso, O. and P. Wuillemin (2000). Top-down construction and repetitive structures representation in Bayesian networks. In *Proceedings of the Thirteenth International Florida Artificial Intelligence Research Society Conference*, pp. 282–286.

Barton, D., T. Saloranta, S. Moe, H. Eggestad, and S. Kuikka (2008). Bayesian belief networks as a meta-modeling tool in integrated river basin management–Pros and cons in evaluating nutrient abatement decisions under uncertainty in a Norwegian river basin. *Ecological Economics 66*(1), 91–104.

Bayes, T. (1764/1958). An essay towards solving a problem in the doctrine of chances. *Biometrika 45*, 296–315.

Beinlich, I., H. Suermondt, R. Chavez, and G. Cooper (1992). The ALARM monitoring system. In *Proceedings of the Second European Conference on Artificial Intelligence in Medicine*, pp. 689–693.

Bell, D. (1988). Disappointment in decision making under uncertainty. In H. R. D.E. Bell and A. Tversky (Eds.), *Decision making: Descriptive, Normative, and Prescriptive Interactions*, pp. 358–384. Cambridge Press, London.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematics Society 35*, 99–110.

Biedermann, A., F. Taroni, O. Delemont, C. Semadeni, and A. Davison (2005a). The evaluation of evidence in the forensic investigation of fire incidents. Part I: an approach using Bayesian networks. *Forensic Science International 147*(1), 49–57.

Biedermann, A., F. Taroni, O. Delemont, C. Semadeni, and A. Davison (2005b). The evaluation of evidence in the forensic investigation of fire incidents. Part II: Practical examples of the use of Bayesian networks. *Forensic Science International 147*(1), 59–69.

Billings, D., A. Davidson, J. Schaeffer, and D. Szafron (2002). The challenge of poker. *Artificial Intelligence 134*, 201–240.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer Verlag.

Blalock, H. (1964). *Causal Inferences in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.

Boehm, B. W. (1988). A spiral model of software development and enhancement. *IEEE Computer*, 61–72.

Boerlage, B. (1995). Link strengths in Bayesian Networks. Master's thesis, Department of Computer Science, University of British Columbia.

Bonafede, C. and P. Giudici (2007). Bayesian networks for enterprise risk assessment. *Physica A: Statistical Mechanics and its Applications 382*(1), 22–28.

Boneh, T. (2010). *Ontology and Bayesian decision networks for supporting the meteorological forecasting process*. Ph. D. thesis, Clayton School of Information Technology, Monash University.

Boneh, T., A. Nicholson, and L. Sonenberg (2006). Matilda: A visual tool for modeling with Bayesian networks. *International Journal of Intelligent Systems 21*(11), 1127–1150.

Borgelt, C. and R. Kruse (2002). *Graphical models: Methods for Data Analysis and Mining*. New York: Wiley.

Borsuk, M. (2008). Bayesian networks. In S. Jorgensen and B. Fath (Eds.), *Encyclopedia of Ecology*, pp. 307–317. Oxford, UK: Elsevier.

Borsuk, M., P. Reichert, and P. Burkhardt-Holm (2002). A Bayesian network for investigating the decline in fish catch in switzerland. In *Integrated Assessment and Decision Support. Proceedings of the 1st biennial meeting of the International Environmental Modeling and Software Society, Lugano, Switzerland*, pp. 108–113.

Borsuk, M., C. Stow, and K. Reckhow (2004). A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modeling 173*(2-3), 219–239.

Bouckaert, R. (1994). Probabilistic network construction using the minimum description length principle. Technical Report RUU-CS-94-27, Dept of Computer Science, Utrecht University.

Boudon, R. (1968). A new look at correlation analysis. In H. Blalock and A. Blalock (Eds.), *Methodology in Social Research*, pp. 199–235. McGraw-Hill, New York.

Boulton, D. (2003). Bayesian opponent modeling in poker. Honours thesis, School of Computer Science and Software Engineering, Monash University.

Boulton, D. and C. Wallace (1969). The information content of a multistate distribution. *Journal of Theoretical Biology 23*, 269–278.

Boutilier, C., N. Friedman, M. Goldszmidt, and D. Koller (1996). Context-specific independence in Bayesian networks. In *UAI96 – Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pp. 115–123.

Boyen, X. and D. Koller (1998). Tractable inference for complex stochastic processes. In *UAI98 – Proceedings of the Fourteenth Conference on Uncertainty in Uncertainty in Artificial Intelligence*, pp. 33–42.

BPP (2006-2010). Online version. http://www.csse.monash.edu.au/bai/poker/.

Bradford, J., C. Needham, A. Bulpitt, and D. Westhead (2006). Insights into protein-protein interfaces using a Bayesian network prediction method. *Journal of Molecular Biology 362*(2), 365–386.

Brightwell, G. and P. Winkler (1990). Counting linear extensions is #p-complete. Technical Report DIMACS 90-49, Dept of Computer Science, Rutgers.

Bromley, J., N. Jackson, O. Clymer, A. Giacomello, and F. Jensen (2005). The use of Hugin® to develop Bayesian networks as an aid to integrated water resource planning. *Environmental Modeling & Software 20*(2), 231–242.

Brooks, F. (1995). *The Mythical Man-Month: Essays on Software Engineering* (Second ed.). Reading, MA: Addison-Wesley.

Buchanan, B. and E. Shortliffe (Eds.) (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley.

Buntine, W. (1991). Theory refinement on Bayesian networks. In *UAI91 – Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pp. 52–69.

Buntine, W. (1996). A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on Knowledge and Data Engineering 8*, 195–210.

Burnside, E., D. Rubin, and R. Shachter (2004). Improving a Bayesian network's ability to predict the probability of malignancy of microcalcifications on mammography. In *International Congress Series*, Volume 1268, pp. 1021–026. Elsevier.

Cadwallader, P. L. (1978). Some causes of the decline in range and abundance of native fish in the Murray-Darling River system. In *The Murray-Darling River System Symposium, Proceedings of the Royal Society of Victoria*.

Cain, J. (2001). Planning improvements in natural resources management: Guidelines for using Bayesian networks to support the planning and management of development programmes in the water sector and beyond. Technical report, Centre for Ecology and Hydrology, Crowmarsh Gifford, Wallingford, Oxon, UK.

Carlton, J. (2000). Bayesian poker. Honours thesis, School of Computer Science and Software Engineering, Monash University. www.csse.monash.edu.au/hons/projects/2000/Jason.Carlton/.

Carnap, R. (1962). *The Logical Foundations of Probability* (Second ed.). University of Chicago.

Cartwright, N. (1989). *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.

Cartwright, N. (2001). What is wrong with Bayes nets? *The Monist 84*, 242–64.

Caruana, R. and A. Niculescu-Mizil (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning*, pp. 161–168.

Chaitin, G. (1966). On the length of programs for computing finite binary sequences. *Journal of the ACM 13*, 547–569.

Chan, H. and A. Darwiche (2003). When do numbers really matter? *Journal of Artificial Intelligence Research 17*, 265–287.

Charitos, T., L. van der Gaag, S. Visscher, K. Schurink, and P. Lucas (2009). A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients. *Expert Systems with Applications 36*(2), 1249–258.

Chee, Y., M. Burgman, and J. Carey (2005). Use of a Bayesian network decision tool to manage environmental flows in the Wimmera river, Victoria. Report No. 4, LWA/MDBC Project UMO43: Delivering Sustainability Through Risk Management, University of Melbourne, Australia.

Cheeseman, P. (1988). An inquiry into computer understanding. *Computational Intelligence 4*, 58–66.

Cheng, H. and G. Hadjisophocleous (2009). The modeling of fire spread in buildings by Bayesian network. *Fire Safety Journal 44*(6), 901–908.

Cheng, J. and M. Druzdzel (2000). AIS-BN: An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks. *Journal of Artificial Intelligence Research 13*, 155–188.

Cherniak, C. (1986). *Minimal Rationality*. Cambridge, MA: MIT Press.

Chickering, D. M. (1995). A tranformational characterization of equivalent Bayesian network structures. In *UAI95 – Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, pp. 87–98.

Chickering, D. M. (1996). Learning equivalence classes of Bayesian network structures. In *UAI96 – Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pp. 150–157.

Chihara, C. and R. Kennedy (1979). The Dutch book argument: Its logical flaws, its subjective sources. *Philosophical Studies 36*, 19–33.

Chow, C. and C. Liu (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory 14*, 462–467.

Conati, C., A. Gertner, and K. VanLehn (2002). Using Bayesian Networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction 12*(4), 371–417.

Conati, C., A. Gertner, K. VanLehn, and M. Druzdzel (1997). On-line student modeling for coached problem solving using Bayesian Networks. In *UM97 – Proceedings of the Sixth International Conference on User Modeling*, pp. 231–242.

Cooper, G. (1984). *NESTOR: A Computer Based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge*. Ph. D. thesis, Medical Computer Science Group, Stanford University, Stanford, CA.

Cooper, G. (1997). A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Mining and Knowledge Discovery 1*, 203–224.

Cooper, G., D. Dash, J. Levander, W. Wong, W. Hogan, and M. Wagner (2006). Bayesian methods for diagnosing outbreaks. In M. Wagner, A. Moore, and R. Aryel (Eds.), *Handbook of Biosurveillance*, pp. 273–288. Academic Press.

Cooper, G. and C. Yoo (1999). Causal discovery from a mixture of experimental and observational data. In *UAI99 – Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 116–125.

Cooper, G. F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence 42*, 393–405.

Cooper, G. F. and E. Herskovits (1991). A Bayesian method for constructing Bayesian belief networks from databases. In *UAI91 – Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pp. 86–94.

Cooper, G. F. and E. Herskovits (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning 9*, 309–347.

Coupe, V., N. Peek, J. Ottenkamp, and J. Habbema (1999). Using sensitivity analysis for efficient quantification of a belief network. *Artificial Intelligence in Medicine 17*, 223–247.

Coupé, V. and L. van der Gaag (2002). Properties of sensitivity analysis of Bayesian belief networks. *Annals of Mathematics and Artificial Intelligence 36*, 323–356.

Coupé, V., L. van der Gaag, and J. Habbema (2000). Sensitivity analysis: an aid for belief-network quantification. *Knowledge Engineering Review 15*, 1–18.

Cousins, S., W. Chen, and M. Frisse (1991a). CABeN: A collection of algorithms for belief networks. Technical Report WUCS-91-25, Department of Computer Science, Washington University.

Cousins, S., W. Chen, and M. Frisse (1991b). CABeN: A collection of algorithms for belief networks. Technical Report WUCS-91-25, Department of Computer Science, Washington University.

Covelo, V., von D. Winterfeldt, and P. Slovic (1986). Risk communication: A review of the literature. *Risk Abstracts 3*(4), 171–182.

Cover, T. and J. Thomas (2006). *Elements of information theory* (2nd ed.). New York: Wiley.

Cowell, R., A. Dawid, T. Hutchinson, and D. Spiegelhalter (1991). A Bayesian expert system for the analysis of an adverse drug reaction. *Artificial Intelligence in Medicine 3*(5), 257–270.

Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. Spiegelhalter (1999). *Probabilistic Networks and Expert Systems*. Statistics for Engineering and Information Science. New York: Springer Verlag.

D'Agostino, R., M. Russell, and D. Huse (2000). Primary and subsequent coronary risk appraisal: new results from the Framingham study. *American Heart Journal 139*, 272–81.

Dagum, P. and A. Galper (1993). Forecasting sleep apnea with dynamic network models. In *UAI93 – Proc of the Ninth Conference on Uncertainty in Artificial Intelligence*, pp. 65–71.

Dagum, P., A. Galper, and E. Horvitz (1992). Dynamic network models for forecasting. In D. Dubois, M. Wellman, B. D'Ambrosio, and P. Smets (Eds.), *UAI92 – Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence*, pp. 41–48.

Dagum, P. and M. Luby (1993). Approximating probabilistic inference in belief networks is NP-hard. *Artificial Intelligence 60*, 141–153.

Dai, H., K. B. Korb, C. S. Wallace, and X. Wu (1997). A study of casual discovery with weak links and small samples. In *IJCAI97 – Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, pp. 1304–1309.

Daly, R., Q. Shen, and S. Aitken (2010). Learning Bayesian networks: Approaches and issues. *The Knowledge Engineering Review*. Forthcoming.

Davies, J. (1995). Mobile monitoring of lower-limb amputees and fall diagnosis in elderly patients. Thesis report. Department of Electrical and Computer Systems Engineering, Monash University.

Dawid, A., J. Mortera, and P. Vicard (2006). Representing and solving complex DNA identification cases using Bayesian networks. In *International Congress Series*, Volume 1288, pp. 484–491. Elsevier.

de Campos, L., J. Fernández-Luna, J. Gámez, and J. Puerta (2000). Ant colony optimization for learning Bayesian networks. *International Journal of Approximate Reasoning 31*, 291–311.

de Finetti, B. (1964). Foresight: Its logical laws, its subjective sources. In Kyburg and Smokler (Eds.), *Studies in Subjective Probability*. New York: Wiley. (Translation of "La prévision" Annales de l'Institut Henri Poincaré, 1937.)

de Laplace, P. S. (1820/1951). *A Philosophical Essay on Probabilities*. Dover.

de Melo, A. and A. Sanchez (2008). Software maintenance project delays prediction using Bayesian networks. *Expert Systems with Applications 34*(2), 908–919.

de Santa Olalla, F. M., A. Dominguez, F. Ortega, A. Artigao, and C. Fabeiro (2007). Bayesian networks in planning a large aquifer in Eastern Mancha, Spain. *Environmental Modeling & Software 22*(8), 1089–1100.

Dean, T. and K. Kanazawa (1989). A model for reasoning about persistence and causation. *Computational Intelligence 5*, 142–150.

Dean, T. and M. P. Wellman (1991). *Planning and Control*. San Mateo, CA.: Morgan Kaufman Publishers.

DeGroot, M. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

DeGroot, M. and M. Schervish (2002). *Probability and Statistics* (Third ed.). Reading, MA: Addison-Wesley.

Demographia. Demographic statistics. http://www.demographia.com.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royals Statistical Society, B 39*, 1–38.

Dietterich, T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation 10*(7), 1895–1923.

Díez, F., J. Mira, E. Iturralde, and S. Zubillaga (1997). Diaval, a Bayesian expert system for echocardiography. *Artificial Intelligence in Medicine 10*(1), 59–73.

Domingos, P. and M. Pazzani (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning 29*(2), 103–130.

Draper, D. (1995). *Localized partial evaluation of belief networks*. Ph. D. thesis, Department of Computer Science, University of Washington.

Druzdzel, M. (1996). Qualitative verbal explanations in Bayesian belief networks. *Artificial Intelligence and Simulation of Behaviour Quarterly 94*, 43–54.

Druzdzel, M. J. (1999, July 18-22). SMILE: Structural modeling, inference, and learning engine and GeNIe: A development environment for graphical decision-theoretic models. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI–99)*, Orlando, FL, pp. 902–903.

Duda, R. and P. Hart (1973). *Pattern classification and scene analysis*. New York: Wiley.

Duda, R. O., P. E. Hart, and N. J. Nilsson (1976). Subjective Bayesian methods for rule-based inference systems. In *AFIPS*, Volume 45, pp. 1075–1082.

Edwards, A. L. (1984). *An Introduction to Linear Regression and Correlation* (Second ed.). New York: W.H. Freeman.

Ellison, A. (1996). An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications 6*, 1036–1046.

Erskine, W. D. (1996). Downstream hydrogeomorphic impacts of Eildon Reservoir on the mid-Goulburn River, Victoria. *Proceedings of the Royal Society of Victoria. 108*, 1–15.

Ettema, J., S. Østergaard, and A. Kristensen (2009). Estimation of probability for the presence of claw and digital skin diseases by combining cow-and herd-level information using a Bayesian network. *Preventive Veterinary Medicine 92*(1-2), 89–98.

Falzon, L. (2006). Using Bayesian network analysis to support centre of gravity analysis in military planning. *European Journal of Operational Research 170*(2), 629–643.

Feigenbaum, E. (1977). The art of artificial intelligence: Themes and case studies of knowledge engineering. In *Fifth International Conference on Artificial Intelligence – IJCAI-77*, San Mateo, CA, pp. 1014–1029. Morgan Kaufmann.

Fernández, A. and A. Salmerón (2008). BayesChess: A computer chess program based on Bayesian networks. *Pattern Recognition Letters 29*(8), 1154–1159.

Findler, N. (1977). Studies in machine cognition using the game of poker. *Communications of the ACM 20*, 230–245.

Fischoff, B. and M. Bar-Hillel (1984). Focusing techniques: A shortcut to improving probability judgments? *Organizational Behavior and Human Performance 34*, 339–359.

Flores, J., A. Nicholson, A. Brunskill, S. Mascaro, and K. Korb (2010). Incorporating expert knowledge when learning Bayesian network structure: Heart failure as a case study. Technical Report 2010/3. Bayesian Intelligence. http://www.Bayesian-intelligence.com/publications.php.

Flynn, J. (2002). The effectiveness of a computer games program in teaching decimal notation. Honours thesis, University of Melbourne.

Forbes, J., T. Huang, K. Kanazawa, and S. Russell (1995). The BATmobile: Towards a Bayesian automated taxi. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, pp. 1878–1885.

Fowler, M. (2003). *UML distilled: A brief guide to the standard object modeling language* (3rd ed.). Addison-Wesley.

Friedman, N., D. Geiger, and M. Goldszmidt (1997). Bayesian network classifiers. *Machine Learning 29*(2), 131–163.

Friedman, N. and M. Goldszmidt (1999). Learning Bayesian networks with local structure. In M. Jordan (Ed.), *Learning in Graphical Models*, pp. 421–459. MIT.

Friedman, N., I. Nachman, and D. Peér (1999). Learning Bayesian network structure from massive datasets: The "sparse candidate" algorithm. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 206–215.

Fung, R. and K.-C. Chang (1989). Weighting and integrating evidence for stochastic simulation in Bayesian networks. In *Proceedings of the Fifth Conference on Uncertainty in Artificial Intelligence (UAI-89)*, pp. 475–482.

Fung, R. and B. del Favero (1994). Backward simulation in Bayesian networks. In *UAI94 – Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pp. 227–234.

Galán, S., F. Aguado, F. Diez, and J. Mira (2002). NasoNet, modeling the spread of nasopharyngeal cancer with networks of probabilistic events in discrete time. *Artificial Intelligence in Medicine 25*(3), 247–264.

Garey, M. and D. Johnson (1979). *Computers and Intractability*. New York: W.H. Freeman.

Geiger, D. and D. Heckerman (1991). Advances in probabilistic reasoning. In *Proceedings of the Proceedings of the Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-91)*, pp. 118–126.

Geman, S., D. Geman, K. Abend, T. Harley, and L. Kanal (1993). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *Journal of Applied Statistics 20*, 25–62.

Getoor, L., J. Rhee, D. Koller, and P. Small (2004). Understanding tuberculosis epidemiology using structured statistical models. *Artificial Intelligence in Medicine 30*(3), 233–56.

Gigerenzer, G. and U. Hoffrage (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review 102*, 684–704.

Ginac, F. (1997). *Customer Oriented Software Quality Assurance*. New York: Prentice Hall.

Gippel, C. J. and B. L. Finlayson (1993). Downstream environmental impacts of regulation of the Goulburn River, Victoria. In *Hydrology and Water Resources Symposium, Newcastle*.

Glymour, C. (1980). *Theory and Evidence*. Princeton, NJ: Princeton University.

Glymour, C. and G. Cooper (Eds.) (1999). *Computation, causation, and discovery*. MIT Press.

Good, I. (1952). Rational decisions. *Jrn. of the Royal Statistical Society, B 14*, 107–114.

Goodman, N. (1973). *Fact, Fiction and Forecast* (Third ed.). Indianapolis: Bobbs-Merrill.

Gowadia, V., C. Farkas, and M. Valtorta (2005). Paid: A probabilistic agent-based intrusion detection system. *Computers & Security 24*(7), 529–545.

Graham, P. (2003). Better Bayesian filtering. http://www.paulgraham.com/better.html.

Group, B. H. S. (2004, December). The Busselton health study. Website. http://bsn.uwa.edu.au/surveys.htm.

Gupta, S. and H. Kim (2008). Linking structural equation modeling to Bayesian networks: Decision support for customer retention in virtual communities. *European Journal of Operational Research 190*(3), 818–833.

Guyon, I., C. Aliferis, G. Cooper, A. Elisseeff, J.-P. Pellet, P. Spirtes, and A. Statnikov (Eds.) (2008). *JMLR workshop and conference proceedings: Causation and prediction challenge (WCCI 2008)*, Volume 3. Journal of Machine Learning Research.

Haddawy, P., J. Jacobson, and C. E. J. Kahn (1997). BANTER: a Bayesian network tutoring shell. *Artificial Intelligence in Medicine 10*, 177–200.

Hagen, R. and E. Sonenberg (1993). Automated classification of student misconceptions in physics. In C. Rowles, H. Liu, and N. Foo (Eds.), *Proceedings of the 1993 Australian Joint Conference on Artificial Intelligence, AI'93*, pp. 153–159. World Scientific.

Hájek, A. (2008). Dutch book arguments. In P. Anand, P. Pattanaik, and C. Puppe (Eds.), *The Oxford Handbook of Rational and Social Choice*. Oxford University Press.

Halpern, J. and J. Pearl (2005a). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science 56*, 843–887.

Halpern, J. and J. Pearl (2005b). Causes and explanations: A structural-model approach. part II: explanations. *The British Journal for the Philosophy of Science 56*, 889–911.

Hand, D. and K. Yu (2001). Idiot's Bayes: Not so stupid after all? *International Statistical Review 69*, 385–398.

Handfield, T., C. Twardy, K. Korb, and G. Oppy (2008). The metaphysics of causal models. *Erkenntnis 68*, 149–168.

Harnad, S. (1989). Minds, machines and Searle. *Journal of Theoretical and Experimental Artificial Intelligence 1*, 5–25.

Hart, B. and C. Pollino (2009). Bayesian modeling for risk-based environmental water allocation. Waterline Report Series No 14, National Water Commission, Canberra, Australia. http://www.nwc.gov.au/resources/documents/Waterlines_14_-_Bayesian_COMPLETE.pdf.

Heckerman, D. (1991). *Probabilistic similarity networks*. Cambridge, MA: MIT Press.

Heckerman, D. (1998). A tutorial on learning with Bayesian networks. In M. Jordan (Ed.), *Learning in Graphical Models*, pp. 301–354. MIT.

Heckerman, D. and D. Geiger (1995). Learning Bayesian networks: A unification for discrete and Gaussian domains. In *UAI95 – Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, San Francisco, pp. 274–284.

Heckerman, D. E. (1986). Probabilistic interpretations for MYCIN's certainty factors. In *Uncertainty in Artificial Intelligence Vol 1*, pp. 167–96. Amsterdam: Elsevier.

Helm, S. (2002). The use of computer games to improve understanding about decimal notation. Honours thesis, University of Melbourne.

Helsper, E. M. and L. C. van der Gaag (2007). Ontologies for probabilistic networks: a case study in the oesophageal-cancer domain. *The Knowledge Engineering Review 22*(1), 67–86.

Henriksen, H., P. Rasmusssen, G. Brandt, D. von Bulow, and F. Jensen (2007). Bayesian networks as a participatory modeling tool for groundwater protection. In *Topics on System Analysis and Integrated Water Resource Management*, pp. 49–72. Elsevier Science Ltd.

Henrion, M. (1988). Propagating uncertainty in Bayesian networks by logic sampling. In *Uncertainty in Artificial Intelligence Vol 2*, pp. 149–163. Amsterdam: North-Holland.

Henrion, M., M. Pradhan, B. Del Favero, K. Huang, G. Provan, and P. O'Rorke (1996). Why is diagnosis using belief networks insensitive to imprecision in probabilities? In *UAI96 – Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pp. 307–314.

Hesslow, G. (1976). Discussion: Two notes on the probabilistic approach to causality. *Philosophy of Science 43*, 290–292.

Holmes, D. and L. Jain (Eds.) (2008). *Innovations in Bayesian Networks: Theory and Applications*. Springer Verlag.

Holte, R. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine Learning 11*, 63–90.

Hoover, K. (2003). Nonstationary time series, cointegration, and the principle of the common cause. *The British Journal for the Philosophy of Science 54*, 527–551.

Hope, L. (2008). *Information Measures for Causal Explanation and Prediction*. Ph. D. thesis, Monash University.

Hope, L. and K. Korb (2002). Bayesian information reward. In *15th Australian Joint Conference on Artificial Intelligence (AI'02)*, pp. 272–283. Springer Verlag.

Hope, L. and K. Korb (2004). A Bayesian metric for evaluating machine learning algorithms. In *17th Australian Joint Conference on Artificial Intelligence (AI'04)*, pp. 991–997. Springer Verlag.

Hope, L., A. Nicholson, and K. Korb (2002). Knowledge engineering tools for probability elicitation. Technical report 2002/111, School of Computer Science and Software Engineering, Monash University.

Hope, L., A. Nicholson, and K. Korb (2006). TakeheartII: A tool to support clinical cardiovascular risk assessment. Technical report TR 2006/209, Clayton School of Information Technology, Monash University.

Horvitz, E. and M. Barry (1995). Display of information for time-critical decision making. In *UAI95 – Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Montreal, Canada, pp. 296–305.

Horvitz, E., J. Breese, D. Heckerman, D. Hovel, and K. Rommelse (1998). The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pp. 256–265.

Horvitz, E., H. Suermondt, and G. Cooper (1989). Bounded conditioning: Flexible inference for decisions under scarce resources. In *Proceedings of the Fifth Workshop on Uncertainty in Artificial Intelligence*, pp. 182–193.

Howard, C. and M. Stumptner (2005). Situation assessment with Object Oriented Probabilistic Relational Models. In *Proceedings of the Seventh International Conference on Enterprise Information Systems (ICEIS 2005)*, pp. 412–418.

Howard, R. and J. Matheson (1981). Influence diagrams. In R. Howard and J. Matheson (Eds.), *Readings in Decision Analysis*, pp. 763–771. Strategic Decisions Group, Menlo Park, CA.

Howson, C. (2001). The logical basis of uncertainty. In D. Corfield and J. Williamson (Eds.), *Foundations of Bayianism*. Dordrecht: Kluwer Academic.

Howson, C. and P. Urbach (2007). *Scientific Reasoning: The Bayesian Approach* (3rd ed.). La Salle, IL: Open Court.

Hume, D. (1739/1962). *A Treatise of Human Nature*. Cleveland: World Publishing.

Husmeier, D., S. Roberts, and R. Dybowski (2003). *Applications of Probabilistic Modeling in Medical Informatics and Bioinformatics*. London: Springer Verlag.

Innocent (2002). Innocent: Fighting miscarriages of justice. http://www.innocent.org.uk.

Jaakkola, T. S. and M. I. Jordan (1999). Variational probabilistic inference and the QMR-DT network. *Journal of Artificial Intelligence Research 10*, 291–322.

Jaynes, E. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics SSC-4*, 227–241.

Jeffrey, R. (1983). *The Logic of Decision* (2nd ed.). New York: McGraw Hill.

Jensen, A. (1995). *A probabilistic model based decision support system for mildew management in winter wheat*. Ph. D. thesis, Aalborg University.

Jensen, F. (1988). Junction trees and decomposable hypergraphs. Research report, Judex Datasystemer A/S, Aalborg, Denmark.

Jensen, F., S. Lauritzen, and K. Olesen (1990). Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly 4*, 269–282.

Jensen, F. V. (1996). *An Introduction to Bayesian Networks*. New York: Springer Verlag.

Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs*. New York: Springer Verlag.

Jensen, F. V., B. Chamberlain, T. Nordahl, and F. Jensen (1991). Analysis in HUGIN of data conflict. In P. P. Bonissone, M. Henrion, L. M. Kanal, and J. F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 6*, pp. 519–528. Elsevier.

Jensen, F. V., U. Kjærulff, K. G. Olesen, and J. Pedersen (1989). Et forprojekt til et ekspertsystem for drift af spildevandsrensning (a prototype expert system for control of waste water treatment). Technical report, Judex Datasystemer A/S, Aalborg, Denmark. In Danish.

Jensen, F. V. and T. D. Nielsen (2007). *Bayesian Networks and Decision Graphs* (2nd ed.). New York: Springer Verlag.

Jensen, K., J. Toftum, and P. Friis-Hansen (2009). A Bayesian network approach to the evaluation of building design and its consequences for employee performance and operational costs. *Building and Environment 44*(3), 456–462.

Jensen, T., A. Kristensen, N. Toft, N. Baadsgaard, S. Østergaard, and H. Houe (2009). An object-oriented Bayesian network modeling the causes of leg disorders in finisher herds. *Preventive Veterinary Medicine 89*(3-4), 237–248.

Jia, H., D. Liu, J. Chen, and X. Liu (2007). A hybrid approach for learning Markov equivalence classes of Bayesian network. *Knowledge Science, Engineering and Management*, 611–616.

Jiang, X., D. Neill, and G. Cooper (2010). A Bayesian network model for spatial event surveillance. *International Journal of Approximate Reasoning 51*(2), 224–239.

Jitnah, N. (1993). Bayesian poker. Honours thesis, Dept. of Computer Science, Monash University.

Jitnah, N. (2000). *Using mutual information for approximate evaluation of Bayesian networks*. Ph. D. thesis, Monash University, School of Computer Science and Software Engineering.

Jordan, M. (1986). An introduction to linear algebra in parallel distributed processing. In J. M. D.E. Rumelhard and the PDP research group (Eds.), *Parallel Distributed Processing, Vol. 1.*, pp. 365–422. Cambridge, MA: MIT Press.

Jordan, M. (Ed.) (1999). *Learning in Graphical Models*. MIT.

K. B. Korb, R. M. I. Z. (1997). A cognitive model of argumentation. In *Proceedings of the Cognitive Science Society Meeting*, pp. 400–405.

Kahn, C. et al. (1997). Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Computers in biology and medicine 27*(1), 19–29.

Kahneman, D. and A. Tversky (1973). On the psychology of prediction. *Psychological Review 80*, 430–454.

Kalman, R. (1960, March). A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Engineering 82*, 34–45.

Kanazawa, K. (1992). *Probability, Time, and Action*. Ph. D. thesis, Brown University, Providence, RI.

Kanazawa, K., D. Koller, and S. Russell (1995). Stochastic simulation algorithms for dynamic probabilistic networks. In *UAI95 – Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pp. 346–351.

Kannel, W. B., J. D. Neaton, D. Wentworth, H. E. Thomas, J. Stamler, and S. B. Hulley (1986). Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screend for the MRFIT. *American Heart Journal 112*, 825–836.

Karzanov, A. and L. Khachiyan (1990). On the conductance of order Markov chains. Technical Report DIMACS 90-60, Dept of Computer Science, Rutgers.

Katz, E., V. Heleg-Shabtai, A. Bardea, I. Willner, H. Rau, W. Haehnel, G. Arroyo-Figueroa, L. Sucar, and A. Villavicencio (1998). Probabilistic temporal reasoning and its application to fossil power plant operation. *Expert Systems with Applications 15*(3), 317–324.

Kim, J. and J. Pearl (1983). A computational model for causal and diagnostic reasoning in inference systems. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 190–193.

Kim, Y. and M. Valtorta (1995). On the detection of conflicts in diagnostic Bayesian networks. In *UAI95 – Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pp. 362–367.

Kipersztok, O. and H. Wang (2001). Another look at sensitivity of Bayesian networks to imprecise probabilities. In *Proceedings of the Eighth Workshop on Artificial Intelligence and Statistics (AISTAT-2001)*, Florida.

Kjærulff, U. (1992). A computational scheme for reasoning in dynamic probabilistic networks. In *UAI92 – Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence*, pp. 121–129.

Kjærulff, U. (1994). Reduction of computation complexity in Bayesian networks through removal of weak dependencies. In *UAI94 – Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pp. 374–382.

Kjærulff, U. (1995). dHugin: A computational system for dynamic time-sliced Bayesian networks. *International Journal of Forecasting, Special Issue on Probability Forecasting 11*, 89–111.

Kjærulff, U. and L. C. van der Gaag (2000). Making sensitivity analysis computationally efficient. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pp. 317–325.

Kjærulff, U. B. and A. L. Madsen (2008). *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Springer Verlag.

Kline, J., A. Novobilski, C. Kabrhel, P. Richman, and D. Courtney (2005). Derivation and validation of a Bayesian network to predict pretest probability of venous thromboembolism. *Annals of Emergency Medicine 45*(3), 282–290.

Knuiman, M. (2005). Personal communication.

Knuiman, M. W., H. T. Vu, and H. Bartholomew (1998). Multivariate risk estimation for coronary heart disease: the Busselton Health Study. *Australian and New Zealand Journal of Public Health 22*, 747–753.

Koller, D. and N. Friedman (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Koller, D. and A. Pfeffer (1997a). Object-Oriented Bayesian networks. In *UAI97 – Proceedings of the Thirteenth Conference on Uncertainty in Uncertainty in Artificial Intelligence*, pp. 302–313.

Koller, D. and A. Pfeffer (1997b). Representations and solutions for game-theoretic problems. *Artificial Intelligence 94*, 167–215.

Koller, D. and M. Sahami (1996). Toward optimal feature selection. In *International Conference on Machine Learning*, pp. 284–292.

Kolmogorov, A. (1933). *Grundbegriffe der Warhscheinlichkeitsrechnung*. Berlin: Springer Verlag.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information and Transmission 1*, 1–7.

Kononenko, I. and I. Bratko (1991). Information-based evaluation criterion for classifier's performance. *Machine Learning 6*, 67–80.

Korb, K. (1991). Searle's Artificial Intelligence program. *Journal of Theoretical and Experimental Artificial Intelligence 3*, 283–296.

Korb, K., L. Hope, and M. Hughes (2001). The evaluation of predictive learners: some theoretical and empirical results. In *ECML'01 – Twelfth European Conference on Machine Learning*, pp. 276–287. Springer Verlag.

Korb, K., C. Kopp, and L. Allison (1997). A statement on higher education policy in Australia. Technical report, 97/318. School of Computer Science and Software Engineering, Monash University.

Korb, K., R. McConachy, and I. Zukerman (1997). A cognitive model of argumentation. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, pp. 400–405.

Korb, K. and E. Nyberg (2006). The power of intervention. *Minds and Machines 16*, 289–302.

Korb, K. B. (1995). Inductive learning and defeasible inference. *Journal for Experimental and Theoretical Artificial Intelligence 7*, 291–324.

Korb, K. B., L. R. Hope, A. E. Nicholson, and K. Axnick (2004). Varieties of causal intervention. In *PRICAI'04—Proceedings of the 8th Pacific Rim International Conference on Artificial Intelligence*, Auckland, New Zealand, pp. 322–331.

Korb, K. B., A. E. Nicholson, and N. Jitnah (1999). Bayesian poker. In *UAI99 – Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 343–350.

Kornfeld, A. (1991, November). Causal diagrams: clarifying uncertainty. *Artificial Intelligence Expert*, 42–49.

Kristensen, K. and I. Rasmussen (2002). The use of a Bayesian network in the design of a decision support system for growing malting barley without use of pesticides. *Computers and Electronics in Agriculture 33*(3), 197–217.

Kuikka, S., M. Hilden, H. Gislason, S. Hanson, H. Sparholt, and O. Varis (1999). Modeling environmentaly driven uncertainties in Baltic cod (Gadus morhua): Management by Bayesian influence diagrams. *Canadian Journal of Fisheries and Aquatic Sciences 54*, 629–641.

Lacave, C., R. Atienza, and F. J. Díez (2000). Graphical explanation in Bayesian networks. *Lecture Notes in Computer Science 1933*, 122–129.

Lacave, C. and F. J. Díez (2002). A review of explanation methods for Bayesian networks. *The Knowledge Engineering Review 17*(2), 107–127.

Lam, W. and F. Bacchus (1993). Learning Bayesian belief networks: An approach based on the MDL principle. *Computational Intelligence 10*, 269–293.

Langseth, H. (2002). *Bayesian networks with applications in reliability analysis*. Ph. D. thesis, Dept. of Mathematical Sciences, NorvegianUniversity of Science and Technology.

Larrañaga, P., C. Kuijpers, R. Murga, and Y. Yurramendi (1996). Learning Bayesian network structures by searching for the bestordering with genetic algorithms. *IEEE Transactions on Systems, Man and Cybernetics, Part A 26*, 487–493.

Laskey, K. and S. Mahoney (1997). Network fragments: Representing knowledge for constructing probabilistic models. In *UAI97 – Proceedings of the Thirteenth Conference on Uncertainty in Uncertainty in Artificial Intelligence*, pp. 334–341.

Laskey, K. and S. Mahoney (2000). Network engineering for agile belief network models. *IEEE: Transactions on Knowledge and Data Engineering 12*(4), 487–498.

Laskey, K. B. (1993). Sensitivity analysis for probability assessments in Bayesian networks. In *UAI93 – Proc of the Ninth Conference on Uncertainty in Artificial Intelligence*, pp. 136–142.

Lauritzen, S. and N. Wermuth (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics 17*, 31–57.

Lauritzen, S. L. and D. J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society 50*(2), 157–224.

Lee, C. and K. Lee (2006). Application of Bayesian network to the probabilistic risk assessment of nuclear waste disposal. *Reliability Engineering and System Safety 91*(5), 515–532.

Lerner, U., E. Segal, and D. Koller (2001). Exact inference in networks with discrete children of continuous parents. In *UAI01 – Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 319–328.

Lewis, D. (1980). A subjectivist's guide to objective chance. In Jeffrey (Ed.), *Studies in inductive logic and probability, volume II*, pp. 263–293. University of California.

Lichtenstein, S., B. Fischhoff, and L. D. Phillips (1982). Calibration of probabilities: The state of the art to 1980. In *Judgment under Uncertainty: Heuristics and Biases*, pp. 306–334. Cambridge.

Liew, D. and S. Rogers (2004, September). Personal communication. http://www.csse.monash.edu.au/~ctwardy/bnepi/Epidem/ 20040930.txt.

Lindley, D. (1972). *Bayesian Statistics: A Review*. Philadelphia: Society for Industrial and Applied Mathematics.

Lindley, D. (1985). *Making Decisions* (2nd ed.). John Wiley.

Loehlin, J. C. (1998). *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis* (Third ed.). Mahwah, N.J.: Lawrence Erlbaum.

Lucas, P., L. van der Gaag, and A. Abu-Hanna (2004). Bayesian networks in biomedicine and health-care. *Artificial Intelligence in Medicine 30*(3), 201–214.

Luger, G. and W. Stubblefield (1993). *Artificial intelligence: Structures and Strategies for Complex Problem Solving* (second ed.). Menlo Park, CA: Benjamin/Cummings.

Mackay, D. (1998). Introduction to Monte Carlo methods. In M. Jordan (Ed.), *Learning in Graphical Models*, pp. 175–204. MIT.

Madigan, D., S. A. Andersson, M. D. Perlman, and C. T. Volinsky (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Communications in Statistics: Theory and Methods 25*, 2493–2519.

Madigan, D., K. Mosurski, and R. G. Almond (1997). Graphical explanation in belief networks. *Journal of Computer and Graphical Statistics 6*, 160–181.

Madigan, D. and A. E. Raftery (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association 89*, 1535–1546.

Madigan, D. and I. York (1995). Bayesian graphical models for discrete data. *International Statistical Review 63*, 215–232.

Maglogiannis, I., E. Zafiropoulos, A. Platis, and C. Lambrinoudakis (2006). Risk analysis of a patient monitoring system using Bayesian network modeling. *Journal of Biomedical Informatics 39*(6), 637–647.

Mahoney, S. and K. Laskey (1999). Representing and combining partially specified CPTs. In *UAI99 – Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, Sweden, pp. 343–350.

Marcot, B. (1999). A process for creating Bayian belief network models of species-environment relations. Technical report, USDA Forest Service, Portland, Oregon.

Marcot, B., R. Holthausen, M. Raphael, M. Rowland, and M. Wisdom (2001). Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management 153*(1-3), 29–42.

Marcot, B., J. Steventon, G. Sutherland, and R. McCann (2006). Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research 36*(12), 3063–3074.

Mayo, M. and A. Mitrovic (2001). Optimising ITS behaviour with Bayesian networks and decision theory. *International Journal of Artificial Intelligence in Education 12*, 124–153.

McConachy, R., K. B. Korb, and I. Zukerman (1998). Deciding what not to say: An attentional-probabilistic approach to argument presentation. In *Proceedings of the Cognitive Science Society Meeting*, pp. 669–674.

McDermott, D. (1987). A critique of pure reason. *Computational Intelligence 3*, 151–237.

McGain, K. (2010). Opponent modeling in Bayesian poker. Honours thesis, Clayton School of IT, Monash University (forthcoming).

McGowan, R. D. (1997). Ambulation monitoring and fall detection using dynamic belief networks. honours thesis. dept. of electrical and computer systems engineering, monash university.

McIntosh, J., K. Stacey, C. Tromp, and D. Lightfoot (2000). Designing constructivist computer games for teaching about decimal numbers. In J. Bana and A. Chapman (Eds.), *Proceedings of the Twenty Third Annual Conference of the Mathematical Education Research Group of Australasia*, pp. 409–416. Freemantle: MERGA, Inc.

McNeil, J. J., A. Peeters, D. Liew, S. Lim, and T. Vos (2001). A model for predicting the future incidence of coronary heart disease within percentiles of coronary heart disease risk. *Journal of Cardiovascular Risk 8*, 31–37.

McNeill, J., R. MacEwan, and D. Crawford (2006). Using GIS and a land use impact model to assess risk of soil erosion in West Gippsland. *APPLIED GIS 2*(3).

Mcroy, S. W., S. S. Ali, and S. M. Haller (1997). Uniform knowledge representation for language processing in the b2 system. *Natural Language Engineering 3*(2), 123–145.

Meek, C. and D. M. Chickering (2003). Practically perfect. In *UAI03 – Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, San Francisco.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics 21*, 1087–1091.

Michie, D., D. Spiegelhalter, C. Taylor, and J. Campbell (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood.

Middleton, B., M. Shwe, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper (1991). Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base II. Evaluation of diagnostic performance. *Methods in Information in Medicine 30*, 256–267.

Miller, A. C., M. M. Merkhofer, R. A. Howard, J. E. Matheson, and T. R. Rice (1976). Development of automated aids for decision analysis. Technical report, SRI International, Menlo Park, CA.

Mitchell, T. (1997). *Machine Learning*. McGraw Hill.

Mittal, A. and A. Kassim (Eds.) (2007). *Bayesian network technologies: Applications and graphical models*. Hershey, PA: IGI Publishing.

Mohan, C., K. Mehrotra, P. Varshney, and J. Yang (2007). Temporal uncertainty reasoning networks for evidence fusion with applications to object detection and tracking. *Information Fusion 8*(3), 281–294.

Montani, S., L. Portinale, A. Bobbio, and D. Codetta-Raiteri (2008). Radyban: A tool for reliability analysis of dynamic fault trees through conversion into dynamic Bayesian networks. *Reliability Engineering & System Safety 93*(7), 922–932.

Montero, J. and L. Sucar (2006). A decision-theoretic video conference system based on gesture recognition. In *Automatic Face and Gesture Recognition, 2006. In FGR 2006, 7th International Conference on Automatic Face and Gesture Recognition*. pp. 387–392.

Monti, S. and G. F. Cooper (1998). A multivariate discretization method for learning hybrid Bayesian networks from data. In *UAI98 – Proceedings of the Fourteenth Conference on Uncertainty in Uncertainty in Artificial Intelligence*, pp. 404–413.

Morgan, M. and M. Henrion (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. London: Cambridge University Press.

Morjaria, M. A., F. J. Rink, W. D. Smith, G. Klempner, G. Burns, and J. Stein (1993). Commercialization of EPRI's generator expert monitoring system. In *Proceedings of the EPRI Conference on Advanced Computer Applications in the Electric Utility Industry*, Phoenix.

Mosteller, F. (1973). *Statistics by Example*. Reading, MA: Addison-Wesley.

Murphy, K., Y. Weiss, and M. I. Jordan (1999). Loopy belief propagation for approximate inference: an empirical study. In *UAI99 – Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 467–475.

Murphy, K. P. (2002). *Dynamic Bayesian networks: Representation, inference and learning*. Ph. D. thesis, Department of Computer Science, University of California, Berkeley.

Murphy, P. and D. Aha (1995). UCI repository of machine learning databases. http://www.ics.uci.edu/˜mlearn/MLRepository.html.

Nadeau, C. and Y. Bengio (2003). Inference for the generalization error. *Machine Learning 52*(3), 239–281.

Nägele, A., M. Dejori, and M. Stetter (2007). Bayesian substructure learning-approximate learning of very large network structures. In *Proceedings of the 18th European Conference on Machine Learning; Lecture Notes in AI*, Volume 4701, pp. 238–249.

Nagl, S., M. Williams, and J. Williamson (2008). Objective Bayesian nets for systems modeling and prognosis in breast cancer. In D. Holmes and L. Jain (Eds.), *Innovations in Bayesian Networks: Theory and Applications*, pp. 131–167. Springer Verlag.

Naticchia, B., A. Fernandez-Gonzalez, and A. Carbonari (2007). Bayesian network model for the design of roofpond equipped buildings. *Energy and Buildings 39*(3), 258–272.

Neapolitan, R. E. (1990). *Probabilistic Reasoning in Expert Systems*. New York: Wiley and Sons.

Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Prentice Hall.

Neil, J. R. and K. B. Korb (1999). The evolution of causal models. In *Third Pacific-Asia Conf on Knowledge Discovery and Datamining (PAKDD-99)*, pp. 432–437. Springer Verlag.

Neil, J. R., C. S. Wallace, and K. B. Korb (1999). Learning Bayesian networks with restricted causal interactions. In *UAI99 – Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 486–493.

Neil, M., N. Fenton, and L. Nielson (2000). Building large-scale Bayesian networks. *The Knowledge Engineering Review 15*(3), 257–284.

Neil, M., N. Fenton, and M. Tailor (2005). Using Bayesian networks to model expected and unexpected operational losses. *Risk Analysis 25*(4), 963–972.

Neil, M., M. Tailor, D. Marquez, N. Fenton, and P. Hearty (2008). Modeling dependable systems using hybrid Bayesian networks. *Reliability Engineering & System Safety 93*(7), 933–939.

Nicholson, A. (1996). Fall diagnosis using dynamic belief networks. In *Proceedings of the Fourth Pacific Rim International Conference on Artificial Intelligence (PRICAI-96)*, pp. 206–217.

Nicholson, A., T. Boneh, T. Wilkin, K. Stacey, L. Sonenberg, and V. Steinle (2001). A case study in knowledge discovery and elicitation in an intelligent tutoring application. In *UAI01 – Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 386–394.

Nicholson, A. and N. Jitnah (1996). Belief network algorithms: a study of performance using domain characterisation. Technical Report 96/249, Department of Computer Science, Monash University.

Nicholson, A. E. (1992). *Monitoring discrete environments using dynamic belief networks*. Ph. D. thesis, Department of Engineering Sciences, Oxford.

Nicholson, A. E. and J. M. Brady (1992a). The data association problem when monitoring robot vehicles using dynamic belief networks. In *ECAI92 – Proceedings of the Tenth European Conference on Artificial Intelligence*, pp. 689–693.

Nicholson, A. E. and J. M. Brady (1992b). Sensor validation using dynamic belief networks. In *UAI92 – Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence*, San Mateo, CA, pp. 207–214.

Nicholson, A. E., C. R. Twardy, K. B. Korb, and L. R. Hope (2008). Decision support for clinical cardiovascular risk assessment. In O. Pourret, P. Naim, and B. Marcot (Eds.), *Bayesian Networks: A Practical Guide to Applications*, Statistics in Practice, pp. 33–52. Wiley.

Nikovski, D. (2000). Constructing Bayesian networks for medical diagnosis from incomplete and partially correct statistics. *IEEE Transactions on Knowledge and Data Engineering 12*(4), 509–516.

Norsys (1994-2010). Netica. http://www.norsys.com.

Novak, J. D. and A. J. Cańas (2008). The theory underlying concept maps and how to construct and use them. Technical report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition.

Nyberg, E. P. and K. B. Korb (2006). Informative interventions. Technical Report 2006/204, School of Computer Science, Monash University.

O'Donnell, R. (2010). *Flexible Causal Discovery with MML*. Ph. D. thesis, Monash University.

O'Donnell, R., K. Korb, and L. Allison (2007). Causal KL: Evaluating causal discovery. Technical Report 2007/207, Clayton School of IT, Monash University.

Olesen, K. (1993). Causal probabilistic networks with both discrete and continuous variables. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 15*(3), 275–279.

Oliver, J. (1993). Decision graphs – an extension of decision trees. In *Proc. fourth int. conf. on artificial intelligence*, pp. 343–350.

Oliver, J. and D. Hand (1994). Averaging over decision stumps. In *European Conference on Machine Learning*, pp. 231–231. Springer Verlag.

Onisko, A., M. Druzdzel, and H. Wasyluk (1998). A probabilistic model for diagnosis of liver disorders. In *Proceedings of the Seventh Symposium on Intelligent Information Systems (IIS-98)*, pp. 379–387.

Otto, L. and C. Kristensen (2004). A biological network describing infection with Mycoplasma hyopneumoniae in swine herds. *Preventive Veterinary Medicine 66*(1-4), 141–161.

Oukhellou, L., E. Côme, L. Bouillaut, and P. Aknin (2008). Combined use of sensor data and structural knowledge processed by Bayesian network: Application to a railway diagnosis aid scheme. *Transportation Research Part C: Emerging Technologies 16*(6), 755–767.

Ozbay, K. and N. Noyan (2006). Estimation of incident clearance times using Bayesian networks approach. *Accident Analysis & Prevention 38*(3), 542–555.

O'Donnell, R., L. Allison, and K. Korb (2006). Learning hybrid Bayesian networks by MML. In *Proceedings of the 19th Australian Joint Conference on AI*, pp. 192–203. Springer Verlag.

O'Donnell, R., A. Nicholson, B. Han, K. Korb, M. Alam, and L. Hope (2006). Causal discovery with prior information. In *Proceedings of the 19th Australian Joint Conference on AI*, pp. 1162–1167. Springer Verlag.

Park, J. D. (2002). Map complexity results and approximation methods. In *UAI02 – Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, pp. 388–396.

Pauker, S. and J. Wong (2005). The influence of influence diagrams in medicine. *Decision Analysis 2*(4), 238–244.

Pearl, J. (1978). An economic basis for certain methods of evaluating probabilistic forecasts. *International Journal of Man-Machine Studies 10*, 175–183.

Pearl, J. (1982). Reverend Bayes on inference engines: a distributed hierarchical approach. In *AAAI82 – Proceedings of the Second National Conference on Artificial Intelligence*, pp. 133–136.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence 29*, 241–288.

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

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge, MA.

Pollino, C., P. Feehan, M. Grace, and B. Hart (2004). Fish communities and habitat changes in the highly modified goulburn river, australia. *Marine & Freshwater Research 55*, 769–780.

Pollino, C. and B. Hart (2006). Bayesian approaches can help make better sense of ecotoxicological information in risk assessments. *Australasian Journal of Ecotoxicology 11*, 56–57.

Pollino, C., A. White, and B. Hart (2007). Examination of conflicts and improved strategies for the management of an endangered eucalypt species using Bayesian networks. *Ecological Modeling 201*, 37–59.

Pollino, C., O. Woodberry, A. Nicholson, and K. Korb (2005). Parameterising Bayesian networks: A case study in ecological risk assessment. In *SIMMOD'05 – Proceedings of the 2005 Conference on Simulation and Modeling*, Bangkok, Thailand, pp. 289–297.

Pollino, C., O. Woodberry, A. Nicholson, K. Korb, and B. T. Hart (2007). Parameterisation of a Bayesian network for use in an ecological risk management case study. *Environmental Modeling and Software 22*(8), 1140–1152.

Poole, D., A. Mackworth, and R. Goebel (1998). *Computational Intelligence: A Logical Approach*. Oxford University Press.

Popper, K. (1959). The propensity interpretation of probability. *British Journal for the Philosophy of Science 10*, 25–42.

Pourret, O., P. Naim, and B. Marcot (Eds.) (2008). *Bayesian Networks: A Practical Guide to Applications*. Statistics in Practice. Wiley.

Pradhan, M., M. Henrion, G. Provan, B. D. Favero, and K. Huang (1996). The sensitivity of belief networks to imprecise probabilities: An experimental investigation. *Artificial Intelligence 85*(1-2), 363–397.

Pradhan, M., G. Provan, B. Middleton, and M. Henrion (1994). Knowledge engineering for large belief networks. In *UAI94 – Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, pp. 484–490.

Promedas. Promedas. http://www.promedas.nl.

Provost, F. and T. Fawcett (2001). Robust classification for imprecise environments. *Machine Learning 42*, 203–231.

Provost, F., T. Fawcett, and R. Kohavi (1998). The case against accuracy estimation for comparing induction algorithms. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pp. 445–453.

Pullar, D. and T. Phan (2007, 10-13 December 2007). Using a Bayesian network in a GIS to model relationships and threats to koala populations close to urban environments. In *Proceedings of the International Congress on Modeling and Simulation MODSIM2007*.

Pynadeth, D. and M. P. Wellman (1995). Accounting for context in plan recognition, with application to traffic monitoring. In *UAI95 – Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pp. 472–481.

Quinlan, J. (1983). Learning efficient classification procedures and their application to chess endgames. In R. Michalski, J. Carbonell, and T. Mitchell (Eds.), *Machine learning: An artificial intelligence approach*. Palo Alto: Tioga Publishing.

Quinlan, J. (1986). Induction of decision trees. *Machine learning 1*, 81–106.

Quinlan, J. (1993). *C4. 5: Programs for machine learning*. Morgan Kaufmann.

Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Computing Surveys 28*, 71–72.

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. New York: Random House.

Ramesh, R., M. Mannan, A. Poo, and S. Keerthi (2003). Thermal error measurement and modeling in machine tools. part ii. hybrid Bayesian network–support vector machine model. *International Journal of Machine Tools and Manufacture 43*(4), 405–419.

Ramsey, F. P. (1931). Truth and probability. In R. Braithwaite (Ed.), *The Foundations of Mathematics and Other Essays*. New York: Humanities Press.

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California.

Reichert, P., M. Borsuk, M. Hostmann, S. Schweizer, C. Sporri, K. Tockner, and B. Truffer (2007). Concepts of decision support for river rehabilitation. *Environmental Modeling & Software 22*(2), 188–201.

Renninger, H. and H. von Hasseln (2002). Object-oriented dynamic Bayesian network-templates for modeling mechatronic systems. In *International Workshop on Principles of Diagnosis (DX-02)*. http://www.dbai.tuwien.ac.at/user/dx2002/proceedings/dx02final21.pdf.

Resnick, L. B., P. Nesher, F. Leonard, M. Magone, S. Omanson, and I. Peled (1989). Conceptual bases of arithmetic errors: The case of decimal fractions. *Journal for Research in Mathematics Education 20*(1), 8–27.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica 14*, 465–471.

Rivas, T., J. Matias, J. Taboada, and A. Arguelles (2007). Application of Bayesian networks to the evaluation of roofing slate quality. *Engineering Geology 94*(1-2), 27–37.

Robinson, R. (1977). Counting unlabeled acyclic digraphs. *Combinatorial Mathematics V*, 28–43.

Ross, B. and E. Zuviria (2007). Evolving dynamic Bayesian networks with multi-objective genetic algorithms. *Applied Intelligence 26*, 13–23.

Russell, S. and P. Norvig (1995). *Artificial Intelligence: A Modern Approach* (First ed.). Prentice Hall Series in Artificial Intelligence. Englewood Cliffs, NJ: Prentice Hall.

Russell, S. and P. Norvig (2010). *Artificial Intelligence: A Modern Approach* (3rd ed.). Prentice Hall Series in Artificial Intelligence. Englewood Cliffs, NJ: Prentice Hall.

Sackur-Grisvard, C. and F. Leonard (1985). Intermediate cognitive organization in the process of learning a mathematical concept: The order of positive decimal numbers. *Cognition and Instruction 2*, 157–174.

Sahin, F., M. Yavuz, Z. Arnavut, and O. Uluyol (2007). Fault diagnosis for airplane engines using Bayesian networks and distributed particle swarm optimization. *Parallel Computing 33*(2), 124–143.

Sakamoto,Y. et al. (1986). *Akaike information criterion statistics*. KTK scientific publishers.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University.

Savage, L. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association 66*(336), 783–801.

Schach, S. (2008). *Object-Oriented software engineering*. New York: McGraw-Hill.

Scheines, R., P. Spirtes, C. Glymour, and C. Meek (1994). *TETRAD II: Tools for Discovery*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*, 461–464.

Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences 3*, 417–457.

Sewell, W. and V. Shah (1968). Social class, parental encouragement, and educational aspirations. *American Journal of Sociology 73*, 559–572.

Shachter, R. and M. Peot (1989). Simulation approaches to general probabilistic inference on belief networks. In *Proceedings of the Fifth Workshop on Uncertainty in Artificial Intelligence*, pp. 311–318.

Shachter, R. D. (1986). Evaluating influence diagrams. *Operations Research 34*, 871–882.

Shachter, R. D. and C. R. Kenley (1989). Gaussian influence diagrams. *Management Science 35*(5), 527–550.

Shafer, G. and P. Shenoy (1990). Probability propagation. *Annals of Mathematics and Artificial Intelligence 2*, 327–352.

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

Shimony, S. E. (1994). Finding MAPs for belief networks is NPhard. *Artificial Intelligence 68*, 399–410.

Shiratori, N. and N. Okude (2007). Bayesian networks layer model to represent anesthetic practice. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (ISIC*, pp. 674–679.

Shwe, M. and G. Cooper (1991). An empirical analysis of likelihood-weighting simulation on a large, multiply connected belief network. *Computers and Biomedical Research 24*, 453–475.

Shwe, M., B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper (1991). Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base I. The probabilistic model and inference algorithms. *Methods in Information in Medicine 30*, 241–255.

Simon, H. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association 49*, 467–479.

Sleeman, D. (1984). Mis-generalization: An explanation of observed mal-rules. In *Sixth Annual Conference of the Cognitive Science Society*, pp. 51–56.

Smith, C., A. Howes, B. Price, and C. McAlpine (2007). Using a Bayesian belief network to predict suitable habitat of an endangered mammalThe Julia Creek dunnart (Sminthopsis douglasi). *Biological Conservation 139*, 333–347.

Smith, S. F. (1983). Flexible learning of problem solving heuristics through adaptive search. In *IJCAI-83*, pp. 422–425.

Smith, W., J. Doctor, J. Meyer, I. Kalet, and M. Phillips (2009). A decision aid for intensity-modulated radiation-therapy plan selection in prostate cancer based on a prognostic Bayesian network and a markov model. *Artificial Intelligence in Medicine 46*(2), 119–130.

Sober, E. (1988). The principle of the common cause. In J. Fetzer (Ed.), *Probability and Causality*. Kluwer, Dordrecht.

Solomonoff, R. (1964). A formal theory of inductive inference, I and II. *Information and Control 7*, 1–22 and 224–254.

Sommerville, I. (2010). *Software Engineering* (Ninth ed.). Reading, MA: Addison-Wesley.

Spiegel, M., J. Schiller, and R. Srinivasan (2008). *Schaum's Outline of Probability and Statistics*. McGraw-Hill.

Spiegelhalter, D. (1986). Probabilistic reasoning in predictive expert systems. In J. Lemmer and L. Kanal (Eds.), *Uncertainty in Artificial Intelligence*. Amsterdam: Elsevier.

Spiegelhalter, D. J. and S. L. Lauritzen (1990). Sequential updating of conditional probabilities on directed graphical structures. *Networks 20*, 579–605.

Spirtes, P., C. Glymour, and R. Scheines (1993). *Causation, Prediction and Search*. Number 81 in Lecture Notes in Statistics. Springer Verlag.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction and Search* (Second ed.). Cambridge, MA: MIT Press.

Srinivas, S. and J. Breese (1989). IDEAL. Technical Report 23, Rockwell International Science Center.

Stacey, K., E. Sonenberg, A. Nicholson, T. Boneh, and V. Steinle (2003). A teaching model exploiting cognitive conflict driven by a Bayesian network. In *Proceedings of the Nineth International Conference on User Modeling (UM'2003)*, Pittsburgh, pp. 352–362.

Stacey, K. and V. Steinle (1998). Refining the classification of students' interpretations of decimal notation. *Hiroshima Journal of Mathematics Education 6*, 49–69.

Stacey, K. and V. Steinle (1999). A longitudinal study of childen's thinking about decimals: A preliminary analysis. In O. Zaslavsky (Ed.), *Proceedings of the Twenty Third Conference of the International Group for the Psychology of Mathematical Education*, Volume 4, Haifa, pp. 233–241. PME.

Stajduhar, I., B. Dalbelo-Basic, and N. Bogunovic (2009). Impact of censoring on learning Bayesian networks in survival modeling. *Artificial Intelligence in Medicine 47*(3), 199–217.

Stassopoulou, A. and M. Dikaiakos (2009). Web robot detection: A probabilistic reasoning approach. *Computer Networks 53*(3), 265–278.

Steel, D. (2006). Homogeneity, selection and the faithfulness condition. *Minds and Machines 16*, 303–317.

Steeneveld, W., L. van der Gaag, H. Barkema, and H. Hogeveen (2010). Simplify the interpretation of alert lists for clinical mastitis in automatic milking systems. *Computers and Electronics in Agriculture 71*(1), 50–56.

Steinle, V. and K. Stacey (1998). The incidence of misconceptions of decimal notation amongst students in Grades 5 to 10. In C. Kanes, M. Goos, and E. Warren (Eds.), *Teaching Mathematics in New Times, MERGA 21*, pp. 548–555. MERGA.

Suermondt, H. J. (1992). *Explanation in Bayesian belief networks*. Ph. D. thesis, Medical Information Sciences, Stanford University.

Sullivan, J. People v. Collins. Problems, cases, and materials on evidence. http://en.wikipedia.org/wiki/People_v._Collins.

Suzuki, J. (1996). Learning Bayesian belief networks based on the Minimum Description Length principle. In *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 462–470.

Tang, A., A. Nicholson, Y. Jin, and J. Han (2007). Using Bayesian belief networks for change impact analysis in architecture design. *Journal of Systems and Software 80*(1), 127–148.

Tang, Z., M. Taylor, P. Lisboa, and M. Dyas (2005). Quantitative risk modeling for new pharmaceutical compounds. *Drug discovery today 10*(22), 1520–1526.

Taylor, B. (2007). Opponent modeling in Bayesian poker player. Honours thesis, Clayton School of Information Technology, Monash University.

Teyssier, M. and D. Koller (2005). Ordering-based search: A simple and effective algorithm for learning Bayesian networks. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pp. 584–590. AUAI Press.

Thomson, S. (1995). Bayesian poker. Honours thesis, Dept of Computer Science, Monash University.

Thrun, S., J. Bala, E. Bloedorn, I. Bratko, B. Cestnik, J. Cheng, K. De Jong, S. Dzeroski, S. Fahlman, D. Fisher, et al. (1991). The MONK's problems: A performance comparison of different learning algorithms. *Revised version*, Technical Report CMU-CS-91-197, Carnegie Mellon University, CS Department.

Ticehurst, J., L. Newham, D. Rissik, R. Letcher, and A. Jakeman (2007). A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales, Australia. *Environmental Modeling & Software 22*(8), 1129–1139.

Tighe, M., C. Pollino, S. Cuddy, and S. Whitfield (2007, 10-13 December 2007). A Bayesian approach to assessing regional climate change pressures on natural resource conditions in the central west of nsw, australia. In *Proceedings of the International Congress on Modeling and Simulation MODSIM2007*, pp. 233–239.

Tsamardinos, I., L. Brown, and C. Aliferis (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning 65*, 31–78.

Tucker, A., V. Vinciotti, X. Liu, and D. Garway-Heath (2005). A spatio-temporal Bayesian network classifier for understanding visual field deterioration. *Artificial Intelligence in Medicine 34*(2), 163–177.

Turing, A. (1950). Computing machinery and intelligence. *Mind 59*, 433–460.

Turney, P. (1995). Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm. *Journal of AI Research*, 369–409.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*, 1124–1131.

Tversky, A. and D. Kahneman (1982). Judgments of and by representativeness. In *Judgment under Uncertainty: Heuristics and Biases*, pp. 84–98. Cambridge.

Twardy, C., A. Nicholson, and K. Korb (2005). Knowledge engineering cardiovascular Bayesian networks from the literature. Technical report TR 2005/170, Clayton School of IT, Monash University.

Twardy, C. R., A. E. Nicholson, K. B. Korb, and J. McNeil (2006). Epidemiological data mining of cardiovascular Bayesian networks. *Electronic Journal of Health Informatics 1*(1), 1–13. http://www.ejhi.net.

Van Berlo, R., E. Van Someren, and M. Reinders (2003). Studying the conditions for learning dynamic Bayesian networks to discover genetic regulatory networks. *Simulation 79*, 689–702.

van der Gaag, L. C. and S. Renooij (2001). Analysing sensitivity data from probabilistic networks. In *UAI01 – Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 530–537.

van der Gaag, L. C., S. Renooij, C. L. M. Witteman, B. M. P. Aleman, and B. G. Taal (1999). How to elicit many probabilities. In *UAI99 – Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 647–654.

van der Gaag, L. C., S. Renooij, C. L. M. Witteman, B. M. P. Aleman, and B. G. Taal (2002). Probabilities for a probabilistic network: A case-study in oesophageal cancer. *Artificial Intelligence in Medicine 25*(2), 123–148.

van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Clarendon Press.

VanLehn, V. and Z. Niu (2001). Bayesian student modeling, user interfaces and feedback: a sensitivity analysis. *International Journal of Artificial Intelligence in Education 12*, 154–184.

Varis, O. (1997). Bayesian decision analysis for environmental and resource management. *Environmental Modeling and Software 12*, 177–185.

Varis, O. and M. Keskinen (2006). Policy analysis for the Tonle Sap lake, Cambodia: A Bayesian network model approach. *International Journal of Water Resources Development 22*(3), 417–431.

Varis, O. and V. Lahtela (2002). Integrated water resources management along the Senegal River: Introducing an analytical framework. *International Journal of Water Resources Development 18*, 501–521.

Venn, J. (1866). *Logic of Chance*.

Verma, T. S. and J. Pearl (1990). Equivalence and synthesis of causal models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, pp. 220–227.

Vicard, P., A. Dawid, J. Mortera, and S. Lauritzen (2008). Estimating mutation rates from paternity casework. *Forensic Science International: Genetics 2*(1), 9–18.

Visscher, S., P. Lucas, C. Schurink, and M. Bonten (2009). Modeling treatment effects in a clinical Bayesian network using Boolean threshold functions. *Artificial Intelligence in Medicine 46*(3), 251–266.

Voie, A., A. Johnsen, A. Strømseng, and K. Longva (2010). Environmental risk assessment of white phosphorus from the use of munitions – a probabilistic approach. *Science of The Total Environment 408*(8), 1833–1841.

von Mises, R. (1919). Grundlage der Warhscheinlichkeitsrechnung. *Mathematische Zeitschrift*.

von Mises, R. (1928/1957). *Probability, Statistics and Truth*. London: Allen and Unwin.

von Neumann, J. and O. Morgenstern (1947). *Theory of Games and Economic Behavior* (second ed.). Princeton, NJ: Princeton University Press.

Von Winterfeldt, D. and W. Edwards (1986). *Decision analysis and behavioral research*. Cambridge University Press.

Wallace, C. and D. Dowe (2000). MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. *Statistics and Computing 10*(1), 73–83.

Wallace, C. S. and D. M. Boulton (1968). An information measure for classification. *The Computer Journal 11*, 185–194.

Wallace, C. S. and D. L. Dowe (1994). Intrinsic classification by MML – the Snob program. In *AI94 – Proceedings of the Seventh Australian Joint Conference on Artificial Intelligence*, Armidale, Australia, pp. 37–44.

Wallace, C. S. and P. R. Freeman (1987). Estimation and inference by compact coding. *Journal of the Royal Statistical Society (Series B) 49*, 240–252.

Wallace, C. S., K. Korb, and H. Dai (1996). Causal discovery via MML. In L. Saitta (Ed.), *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 516–524. Morgan Kaufman.

Wallace, C. S. and K. B. Korb (1999). Learning linear causal models by MML sampling. In A. Gammerman (Ed.), *Causal Models and Intelligent Data Management*. Springer Verlag.

Wang, X., B. Zheng, W. Good, J. King, and Y. Chang (1999). Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. *International Journal of Medical Informatics 54*(2), 115–126.

Waterman, D. A. (1970). A generalization learning technique for automating the learning of heuristics. *Artificial Intelligence 1*, 121–170.

Webb, G., J. Boughton, and Z. Wang (2005). Not so naive Bayes: Aggregating one-dependence estimators. *Machine Learning 58*, 5–24.

Webb, G., F. Zheng, J. Boughton, and K. Ting (2010). Decreasingly naive Bayes: Aggregating n-dependence estimators. *Machine Learning*. Under submission.

Weber, P. and L. Jouffe (2006). Complex system reliability modeling with dynamic Object Oriented Bayesian networks (DOOBN). *Reliability Engineering & System Safety 91*(2), 149–162.

Weber, P., D. Theilliol, C. Aubrun, and A. Evsukoff (2007). Increasing effectiveness of model-based fault diagnosis: A dynamic Bayesian network design for decision making. In *In Proceedings of the Conference on Fault Detection, Supervision and Safety of Technical Processes*, pp. 90–95.

Weidl, G., A. Madsen, and S. Israelson (2005). Applications of object-oriented Bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes. *Computers & chemical engineering 29*(9), 1996–2009.

Wellman, M. (1990). Fundamental concepts of Qualitative Probabilistic Networks. *Artificial Intelligence 44*(3), 257–303.

Wellman, M. P. and C.-L. Liu (1994). State-space abstraction for anytime evaluation of probabilistic networks. In *UAI94 – Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, pp. 567–574.

Wen, Y., K. Korb, and A. Nicholson (2009). Datazapper: Generating incomplete datasets. In *First International Conference on Agents and Artificial Intelligence*, pp. 69–76.

Wermuth, N. (1987). Parametric collapsibility and the lack of moderating effects. *Journal of the Royal Statistical Society B 49*, 353–364.

Weymouth, G., T. Boneh, P. Newham, J. Bally, R. Potts, A. Nicholson, and K. Korb (2007, July). Dealing with uncertainty in fog forecasting for major airports in australia. In *Proceedings of the Fourth International Conference on Fog, Fog Collection and Dew*, Pontificia Universidad Catolica de Chile, Chile, pp. 73–76.

White, A. (2009). *Modeling the Impact of Climate Change on Peatlands in the Bogong High Plains*. Ph. D. thesis, University of Melbourne.

Widarsson, B. and E. Dotzauer (2008). Bayesian network-based early-warning for leakage in recovery boilers. *Applied Thermal Engineering 28*(7), 754–760.

Wikipedia (2010). Conceptual model. http://en.wikipedia.org/wiki/Conceptual_model_(computer_science).

Williamson, J. (2001). Foundations for Bayesian networks. In D. Corfield and J. Williamson (Eds.), *Foundations of Bayesianism*. Dordrecht: Kluwer.

Williamson, J. (2005). *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford: Oxford University.

Willis, D. (2000). Ambulation monitoring and fall detection system using dynamic belief networks. BCSE Thesis report. Department of Electrical and Computer Systems Engineering, Monash University.

Wilson, D., M. Stoddard, and K. Puettmann (2008). Monitoring amphibian populations with incomplete survey information using a Bayesian probabilistic model. *Ecological Modeling 214*(2-4), 210–218.

Winston, P. H. (1977). *Artificial Intelligence*. Reading, MA: Addison-Wesley.

Witten, I. and E. Frank (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). Morgan Kaufmann. http://www.cs.waikato.ac.nz/˜ml/index.html.

Wong, M., W. Lam, and K. Leung (1999). Using evolutionary programming and Minimum Description Length principle for data mining of Bayesian networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 174–178.

Woodberry, O., A. Nicholson, K. Korb, and C. Pollino (2004a). A methodology for parameterising Bayesian networks. Technical report TR 2004/159, School of Computer Science and Software Engineering, Monash University.

Woodberry, O., A. Nicholson, K. Korb, and C. Pollino (2004b). Parameterising Bayesian networks. In G. Webb and X. Yu (Eds.), *Lecture Notes in Artificial Intelligence, (Proceedings of the 17th Australian Joint Conference on Advances in Artificial Intelligence [AI'04], Cairns, Australia, 4-6 December 2004)*, Volume 3339 of *LNCS/LNAI Series*, pp. 1101–1107. Berlin, Germany: Springer Verlag.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research 20*, 557–585.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics 5*, 161–215.

Xenos, M. (2004). Prediction and assessment of student behaviour in open and distance education in computers using Bayesian networks. *Computers & Education 43*(4), 345–359.

Zadora, G. (2009). Evaluation of evidence value of glass fragments by likelihood ratio and Bayesian network approaches. *Analytica Chimica Acta 642*(1-2), 279–290.

Zhou, H. and S. Sakane (2007). Mobile robot localization using active sensing based on Bayesian network inference. *Robotics and Autonomous Systems 55*(4), 292–305.

Zonneveldt, S., K. Korb, and A. Nicholson (2010). Bayesian network classifiers for the German credit data. Technical report, 2010/1, Bayesian Intelligence. http://www.Bayesian-intelligence.com/publications.php.

Zukerman, I., R. McConachy, and K. B. Korb (1996). Perambulations on the way to an architecture for a nice argument generator. In *ECAI'96 Workshop. Gaps and Bridges: New Directions in Planning and Natural Language Generation*, Budapest, Hungary, pp. 31–36.

Zukerman, I., R. McConachy, and K. B. Korb (1998). Bayesian reasoning in an abductive mechanism for argument generation and analysis. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, Madison, Wisconsin, pp. 833–838.

Zweig, G. (2003). Bayesian network structures and inference techniques for automatic speech recognition. *Computer Speech & Language 17*(2-3), 173–193.