*CHAPTER 1*

# INTRODUCTION

**by**
**Glenn Shafer and Judea Pearl**

Most everyday reasoning and decision making is based on uncertain premises. Most of our actions are based on guesses, often requiring explicit weighing of conflicting evidence. The readings in this volume are concerned with methods that have been used in artificial intelligence to manage uncertainty and conflicting evidence.

There are many approaches to managing uncertainty in AI. In addition to the mathematical theory of probability, there are other numerical calculi for the explicit representation of uncertainty, such as MYCIN's certainty factors and Zadeh's possibility measures. (Dempster-Shafer theory is sometimes mentioned in this connection, but we take the view that this theory is merely an alternative way to use probability.) There are proposals to manage uncertainty using modal logic or other symbolic formalisms. And there are ways of designing the architecture of reasoning systems to take uncertainty into account even without explicitly representing it.

In selecting the readings in this volume, we have tried to represent all these approaches, but we have emphasized probability. Historically, the mathematical theory of probability has provided the most successful approach to the formalization of reasoning under uncertainty in a wide variety of fields, from engineering to economics. During the 1960s and 1970s, relatively little probability was used in AI, but during the 1980s, it received increasing attention.

In spite of this increasing attention, much uncertainty and confusion remains about the role of probability in AI. It is not always clear how the traditional assumptions made in applying probability can be reconciled with the realities of computer programming or the realities of the problems that concern AI. It is not clear how the concepts of probability are related to the intuitions about uncertainty and plausible reasoning that have been developed in AI over the last thirty years. And it is not clear how probability theory can be related to the other formalisms that AI has used, especially symbolic logic.

We hope that the articles in this volume will help clarify these questions. Some of these articles bring into the AI literature for the first time debates about the meaning of probability that need to be taken into account in order for the potential role of probability in AI to be fairly assessed. Others point to interesting connections between probability and other approaches. The majority of the articles—26 out of 47—are primarily concerned with probabilistic approaches—Bayesian, frequentist, and Dempster-Shafer. Of the 21 other articles, 10 are primarily concerned with non-probabilistic approaches, and 11 compare or attempt to integrate probabilistic and non-probabilistic approaches.

Our emphasis on probability is due in part to our own interests. We are both proponents of probability. Judea Pearl is well known as a proponent of Bayesian approaches in AI, and Glenn Shafer is one of the founders of the Dempster-Shafer theory. We do not want to overstate these biases, however, because not all problems of uncertainty in AI lend themselves to probability. Other approaches are often required, and we hope that the articles in this volume will be informative and useful even to those who are working with problems that do require other approaches.

In the remainder of this introduction, we will review the role of probability in AI in more detail, and we will discuss the purposes of each of the chapters that follow.

### *The Emergence of Probability Theory in Artificial Intelligence*

A few scholars have always advocated a role for probability in AI. Richard Duda, Peter Hart, and Nils Nilsson are perhaps the best known early AI researchers to develop important applications of probability, first in pattern recognition and learning (Duda and Hart 1973, Nilsson 1965), and then in expert systems (see their article in Chapter 5). Others have advocated a wider role of probability in AI; Ray Solomonoff, for example, has long argued that AI should be based in general on the use of algorithmic probability to learn from experience (Solomonoff 1986). Most of the formal work in AI before the 1980s, however, was based on symbolic logic rather than probability theory.

There are obvious reasons for this disinterest in probability. Probabilities are numbers, and number crunching is just what AI was supposed not to be. In the 1950s, computers were used mainly for number crunching. They were impressively good at this, but they were not able to explain what they were doing. Thus they were not intelligent. Artificial intelligence required going beyond number crunching to symbolic manipulation.

It should also be noted that the dominant interpretation of probability in the 1950s, in almost all applied fields, was frequentist. Probability was seen, even by its proponents, as a technique that was appropriate only when statistical data were available.

In this environment, symbolic logic seemed to provide the only appropriate formal mathematics for AI. Not everyone found symbolic logic to be useful, but those who did not emphasized ad hoc programming rather than alternative formalisms. To a large extent, the field was divided between the logicists, whose archetype was John McCarthy, and non-logicists, whose archetype was Marvin Minsky (Israel 1983).

Probability came back into AI on the practical side in the 1970s, when attempts were made to apply AI ideas to expert systems for large-scale practical problems, such as vision, speech recognition, medical diagnosis, and geological exploration. Numerical scoring of one kind or another seems to be inevitable in such problems, both as a device to weigh conflicting arguments and as an aid to selecting the most promising from many possible lines of reasoning.

This re-emergence of probability in AI was aided by philosophical changes in the interpretation of probability. By the 1970s, frequentism was no longer the only respectable interpretation of probability. It was being contested by a resurgent subjective interpretation, which opened the possibility of using probabilities even when frequency data was not available.

The expert systems that used probability in the 1970s paid relatively little attention to the theory of probability. They used numbers to score the strength of evidence, but they did not use ideas from probability theory to organize inference. The primary organizing idea was still symbolic logic.

Experience with the systems of the 1970s led, however, to deeper interest in probability. As we explain in the introduction to Chapter 6, it turned out to be dangerous to implement the certainty factors of MYCIN and the likelihood propagation ideas of PROSPECTOR in general inference engines. Probabilities could not simply tag along as numbers attached to if-then rules. The results of probability calculations would be sensible only if these calculations followed principles from probability theory.

This deeper interest in probability has borne fruit in several ways. On the practical side, we now have a number of practical systems that use probability ideas, especially in medicine. These include the HUGIN system, described in Chapter 6, and the Gister and DELIEF systems, described in Chapter 7. On the scholarly side, we see numerous articles on probability in AI journals, and probability is flourishing not only in standard AI forums but also in an annual Uncertainty in AI Workshop (see Kanal and Lemmer 1986, Lemmer and Kanal 1988, Kanal et al. 1989, Shachter et al. 1990).

In spite of this high level of activity, there is still no consensus on the theoretical and practical role of probability in AI. There are many reasons for this. In many cases, we can afford to ignore uncertainties. In other cases, probability ideas can be used implicitly rather than explicitly. Most importantly, perhaps, the thirty-year momentum of logic as the overriding formal framework of AI makes it difficult for probability to be accepted fully without a consensus on how probability and logic should fit together.

### *What Can Artificial Intelligence Learn from Probability Theory?*

As we have just noted, the experience of the 1970s forced AI to learn from probability theory about the structure of plausible reasoning. The modularity of if-then rules contributed to the success of the expert systems of the late 1970s and early 1980s (Davis and King 1984), but this modularity could not be sustained in problems where exceptions were too important to ignore and too numerous to list explicitly. Unlike demonstrative reasoning, which produces correct results no matter what path deductions take and no matter how often they are repeated, plausible reasoning must be structured so that inferences are made in an appropriate order and so that repetition of an inference is not interpreted as corroboration.

In terms of probability theory, this need for structure can be explained in terms of conditional independence. The propagation of probability in the course of inference is legitimate if the propagation follows separate paths in a tree where separation represents conditional independence. This point is elaborated in a number of articles in this volume, including the article by Judea Pearl, Dan Geiger, and Thomas Verma in Chapter 2, the article by David Heckerman in Chapter 5, and the articles by Pearl and by Steffen Lauritzen and David Spiegelhalter in Chapter 6.

Once this concept of probabilistic structure became clear, it also became clear that it was already being used, at an intuitive level, in AI systems. In other words, this aspect of probability formalizes actual practice.

This can be illustrated using the MUM architecture for the diagnosis of angina, which is described in the article by Paul Cohen in Chapter 4. Figure 1 is a very simplified depiction of the structure of the inferences made following a MUM workup. As this figure indicates, a physician working with MUM combines evidence from a physical examination and a health history to get a judgment about how much at risk of heart disease the patient is, and then he or she combines this with the patient's description of an apparent angina episode to get a judgment about whether the patient really has angina.

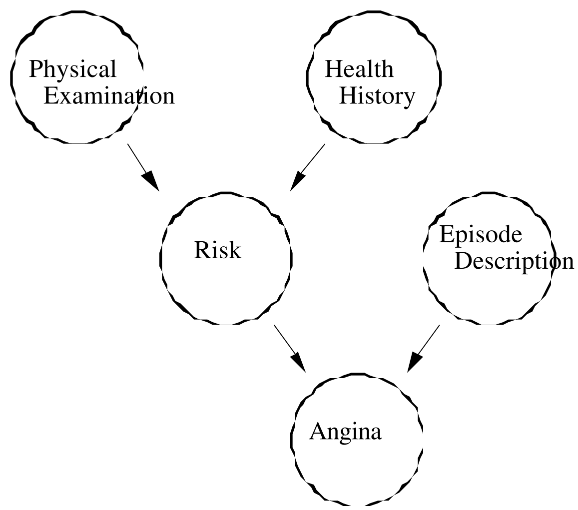


**Figure 1. A Simplification of a MUM workup.**

This structure of Figure 1 can be explained in terms of probability theory by saying that Physical Examination and Health History are conditionally independent of Episode Description given Risk. It is this conditional independence that justifies forgetting about the details of Physical Examination and Health History when combining what they told us about Risk with the information from Episode Description.

Such intuitive exploitation of conditional independence structures appears throughout the AI literature on managing uncertainty. We will find it repeatedly in the articles in Chapters 4 and 8, in systems that do not explicitly use probability ideas or even numbers.

Once examples like this one are understood, it becomes clear that probability theory is more fundamentally concerned with the structure of reasoning and causation than with numbers. The combination of evidence in MUM does not use numbers; instead it uses rules for combining verbal descriptions of evidence strength. Yet the structure is probabilistic.

The idea of probabilistic structure is also relevant to McCarthy and Hayes's famous frame problem, discussed by Raymond Reiter in section 6.3.2 of his article in Chapter 8. As Reiter explains, it is an overwhelming task to list as explicit frame axioms all the irrelevancies in a situation—all the things that should not be directly affected by a given action or change. But as Pearl, Geiger, and Verma explain in their article in Chapter 1, these irrelevancies are implicit rather than explicit in the probabilistic approach; they are represented by missing links in the conditional-independence structure.

Artificial intelligence may also be able to profit from the debate about the interpretation of probability that has been conducted in mathematical statistics, philosophy, and other fields. During the past thirty years, this debate has led to the development of Bayesian and other subjective approaches to probability that are quite distinct from the frequentist approach. Yet much of the discussion of probability in AI continues to assume the frequentist interpretation that was dominant in the 1950s. In his article in Chapter 8, for example, Reiter dismisses probabilistic approaches by arguing that adequate frequency data is usually not available and that even when it is, typicality differs from frequency.

### *What Can Probability Learn from Artificial Intelligence?*

That structure is more important than numbers in probability needs to be taught to all users of probability, not just to workers in AI. The lesson is familiar, in some sense, to users of complex statistical models and to workers in applied probability, where the greatest effort goes into analyzing structure. Unfortunately, it is not always evident to research workers new to probability.

The fact that this intimate connection between probability and structure can be overlooked is an indictment of the way probability is taught. Most teaching in probability and statistics is relatively abstract. Most introductions to probability are patterned after what mathematicians call measure theory. Statistical inference is taught using only the simplest models, so that the logic and mathematics of inference can be emphasized. More complicated and structured models are left to advanced courses in stochastic processes or specialized courses in statistics, such as path analysis.

The situation is at its worse in the teaching of subjective probability, which should be most relevant to AI. In a course of study emphasizing objective probability, one comes relatively quickly to stochastic processes and other models of applied probability. But because the revival of subjective probability in the past thirty years has involved so much philosophical controversy, the teaching of subjective probability tends to emphasize axiomatizations and to avoid the study of complicated structure in favor of exhaustive analysis of simple models, where the differences between Bayesian and frequentist reasoning are clearest.

Fortunately, we can list a number of developments that promise to bring more emphasis on structure into the introductory teaching of probability:

- The axiomatic study of conditional independence, pioneered by Judea Pearl and summarized in the article by Pearl, Geiger, and Verma in Chapter 2.
- The constructive interpretation of probability, discussed further in the article by Shafer and Tversky in Chapter 2.
- The evolution, within decision analysis, from decision trees to influence diagrams. This is discussed in the introduction to Chapter 3 of this volume, and in the article by Shachter reprinted there.
- The revival of structural approaches to the definition of conditional probability. See Shafer (1985).
- The work on local computation in probability networks. This work is described in Chapters 5, 6, and 7 of these readings and in Pearl (1988).

These developments should also help bring the relevance of probability to AI more into focus in the years to come.

Beyond the increased emphasis on structure that the AI experience has brought to probability, probability may also benefit from the debate in AI over the semantics of knowledge representation. As Terry Winograd explains in his article in Chapter 8, logicists in AI have tried to retain the model-theoretic semantics of the predicate calculus, while non-logicists, such as Winograd himself, believe that many useful statements have no meaning "in any semantic system that fails to deal explicitly with the reasoning process." Parallels have arisen in probability. On the one hand, most frequentists and most subjectivists have insisted, like the logicists, on process-free semantics; the frequentists have insisted on a semantics in terms of populations, while the subjectivists have insisted on a semantics in terms of willingness to bet. Yet more process-oriented approaches have repeatedly appeared. Jerzy Neyman was famous within frequentist statistics for his insistence on a process-oriented interpretation of tests and confidence intervals (Neyman 1967), and the constructive interpretation of probability advanced by Shafer and Tversky in their article in Chapter 2 is also process-oriented. A clearer understanding of the AI debate may clarify these issues. The more deeply we understand the processes needed for various reasoning tasks, the better we can formulate process-oriented semantics for the very notion of probability.

Probability may also profit from the attention that AI has given to the structure of convention in human communication. For example, having been told that As are typically Cs and Bs are typically Cs, we tend to presume that As which are also Bs are typically Cs. This convention is useful in knowledge elicitation, and it is embedded in almost every non-monotonic logic. Yet it is not directly expressible in terms of conditional independence or its graphical representation. This point is discussed further in Chapter 9.

In the long run, the most important contribution of AI to probability should be in forcing more serious consideration of the problem of model construction. Discussions of probability theory have traditionally started with the assumption that we have a probability structure—and even probability numbers. This is not the starting point for AI, and as we learn more from attempts to implement probability in AI, it cannot remain the starting point for probability theory in general.

### *An Overview of the Readings*

The unsettled nature of current thinking about the management of uncertainty in AI is reflected in the structure of this volume. The eight chapters into which we have grouped the articles are coherent but relatively independent.

We have placed more general topics first, in the hope that this will be helpful to most readers. The general background on probability provided by Chapters 2 and 3 should help readers understand the more specific topics covered in Chapters 5, 6, and 7. And the earlier chapters on probability will certainly be useful in understanding the attempts to integrate probability and logic covered in Chapter 9.

This organization may not serve all readers equally well, especially as some topics appear in more than one chapter. Depending on their interests, readers may profitably begin with any chapter. And some articles in later chapters—the first article by Judea Pearl in Chapter 6, for example—may be helpful in understanding earlier chapters.

Chapter 2 is designed to bring to AI what other disciplines have learned about the meaning of probability. Some of the articles in this chapter cover the debate between frequentists and subjectivists in mathematical statistics, as well as the more recent debate about the descriptive and normative aspects of the theory. Other articles explain what is meant by saying that probability is constructive and that probability is more about structure than about numbers.

Chapter 3 is concerned with using probabilities to make decisions. The emphasis is on Bayesian decision theory, its relation to human practice, and its implementation in AI.

Chapter 4 is concerned with the role of system architecture in managing uncertainty in AI. As we will see in this chapter, there is more to the control of uncertain inference than conditional-independence structures, even in systems that explicitly use probability.

Chapter 5 reports on how probabilities and related numerical methods have actually been used in expert systems. The emphasis is on pioneer articles—articles that describe early systems such as MYCIN, PROSPECTOR, PIP, INTERNIST, and CASNET. We also include an article on HUGIN, a recent Bayesian system.

Chapters 6, 7, and 8 present competing theories for the explicit representation and handling of uncertainty. Chapters 6 and 7 present two probabilistic theories—the Bayesian theory and the Dempster-Shafer theory. Chapter 8 presents several non-numerical approaches.

Chapter 9 presents a number of ideas about how we might integrate probability and logic. Although the articles in this chapter take rather disparate approaches, they together point towards a possible consensus, in which qualitative relationships can be seen as abstractions from probability ideas, and numerical probabilities can be seen as supplementary to traditionally symbolic systems.

### *References*


Buchanan, Bruce G., and Edward H. Shortliffe (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley: Reading, Massachusetts.

Davis, Randall, and Jonathan J. King (1984). The origin of rule-based systems in AI. In Buchanan and Shortliffe (1984), pp. 20-52.

Duda, Richard O., and Peter E. Hart (1973). *Pattern Classification and Scene Analysis*. Wiley: New York.

Israel, David (1983). The role of logic in knowledge representation. *IEEE Computer* October 1983, pp. 37-41.

Kanal, Laveen N., and John F. Lemmer, eds. (1986). *Uncertainty in Artificial Intelligence*. North-Holland: Amsterdam.

Kanal, Laveen N., Tod S. Levitt, and John F. Lemmer, eds. (1989) *Uncertainty in Artificial Intelligence 3*. North-Holland: Amsterdam.

Lemmer, John F., and Laveen N. Kanal, eds. (1988). *Uncertainty in Artificial Intelligence 2*. North-Holland: Amsterdam.

Neyman, Jerzy (1967). *A Selection of Early Statistical Papers of J. Neyman*. Cambridge University Press.

Nilsson, Nils J. (1965). *Learning Machines*. New York: McGraw-Hill.

Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.

Shachter, Ross, Tod S. Levitt, John F. Lemmer, and Laveen N. Kanal, eds. (1990). *Uncertainty in Artificial Intelligence 4*. North-Holland: Amsterdam.

Shafer, Glenn (1985). Conditional probability. *International Statistical Review* **53** 261-277.

Solomonoff, Ray (1986). The application of algorithmic probability to problems in artificial intelligence. In Kanal and Lemmer (1986), pp. 473-491.