Intelligent Information Systems V
Proceedings of the Workshop held in
Dęblin, Poland, June 2-5, 1996

pages 12- 24

Explanation in Probabilistic Systems: Is It Feasible? Will It Work?

Marek J. Druzdzel

University of Pittsburgh
Department of Information Science
and Intelligent Systems Program
Pittsburgh, PA 15260
marek@lis.pitt.edu

Abstract. Reasoning within such domains as engineering, science, management, or medicine is traditionally based on formal methods employing probabilistic treatment of uncertainty. It seems natural to base artificial reasoning systems in these domains on the normative foundations of probability theory. Two usual objections to this approach are (1) probabilistic inference is computationally intractable in the worst case, and (2) probability theory is incomprehensible for humans and, hence, probabilistic systems may be hardly usable. The first objection has been addressed effectively in the last decade by a variety of efficient exact and approximate schemes for probabilistic reasoning, applied in several practical systems. In this paper, I review the state of the art with respect to the second objection.
First I argue that the observed discrepancies between human and probabilistic reasoning and the anticipated difficulties in building user interfaces are not a good reason for rejecting probability theory. On the contrary — they provide motivation for a normative treatment of uncertainty. I point out that probability theory rests on qualitative foundations that capture essential properties of a domain along with such concepts such as relevance and conflicting evidence. In addition, graphical probabilistic models, as opposed to rule-based systems, integrate numerical and structural properties of a domain and provide a natural representation of causality. Finally, availability of a full quantitative specification of a model allows for manipulating the level of precision for both reasoning and explanation.

Keywords: reasoning under uncertainty, user interfaces, explanation

1 INTRODUCTION

The purpose of computing is insight, not numbers.
Richard Wesley Hamming

Reasoning within such domains as engineering, science, management, or medicine is traditionally based on formal, mathematical methods employing probabilistic treatment of uncertainty. These methods have proven over time to lead to results of predictable reliability and have become integral parts of both teaching and practice in each of the fields. It seems natural to base artificial reasoning systems in these domains on the normative foundations of probability theory and decision theory. Two usual objections to this approach are (1) probabilistic inference is computationally intractable in the worst case, and (2) probability theory is incomprehensible for humans and, hence, probabilistic systems may be hardly usable. The first objection has been addressed effectively in the last decade by a variety of efficient exact and approximate schemes for probabilistic reasoning, applied in several practical systems. In addition, theoretical studies identified major incompatibilities between rule-based systems and the task of reasoning under uncertainty [24]. Designing effective user interfaces to probabilistic systems remains a major research challenge.

The need for effective user interfaces is especially clear in the domain of decision support systems, where Hamming's quote (above) readily paraphrases into *The purpose of decision support is insight, not advice.* While probabilistic systems can be expected to give better quality advice under uncertainty than heuristic systems,[1] their impact on actual decisions will depend on how successful they are in increasing users' insight into these decisions. Without understanding, the users may accept or reject system's advice for wrong reasons and the combined decision making performance may deteriorate even below unaided performance [21]. System's justification of its advice is also important because of ethical or legal constraints on the user — a physician using a diagnostic system is not supposed to follow an oracle advice blindly, even if the system has shown reliable performance over time. In addition, a system based on normative principles can play a tutoring role — one might hope that its user will learn the domain model and how to reason with it over time and improve his or her own thinking.

Probabilistic reasoning is often hard to comprehend for humans. Even in single step Bayesian updating, the magnitude of change in belief and sometimes even its direction may be experienced counterintuitive. Humans do not appear to think normatively, to maximize, or to minimize the way normative theories do. Are there any reasons at all to hope that a normative system can make its advice comprehensible? Questions as above have understandably worried AI researchers for some time (e.g., [2, 28]). In this paper, I argue that although the state of the art in explanation of probabilistic inference perhaps does not yet match that of explanations for rule-based systems, there are good reasons

[1] I have encountered some disagreement on this issue among AI researchers. I invite those readers who are not convinced to go through a mental (!) exercise of flying an airplane that has been designed using heuristic methods rather than the laws of physics enhanced by probabilistic reliability assessment.

for optimism.

The first part of this paper argues against rejection of probability theory on the grounds that it is remote from human reasoning under uncertainty. It reviews the available empirical evidence concerning human judgment under uncertainty and, to put things in a proper perspective, it reports empirical evidence for discrepancies between human reasoning and first order logic. I argue that the evidence in both cases does not support rejecting the formalisms, but rather is a strong argument for supporting unaided human intuition. A prudent response to our demonstrable weakness in mental arithmetics has been not rejection of elementary algebra but instead use of aids such as calculators. The second part of the paper shows that probabilistic representations offer several important advantages over rule-based systems with respect to building user interfaces. It shows that probability theory allows for capturing the most essential qualitative properties of a domain, along with its causal structure. Independence and relevance are represented in an efficient and natural way. Availability of a full quantitative specification of a model allows for manipulating the level of precision both for reasoning and for explanations.

2 HUMAN REASONING UNDER UNCERTAINTY

Human judgment under uncertainty has been a major focus of behavioral research in the last quarter of the century. A central motivation for this work has been the desire to compare human judgment against the "gold standard" of the normative probability theory and decision theory in order to identify places where human decision making could be improved. Numerous studies have shown that people usually rely on rules of thumb, or *heuristics*, in their estimations of the likelihood of uncertain events. These heuristics make estimation of probabilities computationally tractable, but often result in systematic biases and inconsistencies.

A concise summary of the position that is known today as the "heuristics and biases" school of thought is given by Tversky and Kahneman:

> The subjective assessment of probability resembles the subjective assessment of physical quantities such as distance or size. These judgments are all based on data of limited validity, which are processed according to heuristic rules. For example, the apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. This rule has some validity, because in any given scene the more distant objects are seen less sharply than nearer objects. However, the reliance on this rule leads to systematic errors in the estimation of distance. Specifically, distances are often overestimated when visibility is poor because the contours of objects are blurred. On the other hand, distances are often underestimated when visibility is good because the objects are seen sharply. Thus, the reliance on clarity as an indication of distance leads to common biases. Such biases are also found in the intuitive judgment of probability. [19, page 3]

It is these biases that are believed to provide the most information about cognitive processes used in judgment. Designing experiments that lead to errors in human reasoning and observing the nature of these errors allows to infer the heuristic procedures that most likely led to them.

One of such heuristics is *representativeness*[2], which amounts to judging representative instances of a category to be more probable than unrepresentative instances. Another is *availability* of examples of the judged events in memory. Availability to memory depends of course on the probability, as this usually affects availability. However, as availability is also affected by recency and emotional impact, it is not always a good guide to an objective value of probability. Third well studied heuristic is *anchoring and adjustment*, which amounts to estimating probability of an event by starting from an initial value (e.g., known probability of a similar event) and adjusting that value to yield the final answer. This procedure is not reliable, as adjustments are typically insufficient and different starting points, evoked for example by the framing of the problem, lead to different estimates. Another heuristic, labeled by Kahneman and Tversky *simulation heuristic* (this is also known as *scenario thinking* [6]), amounts to generation of plausible stories and using the ease with which they come to mind in judging the probability of the component events.

Use of heuristics often leads to inconsistencies and systematic errors. One such systematic error is known as *conjunction fallacy*, which is a tendency to judge a conjunction of an unusual and a likely event to be more likely than the unusual event alone. Other violations are neglecting the effect of the base rate in probability updates (*base-rate fallacy*) and judging small sample sizes to be equally informative as large sample sizes (*insensitivity to sample size*). Yet other is *misconception of chance*, e.g., judging sequences of outcomes that look random more likely than sequences that appear less random (e.g., the sequence of coin tosses *HTTHTH*, is judged to be more likely than the sequence *HHHTTT*). Other documented biases are *insensitivity to predictability*, *illusion of validity*, and *misconceptions of regression*.

Overall, although one might argue about the extent of the errors, validity of the laboratory results in real life, and the conclusions with respect to human rationality, it seems rather clear that human decision makers depart from the idealized rationality of normative theory and rely on a variety of heuristics that reduce the cognitive load at the expense of optimal decision making. It is not only naive subjects who demonstrate apparent errors and inconsistencies in judgment. Professionals, such as practicing physicians, have been shown to use essentially the same judgmental heuristics and be prone to the same biases, although here the degree of departure from the normatively prescribed judgment seems to decrease with experience (e.g., physicians in real diagnostic context have been observed to take base rates into consideration in their judgment of probability of a disease in their patients [4]).

The behavioral decision theory literature provides us with detailed information about some aspects of human reasoning under uncertainty and, in particular, the discrepancies between human reasoning and probability theory. Proponents of alternative approaches to uncertain reasoning have sometimes used these findings as a justification for their formalisms on the grounds that they provide better models of human reasoning than

[2] Most of this section has a rather concise character. Formal discussion of the most important research results along with experimental data can be found in an anthology edited by Kahneman, Slovic, and Tversky [19]. Another excellent source is an accessible book by Dawes [6].

probability theory. However, despite common claims that these models are compatible with the way people think, there is little rigorous empirical evidence that this is the case, as none of the other formalisms have been subjected to as extensive experimental investigation as probability theory. The findings of the "heuristics and biases" literature can be viewed as numerous valuable data points guiding the work on user interfaces. This is, in fact, how these findings are viewed in behavioral decision theory literature: they are not a reason to question the validity of probability theory and decision theory for decision support but rather a source of valuable hints how to improve human intuitive judgment. These findings have had a large impact on the actual decision-analytic practice and led to many improvements of the procedures used in knowledge elicitation.

3 SHOULD NORMATIVE THEORIES BE REJECTED?

Given the empirical evidence concerning human reasoning under uncertainty, a question that has been asked repeatedly, is whether probability theory should be rejected on the grounds of incompatibility with human reasoning. I will argue that mere incompatibility of normative theories with human reasoning are not sufficient grounds for rejecting them. There is little as established in artificial intelligence research as logic. Still, one easily forgets that there are considerable discrepancies between human reasoning and logic. Wason and Johnson-Laird [29] have investigated human performance on various isomorphs of the following task. The experimenter lays out in front of a subject four cards displaying the following symbols:

$$\boxed{\text{E}} \quad \boxed{\text{K}} \quad \boxed{4} \quad \boxed{7}$$

The subject is told that each card has a number on one side and a letter on the other side. The task is to select those cards that need to be turned over in order to find out whether the rule "if a card has a vowel on one side then it has an even number on the other side" is true or false. Although the problem is easy to understand, it seems to be hard for most subjects to solve: although most subjects turn over the card with a vowel, some decide to turn over the card with an even number, and only around 12% of subjects are reported to have realized the need to turn over the card bearing an odd number.[3] Some isomorphs of this problem, for example cards

[3] The correct answer consists of turning over the cards $\boxed{\text{E}}$ and $\boxed{7}$.

Manchester	*Sheffield*	*Train*	*Car*

and the rule "every time I go to Manchester I travel by train," are much easier for subjects to solve. The authors report that as many as 60% subjects turn over correctly the card $\boxed{\mathit{Car}}$. Logically, this choice is equivalent to the choice of the card bearing an odd number, which was turned over by only 12% of the subjects.

Results of these, and other experiments[4] suggest strongly that human reasoning is not based on abstract, domain independent rules, but rather depends heavily on the content. Logic, which provides domain independent rules, does not seem to be a good descriptive formalism for human reasoning. Despite this, few would find the above results a reason for rejecting logic as a basis for some types of operations in artificial reasoning schemes. Difficulty that humans experience with probabilistic tasks is not a compelling reason for rejecting probability theory. The "heuristics and biases" research shows exactly the possible dangers of replicating human behavior — along with replicating computationally efficient heuristics, we might replicate the undesirable judgmental biases. The goal of a decision support system is supporting unaided human intuition, just as the goal of using a calculator is aiding human's limited capacity for mental arithmetic. It is more appropriate to view human biases and inconsistencies not as problems of probability theory but perhaps as sources of motivation for normative decision support.

4 HUMANS AS QUALITATIVE BAYESIANS

While humans have been shown to err in their quantitative judgments, incompatibility between qualitative aspects of reasoning under uncertainty and qualitative probabilistic statements are harder to find. In fact, there are several common reasoning patterns that seem to be modeled well by probability theory and are troublesome for other formalisms. Henrion [16] lists several such patterns and among them human ability to mix predictive and diagnostic reasoning, ability to discount correlated sources of evidence, and intercausal reasoning. And so, presence of predictive and diagnostic rules in the same knowledge base introduces the problem of possible infinite inference cycles. For example, higher likelihood of fever increases the likelihood of flu and higher likelihood of flu, in turn, increases the likelihood of fever. It appears that this problem cannot be solved elegantly at a purely declarative level and requires rules whose function is control of reasoning and not domain description. Probabilistic models, such as those based on Bayesian belief networks, do not suffer from this problem and can easily mix predictive and diagnostic inference. Rule-based systems have also a hard time with modeling dependent sources of evidence. Recall hearing radio broadcasts about thousands of people dead

[4] For example, a study by Kotovsky, Hayes and Simon [20] reports dramatic performance differences among different isomorphs of Tower of Hanoi puzzle.

in Chernobyl, shortly thereafter watching TV broadcasts and reading newspaper articles that stated the same. These tree apparently independent sources of news will normally strongly increase our belief that the news is true. When we learn subsequently that each of the media stories is based on a single telephone conversation with a western journalist located in Moscow (thus not even in Ukraine, where Chernobyl is located!), we easily discount the evidence. This reasoning pattern is quite troublesome for systems based on rules enhanced with measures of uncertainty. Intercausal reasoning occurs when several possible causes of an observed common effect become dependent. Suppose we observe fever. This may be caused by flu, appendicitis, cholera, and many other causes. Additional evidence for flu makes all these other caused less likely: flu explains the fever and, hence, explains away all its other causes. Intercausal reasoning has received a considerable interest and has been applied in qualitative reasoning schemes and in qualitative verbal explanations of reasoning [8, 11, 12, 17].

One more empirical finding about human judgment under uncertainty finds surprising theoretical support in probability theory. In my earlier work [7], I conducted a series of empirical studies involving analysis of verbal protocols of subjects solving simple belief updating problems under uncertainty. Several of the subjects exhibited what had been earlier known as scenario thinking — the subjects considered only selected, most probable scenarios involving an outcome or its complement and seemed to weight these to judge the likelihood of that outcome. This turns out to be not only compatible with the normative approach, but constitutes an interesting approximate algorithm for belief updating. Belief updating amounts simply to summation over different possible scenarios and relative weighting of the different sums. The only departure from an exact algorithm is that the subjects considered not more than a few scenarios, a tiny fraction of the number of possible scenarios in a complex problem. Later [9] I demonstrated theoretically that the probability of any scenario within a model often can be seen as drawn from a highly skewed lognormal distribution. Effectively, usually a small fraction of all scenarios covers most of the probability space, while the remaining scenarios are extremely unlikely. I verified this theoretical prediction studying the properties of subsets of a medical model consisting of 37 random variables. It was not unusual to observe a single scenario covering as much as 60% of the total probability state. What it practically means is that despite uncertainty, there is at any point usually a few very likely states of the model. These states explain for all practical purposes almost all uncertainty and, given people's propensity to scenario thinking, can be used for explanations. This is, in fact, the main idea behind scenario-based explanations that Henrion and I [8, 17] proposed. The approaches based on the Most Probable Explanation paradigm introduced by Pearl [24, Section 5.6] and developed further by Charniak and Shimony [3] explore the same property of uncertain models and involve searching for the most probable deterministic state of the world that is compatible with the observed evidence.

The qualitative character of human knowledge can also be explored at the other end of human interface to probabilistic systems, i.e., in model building. Most qualitative statements about prior and conditional probabilities, including relations among them, can be interpreted as constraints on the underlying joint probability distribution and used in the process of probability elicitation. Van der Gaag and I [15] have recently shown how to interpret uncertain knowledge expressed in qualitative or fuzzy terms in probabilistic framework for the purpose of knowledge elicitation.

5 PROBABILITY, CAUSALITY, AND EXPLANATION

It seems to be an accepted view in psychology that humans attempt to achieve a coherent interpretation of the events that they observe and that they often organize their knowledge in schemas consisting of cause-effect relations. Despite the lack of a widely agreed upon formal definition of what causality is, people do have strong intuitions concerning presence or absence of cause–effect relations in particular instances. People seem to use systematic rules and strategies for assessing causality, both in science and everyday life. The irresistible tendency to perceive sequences of events in terms of causal relations, even if the subject is fully aware that the relation between the events is incidental, has been demonstrated in a multitude of situations.

Quite understandably, experiments with early decision support systems, such as Mycin, have indicated that rules alone are not sufficient for generating understandable explanations.[5]. Swartout [27] proposes equipping programs with a domain model and representation of problem-solving control strategies for the purpose of explanation. The need for a domain model rises from the inability of traditional rule-based systems to explicitly examine their control strategy and their behavior and their inability to justify system's rules.

In probability theory, similarly to most mathematics, existence of cause and effect relations between variables is immaterial: all probabilistic dependencies are expressed by conditional probabilities. And so, $Pr(A|B)$ denotes the probability of A given that B has been observed. This can be easily reversed by Bayes theorem to yield $Pr(B|A)$. It is worth noting that rules are perfectly as well. A rule "if A than B" is equivalent to a rule "if not B than not A". Human representation of uncertainty, on the other hand, seems to be close to the representation of causality. People have been observed to use their knowledge about causal relations in probability judgments, especially in judgment of conditional probabilities, confusing the purely informative character of conditioning with causality.

Scarcity of references to causality in most statistics textbooks and the disclaimers that usually surround the term "causation" might create the impression that causality forms a negative and unnecessary ballast on human mind that cannot be reconciled with the probabilistic approach. In fact, *causality and probability are closely related.* While probabilistic relations indeed do not imply causality, causality normally implies a pattern of probabilistic interdependencies and these, in turn, provide clues about causality. In fact, a generally accepted necessary condition for causality is statistical dependence. For A to be considered a cause of B in a context S, it is necessary that $Pr(B|AS) \neq Pr(B|\overline{A}S)$, i.e., the presence of A must have impact on the probability of B.

Modern representations of probabilistic models — directed graphs such as Bayesian belief networks or influence diagrams — readily combine the symmetric view of probabilistic dependence with the asymmetry of causality. A directed graph can be given causal interpretation and can be viewed as a structural model of the underlying domain.

Simon and I [13] tied the work on structural equation models in econometrics and AI to probabilistic models and formulated the semantic conditions under which a Bayesian belief network is causal.

Explanation has been postulated to be a complex problem solving process depending not only on the actual line of reasoning, but also on additional knowledge of the domain [31]. Some researchers have even postulated decoupling of the knowledge used for problem solving from the knowledge used for explanations [30]. This, in my opinion, carries the danger of possible inconsistencies between the two knowledge bases and the fact that the explanation system, although intended to provide insight into system's advice, may be plainly lying to the user. Probabilistic models are able to represent both a domain's structure and its numerical properties. This avoids the need for additional structural domain model and the related problem of possible inconsistencies between the model used for reasoning and the additional model used for the purpose of explanations. Knowing the causal structure of a domain allows for predicting the effects of action, i.e., manipulating the model. For example, reasoning within a model involving a barometer and the weather might lead to absurd results if there is no knowledge of causal interactions. Given their interdependence, an agent might feel tempted to manipulate the barometer in order to impact on the weather. Only the knowledge of the direction of causal relation allows for concluding that manipulation of the weather will have impact on the barometer reading and manipulation of the barometer will have no impact on the weather. Obviously, knowledge of causal interactions will, as noted by Clancey [5] and Swartout [27], enhance the explanations greatly. Graphical structure of a Bayesian belief network model is in itself an important component of an explanation and has been used as a part of the user interface in several practical decision support systems shells.

6 FLEXIBILITY IN THE LEVEL OF REASONING

Users of computer systems often feel lost in the amount of information that they are provided. They have been observed to become especially disoriented when working with models of very large problems. An important issue in explanation design is structuring and simplification. An explanation that tells the user indiscriminately everything that happened in the system is for sufficiently complex systems worthless. For an explanation to be effective, it should perform for the user the task of extracting the most essential knowledge and be as simple as possible.

Similarly to what happens in a rule-based system, only a subset of a large probabilistic model is at any point important for explaining a reasoning step. A probabilistic model contains sufficient information to determine what is normatively relevant and what is normatively important for an explanation. Relevance in probabilistic models has a natural interpretation and probability theory supplies effective tools that aid in determining what is at any given point most crucial for the inference. Parts of the model that are probabilistically independent from the explanandum event are clearly not relevant (this independence can be easily established and communicated to the user upon inquiry). Suermondt and I [14] recently summarized our work on the methods that can be used for such reduction in probabilistic models. Each of these methods is fairly well understood theoretically and practically implemented.

Simplification for the sake of explanations can be viewed as the process of decreasing the precision of the representation. It is always possible to give up some precision in order to make explanations simpler. Availability of the full numerical specification of the model allows to move freely among the various possible levels of specification while controlling precision and preserving axiomatic correctness. Numerical properties of probabilistic domains can be encoded at a variety of levels of precision, ranging from point probabilities, through probability intervals, to merely signs of influences. Probabilistic reasoning at the quantitative level is often hard to follow. Availability of a full quantitative specification of a model allows for manipulating the level of precision both for reasoning and for explanations. It is possible to translate quantitative reasoning into a qualitative explanation method. This is the foundation of qualitative belief propagation-based explanations [17]. Qualitative belief propagation traces the signs of change in probability on the paths from the evidence to nodes of interest. These signs are then translated into verbal form [8, 10, 17]. Qualitative explanations can be generated by extracting qualitative properties of a fully specified quantitative network and running an extremely efficient algorithm that determines the signs of changes in the network [11]. Suermondt [26] provides a thorough quantitative treatment of several issues that are complementary to the QBP explanations.

Explanations that are based on a less precise abstraction of the model provide an approximate, but correct picture of the model. This makes it easier to debug and improve the decision support system. Possible disagreement between the system and its user can be always reduced to a disagreement over the model. This differs from the approach taken by some alternative schemes for reasoning under uncertainty, where simplicity of reasoning is often achieved by making simplifying, often arbitrary assumptions (such as independence assumptions embedded in Dempster–Shafer theory and possibility theory) [32]. Ultimately, it is hard to determine in these schemes whether possibly counterintuitive or wrong advice is the result of errors in the model or errors introduced by the reasoning algorithm.

7 USER MODELING

A major factor that is argued to influence the effectiveness of user–computer interaction is the system's ability to adjust its behavior to the specific needs and expectations of the user. The mechanism underlying system's ability to adapt itself to individual differences among users is called "user modeling." Along with simplification, user modeling means exploration of relations between relevant knowledge and whatever guesses are available about the user's knowledge state. In its explanations, the system follows gradients along which the user is likely to comprehend and integrate the new knowledge. A system usually has some stored knowledge about different classes of users with respect to need for system's guidance. Users may either voluntarily classify themselves at the beginning of the interaction or the system can monitor their behavior and make assessments and guesses concerning their level of expertise [22, 25].

Graphical models, sharing the structure of directed probabilistic graphs, are a popular tool for modeling human knowledge, used in what is known as *mental models* (e.g., [1]). Using the probabilistic methods, it is possible to model the knowledge of a human user by means of the tools that are used to model the system's domain [18]. This facilitates such operations as comparing the differences between the system and user's beliefs and addressing these differences. Graphical probabilistic models have been also used for student modeling in intelligent tutoring systems [23].

8 CONCLUSION

Explanation is an important component of decision support systems. Its main role is to increase user's insight into the system advice. As the human user is the ultimate decision maker, this can be expected to have a positive impact on the quality of the resulting decisions. In addition, explanation has a tutoring role and helps in finding possible problems with the domain model.

I have argued that explaining probabilistic inference is not as hopeless as popularly believed. Directed graphical representations model explicitly independences and tie probability with causality, allowing for a concise and insightful representation of uncertain domains. The structural properties of such domains directly support user interfaces. Full numerical representation of a domain allows for manipulating with the level of precision for the sake of user interfaces.

Viewing normative methods as formal procedures that have as goal providing a numerical recommendation does not do them a complete justice. The goal of decision analysis, which is the art and science of applying decision theory to aid decision making in the real world, is providing insight into the decision and this insight, consisting of the analysis of all relevant factors, their uncertainty, and criticality of some assumptions, is even more important than the actual recommendation.

Acknowledgments

Support for this work was provided by the National Science Foundation under grant IRI–9624629 to University of Pittsburgh and by DARPA's Computer Aided Education and Training Initiative under grant N66001–95–C–8367.

References

1. R. Axelrod. The cognitive mapping approach to decision making. In R. Axelrod, editor, *Structure of Decision*, pages 221–250. Princeton University Press, Princeton, NJ, 1976.
2. A. Barr and E.A. Feigenbaum. *The Handbook of Artificial Intelligence, Volume 1.* Morgan Kaufmann, Los Altos, CA, 1982.
3. Eugene Charniak and Solomon E. Shimony. Probabilistic semantics for cost based abduction. In *Proceedings of the 9th National Conference on Artificial Intelligence, AAAI–90*, pages 106–111, Boston, MA, 1990. American Association for Artificial Intelligence.

4. Jay J.J. Christensen-Szalanski. Experience and the base rate fallacy. *Organizational Behavior and Human Performance*, 29(2):270–278, April 1982.
5. William J. Clancey. Use of MYCIN's rules for tutoring. In Bruce G. Buchanan and Edward H. Shortliffe, editors, *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, chapter 26, pages 464–489. Addison-Wesley, Reading, MA, 1984.
6. Robyn M. Dawes. *Rational Choice in an Uncertain World.* Hartcourt Brace Jovanovich, Publishers, 1988.
7. Marek J. Druzdzel. Towards process models of judgment under uncertainty. Technical Report CMU–EPP–1990–03–03, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, October 1989.
8. Marek J. Druzdzel. *Probabilistic Reasoning in Decision Support Systems: From Computation to Common Sense.* PhD thesis, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, December 1992.
9. Marek J. Druzdzel. Some properties of joint probability distributions. In *Proceedings of the Tenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–94)*, pages 187–194, Seattle, WA, 1994.
10. Marek J. Druzdzel. Qualitative verbal explanations in Bayesian belief networks. *Artificial Intelligence and Simulation of Behaviour Quarterly*, 94:43–54, 1996.
11. Marek J. Druzdzel and Max Henrion. Efficient reasoning in qualitative probabilistic networks. In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI–93)*, pages 548–553, Washington, D.C., 1993.
12. Marek J. Druzdzel and Max Henrion. Intercausal reasoning with uninstantiated ancestor nodes. In *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI–93)*, pages 317–325, Washington, D.C., 1993.
13. Marek J. Druzdzel and Herbert A. Simon. Causality in Bayesian belief networks. In *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI–93)*, pages 3–11, Washington, D.C., 1993.
14. Marek J. Druzdzel and Henri J. Suermondt. Relevance in probabilistic models: "Backyards" in a "small world". In *Working notes of the AAAI–1994 Fall Symposium Series: Relevance*, pages 60–63, New Orleans, LA (An extended version of this paper is currently under review.), 1994.
15. Marek J. Druzdzel and Linda C. van der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI–95)*, pages 141–148, Montreal, Canada, 1995.
16. Max Henrion. Uncertainty in artificial intelligence: Is probability epistemologically and heuristically adequate? In Jeryl Mumpower, Ortwin Renn, Lawrence D. Phillips, and V.R.R. Uppuluri, editors, *Expert Judgment and Expert Systems, Proceedings of the NATO Advanced Research Workshop on Expert Judgment and Expert Systems*, pages 105–129, Porto, Portugal, 1986. Berlin, Germany: Springer Verlag.
17. Max Henrion and Marek J. Druzdzel. Qualitative propagation and scenario-based approaches to explanation of probabilistic reasoning. In P.P. Bonissone, M. Henrion, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 17–32. Elsevier Science Publishers B.V., North Holland, 1991.
18. Ronald A. Howard. Knowledge maps. *Management Science*, 35(8):903–922, August 1989.
19. Daniel Kahneman, Paul Slovic, and Amos Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge University Press, Cambridge, 1982.

20. Kenneth Kotovsky, John R. Hayes, and Herbert A. Simon. Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17(2):248–294, April 1985.
21. Paul E. Lehner, Theresa M. Mullin, and Marvin S. Cohen. A probability analysis of the usefulness of decision aids. In M. Henrion, R.D. Shachter, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*, pages 427–436. Elsevier Science Publishers B.V. (North Holland), 1990.
22. Clayton Lewis. Understanding what's happening in system interactions. In Donald A. Norman and Stephen W. Draper, editors, *User Centered System Design: New Perspectives on Human-Computer Interaction*, pages 171–185. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.
23. Joel Martin and Kurt VanLehn. Student assessment using Bayesian nets. *International Journal of Human-Computer Studies*, 42:575–591, 1995.
24. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988.
25. Mary S. Riley. User understanding. In Donald A. Norman and Stephen W. Draper, editors, *User Centered System Design: New Perspectives on Human-Computer Interaction*, pages 157–169. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.
26. Henri J. Suermondt. *Explanation in Bayesian Belief Networks.* PhD thesis, Department of Computer Science and Medicine, Stanford University, Stanford, CA, March 1992.
27. William R. Swartout. XPLAIN: a system for creating and explaining expert consulting programs. *Artificial Intelligence*, 21(3):285–325, September 1983.
28. Peter Szolovits. Artificial intelligence in medicine. In Peter Szolovits, editor, *Artificial Intelligence in Medicine*, pages 1–19. Westview Press, Boulder, Colorado, 1982.
29. Peter C. Wason and Philip N. Johnson-Laird. *Psychology of Reasoning: Structure and Content.* Harvard University Press, Cambridge, MA, 1972.
30. Michael R. Wick. The 1988 AAAI Workshop on Explanation. *AI Magazine*, 10(3):22–26, Fall 1989.
31. Michael R. Wick and William B. Thompson. Reconstructive expert system explanation. *Artificial Intelligence*, 54(1–2):33–70, March 1992.
32. Ben P. Wise and Max Henrion. A framework for comparing uncertain inference systems to probability. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 69–83. Elsevier Science Publishers B.V. (North Holland), 1986.