69

Pople, Harry W. Jr. "Heuristic Methods for Imposing Structure on Ill-structured Problems," in *Artificial Intelligence in Medicine*, P. Szolovits, ed., Westview Press (1982), pp. 119–190.

Shortliffe, E. H. *Computer-Based Medical Consultations: MYCIN*, Elsevier- North Holland, New York (1976).

69

Auditor Productivity in the Year 2000: 1987 Proceedings of the Arthur Young Professors' Roundtable. Andrew Bailey, ed. pp. 121-128, 1988

DISCUSSION OF INTEGRATION OF TWO KINDS OF EXPERTISE: INTELLIGENT EXPLORATORY BEHAVIOR AND NORMATIVE ALGORITHMS

GLENN SHAFER

University of Kansas

I am honored by this opportunity to comment on the paper by Chandrasekaran and Dillard. Professor Dillard is a leader in the effort to provide automated assistance for the audit task, and Professor Chandrasekaran is a leader in deepening our understanding of the potential of knowledge-based systems in general.

This paper brings together two themes for which Chandrasekaran is well known: the importance of studying generic tasks, and the value of integrating AI and OR techniques. On both of these very sensible themes, he deserves our applause. His work on them continues to be informative and influential. I will comment on the second theme, the integration of AI and OR.

I will argue that the terminology that Chandrasekaran uses in discussing this integration is not helpful. The algorithms that we find in AI are not particularly "intelligent," and the algorithms that we find in statistics and operations research are hardly always "normative" or optimal. Algorithms are algorithms, and we are seldom able, in any field, to label an algorithm optimal.

I will also argue for a wider role for probability than the role Chandrasekaran sees for it. It is a mistake to think that we are not using probability ideas just because we do not use numbers. Probability is not really about numbers; it is about the structure of reasoning.

The Role of Optimality. A great deal is said about optimality in the theory of statistics and in the theory of operations research. Does this mean that practical statisticians, decision analysts, and other management scientists generally use optimal methods? Or that they should do so?

Optimality is also discussed in the theory of algorithms in computer science. Does this mean that programmers use optimal algorithms? Or that they should do so?

Of course not. Every field of engineering must seek frameworks for evaluation—frameworks that allow us to compare methods and make some judgments about which are better. Artificial intelligence is no exception. Most of the literature in artificial intelligence, like most of the literature in statistics, operations research, and other branches of computer science, is about such comparisons of methods.

Statistics and operations research frameworks sometimes allow us to identify optimal methods—methods that are the very best. We are fond of these frameworks. We emphasize them in our introductory courses. But practical statisticians soon learn than they do not take us very far. In the real world, we can never make the assumptions that are needed to be certain that our methods are optimal. Usually we cannot even make enough assumptions to be certain they are "admissible." We simply look for methods that are reasonable. That is the real purpose of the theory—to give us some pointers to recognize when some methods might be more reasonable than others.

In summary, the concern for optimality does not distinguish statistics and operations research from applied AI.

What is Normative? Chandrasekaran also uses the word "normative" to characterize the methods of statistics and operations research. Where does this word come from?

The word actually belongs not to the probabilists but to the logicians. In the latter half of the nineteenth century, logicians began to distinguish between the "normative" and "descriptive" interpretations of symbolic logic. Logic may not be descriptive, they said, but it is normative. People may not think that way. But they should.

It is only relatively recently that this distinction between "normative" and "descriptive" was introduced into statistics. The culprit was L.J. Savage, who argued in his book, *The Foundations of Statistics*, published in 1954, that it was normative to make judgments using Bayesian rules. People might not always think that way, he said, but they should.

The majority of statisticians, theoretical and applied, are not Bayesians. This majority has never claimed the label "normative" for its methods. I myself, since I am an advocate of a theory of subjective probability that is not Bayesian, have argued at length against Savage's normative claims; see, for example, Shafer (1986). And even many Bayesians dissent from Savage's claims of normativeness for their methods.

The word "normative" is also inappropriate for most of the non-statistical methods of operations research. What would it mean to call the simplex method "normative"? The method has some optimality properties, but then so do hammers and screwdrivers. What would it mean to call a hammer "normative"?

It is true there is a cadre of Bayesian statisticians and decision analysts who are telling the AI community that Bayesian methods are normative. But in spite of the amount of noise they make, these people do not represent a majority in the statistics and operations research communities.

In summary, applied statistics and operations research is no more "normative" than applied AI. Indeed, to the extent that applied AI is applied logic, it has a greater historical claim to the word "normative" than applied statistics and operations research do.

Are the Methods of Applied AI Intelligent? I am not among those who ridicule artificial intelligence by saying that its achievements have nothing to do with intelligence. The academic AI community has been very productive, and it is no criticism of this community that the things they have taught us should now seem to lie outside the mystery of intelligence.

All the same, I think it confuses rather than clarifies to use the adjective "intelligent" to distinguish rule-based systems and other practical AI ideas from other algorithms. Algorithms are algorithms.

Rule-based systems were born amidst a burst of enthusiasm for the idea that "people think like that," but that burst was short-lived. Now that we understand rule-based systems so well, it seems more to the point to emphasize ways in which "people are different from this." For one thing, human reasoning is surely more associative.

Rule-based systems are one way of integrating relatively unstructured information, and people are not really very good at integrating unstructured information.

Consider these two examples: (1) By linearly regressing the predictions of clinical psychologists on the cues that these psychologists use, one can obtain a prediction equation that predicts better than the psychologists themselves (Dawes, 1979); (2) By formalizing the rules of thumb used by cotton farmers, one can obtain an expert systems that makes planting, irrigation, and pesticide decisions better than the farmers themselves (COMAX).

Both these examples illustrate Dawes's point that "people—especially the experts in a field—are much better at selecting and coding information than they are at integrating it" (p. 394). The computer is better at integration, whether it does it using arithmetic, as in the first example, or logic, as in the second example.

How to distinguish? How, then, can we distinguish the algorithms of AI from those of OR?

We might talk about applied logic as opposed to applied arithmetic. Or we might emphasize typical differences in the data used. A knowledge base may not have the kind of repetitive data needed by a statistical method.

Perhaps, however, we do not need to distinguish. Perhaps an algorithm is an algorithm.

A Catholic Conception of Our Enterprise. There is a common enterprise shared by applied statisticians, decision analysts, applied AI workers, and psychologists working in the area of judgment and decision making.

At the foundation of this enterprise are formal mathematical decision techniques—algorithms and heuristics; these are two names for the same thing. But the emphasis should be on algorithm as process, not on the optimality of its result; in Simon's words, procedural rather than substantive rationality.

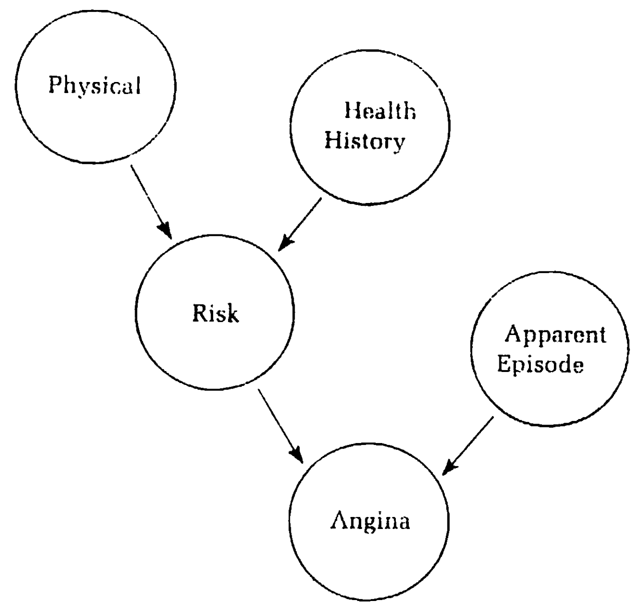
The role of the theorist is to devise and to study these algorithms—to try to create frameworks in which they can be evaluated theoretically—sometimes even to create frameworks in which they can be shown to be optimal.

The role of the applied psychologist is to study how people can and do use these algorithms—with and without assistance. Which algorithms do people tend to use and under what circumstances? Which can they learn to use effectively?

The role of the practitioner is then to put these algorithms to use, taking what account he or she can of the abstract work of the theorist and the empirical work of the psychologist.

Is Probability About Numbers or Structure? A year or two ago, I was debating with an AI colleague the familiar issue of whether probability is needed in knowledge-based systems. He proceeded to tell me about a fragment of a system that he was building without probability. There was no probability in it, he said, because there were no numbers. Instead of numbers, he used a half dozen words: certain, nearly certain, very probable, and so on.

Then, he drew me as picture of how he combined his evidence, a picture something like this:



He combined evidence from a physical and a health history to get a judgment about how much at risk of heart disease the patient was, and then he combined this with the patient's description of an apparent angina episode to get a judgment of whether the patient really had angina. It was this structure that was really important, he explained to me, not probabilities.

This was too much for me. Probability, I retorted, is about structure, not about numbers. Probability is the theory of structures like this one. You have to use probability language to explain this structure. You have to say, "I can forget about Physical and Health History when I am combining Risk with Description of Episode because Physical and Health History are independent of Description given Risk."

When I made that speech, I felt I was telling my friend something he should have known, but since then I have had time to ask myself how any worker in AI, or anyone else who has not studied probability for ten years, could possibly know it. Is there anything in our teaching of probability that could alert the student to the fact that probability is really about structure, not about numbers?

Look at how we teach probability. We start with the axioms of what the mathematicians call measure theory, together with the definition of conditional probability, $P(A|B) = P(A\&B)/P(B)$. No structure there.

When do we get to structure? We get to it as a very special topic in some very advanced courses. An advanced course in stochastic processes might get to the topic of Markov trees and Markov fields. In the fourth or fifth course in statistical inference, we get to path analysis. And in a course in Bayesian inference, we might just occasionally leave the conundrums of parametric inference aside long enough to talk about hierarchical inference.

It is time—and I have been preaching this to the statisticians and management scientists—that we start bringing the idea of structure into our elementary teaching of probability. Here are four ways to do this.

Idea #1: Replace the definition of conditional probability with a theorem. (Shafer, 1985.)

Idea #2: Replace the deFinetti-Savage arguments for normativeness with the constructive philosophy. (Shafer and Tversky, 1985.)

Idea #3: Emphasize the axiomatic study of conditional independence. (Pearl and Verma, 1987.)

Idea #4: Give the theory of Bayesian hierarchical inference priority over Bayesian parametric inference. (Kelly and Barclay, 1973; Kiiveri, Speed, and Karlin, 1984; Pearl, 1986; Shafer, Shenoy, and Mellouli, 1987.)

The Pitfalls of Ecumenical Work. In trying to bring together AI and OR, we are doing more than combining techniques. We are bringing together two academic communities—two different cultures. We are doing ecumenical work.

Such work is always difficult. It is a struggle to establish a common vocabulary and agree on a conception of a common task. Modes of expression that seem innocent to a representative of one culture can be full of tendentious and invidious meaning to a representative of another, and human beings are always tempted to compare the ideals and best practice of their own culture to the excesses and shortcomings of a foreign one.

The hardest aspect of appreciating a foreign culture is understanding the delicate relation between theory and practice. A practical person learns a myriad of ways of adapting his culture's theory and ideals to reality. It is often hard for the foreigner to see the earnestness of this adaptation, but it is often easy for him to see absurdities in the ideals and hypocrisies in the practice.

Imagine what happens when an applied statistician looks at AI, and what happens when a practitioner of applied AI looks at statistics or OR. Both will look for understanding to the theory of the other field, and both will be astonished by the excessive formality and intemperate claims that they see. The statistician is apt to be astonished by the doctrine—earnestly argued by many theorists in AI—that all thought can be encompassed by symbolic logic. The AI practitioner is apt to be astonished by the doctrine—argued equally earnestly by many theorists in statistics and operations research—that it is always "normative" to use numbers. The practitioners know how much salt to take with their own theorists' claims, but it is hard for them to get just the right measure of salt to take with the claims of the theorists in the other field.

My suggestion is that practitioners in both fields will find it easier to see how much they have in common if they dump the invidious rhetoric of their theorists. Who needs the talk about "intelligence" or the talk about "normativeness"? Let's talk about algorithms.

REFERENCES

Dawes, Robyn M. The Robust Beauty of Improper Linear Models in Decision Making. *American Psychologist* 34 (1979)

Kiiveri, Harri, T. P. Speed, and J. B. Carlin. Recursive Casual Models. *Journal of the Australian Mathematical Society, Series A* 36 (1984), pp. 30–52.

Kelly, Clinton W., III and Scott Barclay. A General Belief Model for Hierarchical Inference. *Organizational Behavior and Human Performance*, 10 (1973), pp. 388–403.

Pearl, Judea. Fusion, Propagation, and Structuring in Bayesian Networks. *Artifical Intelligence*, 29 (1986), pp. 241–288.

——— and Thomas Verna. The Logic of Representing Dependencies by Directed Graphs. *Proceedings of the Sixth National Conference on Artificial Intelligence*, Seattle. Volume 1 (1987), pp. 374–379.

Savage, L. J. *The Foundations of Statistics*. Wiley (1954).

Shafer, Glenn. Conditional Probability. *International Statistical Review*, 53 (1985), pp. 261–277.

———. Savage Revisited (with discussion). *Statistical Science* 1 (1986), pp. 463–501.

———, Prakash P. Shenoy, and Khaled Mellouli. Propagating Belief Functions in Qualitative Markov Trees. *International Journal of Approximate Reasoning*. Vol. 1 (1987), pp. 349–400.

——— and Amos Tversky. Languages and Designs for Probability Judgment. *Cognitive Science* 9 (1985), pp. 309–339.

Simon, Herbert A. Rationality in Psychology and Economics. *Journal of Business*, 59 (1986), pp. S209–S224.