Artificial Intelligence and Law **9:** 189–197, 2001.

Review Essay

Complexity, transparency, and the warranted use of formal systems in legal factfinding

VERN R. WALKER
Hofstra University School of Law, 121 Hofstra University, Hempstead, New York 10549, USA
E-mail: lawvrw@hofstra.edu

Joseph B. Kadane and David A. Schum, *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*, New York: Wiley, 1996. xvi + 366 pp. (cloth). ISBN 0471-1141-828.

Glenn Shafer, *The Art of Causal Conjecture*, Cambridge, Mass.: MIT Press, 1996. xx + 511 pp. ISBN 0262-1936-8X.

Two recent books invite us to reflect on the use of formal probability systems in legal factfinding. More generally, they make evident certain difficulties for applying artificial intelligence systems in such factfinding. The two books are: Joseph B. Kadane and David A. Schum, *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*, and Glenn Shafer, *The Art of Causal Conjecture*. Out of the many difficulties in applying formal systems, I focus in this brief review on using them to assist factfinding in a legal context. The consideration added by this context is that a finding of fact, and the use of a formal probability system to support it, must be transparently warranted by the evidence.

To set the stage, we must understand something about the context of legal factfinding. We often justify legal decisions by relying on propositions that are considered to be true for purposes of the decision. Those propositions form the factual basis for the decision. The propositions themselves are "found" to be true by people authorized to make such "findings of fact", and such findings are usually made in the course of some well-defined legal procedure. This process of factfinding occurs in all branches of government. The legislative branch may hold hearings and record its findings in a statute or in the legislative history for the statute. The executive branch may conduct rulemaking or adjudicatory proceedings, in which administrative officials record their official findings of fact. Judges in the judicial branch may conduct evidentiary hearings or trials that result in findings or verdicts. The procedures involved may be peculiar to each type of legal institution, but one central objective of factfinding is the same in each setting. Factfinding helps to provide legitimacy for governmental action by identifying for potentially affected parties the propositions considered to be true by the decisionmaker, the evidence on which those conclusions rest, and the reasoning connecting that evidence to those conclusions. Findings provide legitimacy only if the evidentiary basis for them and the inferences to them are reasonably transparent.

With such factfinding contexts in mind, we can address the question: Are there special problems for applying formal systems, or artificial intelligence systems generally, in such factfinding contexts? The answer, I believe, is yes. Moreover, the books by Kadane and Schum and by Shafer help us to see why this is so. The peculiar problems arise precisely because a finding must be reasonably transparent. If a formal system is used in reasoning to a finding, then that reasoning must also be transparent This transparency requirement compares favorably to the Euclidean and Cartesian ideal of mathematical proof. Both modes of providing warrant have transparency as one of the essential conditions of an acceptable line of reasoning. But the mathematical ideal can afford to presume an artificially constructed language, as well as a specially trained and patient audience, in a way that legal justification cannot. Legal factfinding, in order to be effective, must be transparent to ordinary, intelligent and educated citizens. It must provide a justification that is reasonably transparent to potentially affected parties.

1. Models of inferences and events

Perhaps there is no better illustration of what is at stake in legal factfinding than the murder trial of Nicola Sacco and Bartolomeo Vanzetti in Massachusetts in 1921. Both defendants were convicted of first-degree murder and they were executed in 1927. Much of the history surrounding this controversial episode is recounted in the book by Kadane and Schum, but that history is not the major focus of the book. Rather, the authors use the extensive evidentiary record in the legal case to apply many of the analytical methods described by David Schum in his earlier book, *The Evidential Foundations of Probabilistic Reasoning* (John Wiley & Sons, Inc., 1994). They use those methods to provide an organization for the extensive Sacco and Vanzetti trial evidence (testimony from 169 witnesses and 85 exhibits) and the post-trial evidence, for a total of "395 substantively different items of evidence" (Kadane and Schum, pp. 13, 23, 78–80). They organize this evidentiary record into a single "evidence chart" about 18 feet long, consisting of 28 sector charts that map the reasoning within important sub-areas (Kadane and Schum, p. 88). In this brief review of their work, I cannot provide many details of those methods. What is important for present purposes is that in these evidence charts the authors purport to take into account all known and relevant evidence concerning the guilt or innocence of Sacco and Vanzetti (Kadane and Schum, pp. 82–83). Moreover, they construct what they consider to be plausible chains of reasoning connecting all of these items of evidence to the ultimate factual issues in the case, and they assess the "probative force" of some of the evidence. More importantly, they illustrate a method for assessing the probative force of individual items of evidence, as well as the cumulative probative value of combinations of evidence.

An evidence chart, which is an adaptation of the highly original work by John Henry Wigmore, consists of a directed acyclic graph, or tree diagram. It models the inferences from items of evidence to intermediate and ultimate conclusions. Nodes in the tree represent propositions that are capable of being either true or false. We can think of a single node whose truth-value is uncertain, such as a node representing an intermediate or ultimate conclusion, as representing a pair of propositions, "P" and its negation "P^C". The arcs or arrows connecting the nodes represent "probabilistic linkages" of inferential support between propositions. These arrows of inference run from the nodes at the ends of the branches (the evidence) to the root node for the tree (the ultimate factual issue or hypothesis). Any two nodes in the same tree either lie within a directional path (that is, they are connected by a single arrow or by a sequence of arrows) or they occur in different paths representing different lines of reasoning. What unifies the reasoning into a single tree is that all the evidence is important (or at least relevant) for inferring the truth or falsehood of the ultimate issue (represented by the tree's root node). A chart is "acyclic" if there are no cycles – that is, there are no ordered paths from any node that will return you to the node from which you began (Kadane and Schum, p. 71). All reasoning within a single tree runs from the evidence to the ultimate issue of fact to be decided.

Glenn Shafer uses a formally similar tree model, which is acyclic in the same sense (Shafer, pp. 230, 466). However, Shafer gives his tree models a different interpretation. Shafer constructs "event trees", with "situations" represented by nodes and "Humean events" as the steps between nodes (Shafer, pp. 23–26, 32–35). Such an event might be the fall of a rock at 2:00 p.m. on July 18, 1996, on Mount Albany, Maine, or the event of a particular young man named Mark dropping out of college (Shafer, pp. 43–49). In this situational interpretation of the model, an arrow between nodes represents the occurrence of an event. Therefore, it is normal for us to think of the arrows as "running" in the same direction as time, from the root node of the tree to the nodes at the ends of the branches.

The nodes other than the root node represent the situations that might occur if the root event happens. When an event at some point in a path "happens", then some events elsewhere in the tree become certain, some become impossible, and others are still possible but unresolved. An event tree therefore models the possible outcomes if some specified event occurs. A "probability tree" is an event tree that also has probability assignments for its branching events. Human experience can be characterized in part as a shift in our information about events as "nature" moves from situation to situation along a path in "nature's probability tree". As we experience change, we become aware of which events happen and we receive information about them. What may also change over time is our ability to predict other events.

Shafer therefore uses probability trees to model the real world of our experience. Probability trees do not model the world as present to an omniscient God, but the world of events as it is potentially knowable by us, framed by our human concepts and our science. To understand the real world is to apprehend the probabilistic structure of the world as it would appear to an observer with the most complete empirical knowledge available at the time. As Shafer puts it, the ideal of empirical knowledge is to model the world as "nature observes it" (Shafer, p. 9). The goal is to develop as our own model of the world partial descriptions of the probability tree that nature possesses for those same events (Shafer, p. 29). As individuals with limited knowledge, we try to conform our personal probability trees for limited sets of events to the relevant portions of nature's single probability tree for all events. Once we understand human knowledge in these terms, then we should be able to redefine in probabilistic terms all of the concepts and techniques needed for achieving this goal. This is what Shafer sets out to do. If the project is successful, we should understand why concepts of causal regularity are useful devices in the quest for empirical knowledge, and we might be able to devise more effective and more efficient techniques for making causal conjectures.

Although the above account of empirical knowledge may seem unduly metaphysical, Shafer uses it to balance the strict formalism that dominates most of his presentation. It is easy to become disoriented as Shafer abstracts each formal system to a yet more general system, and keeps redefining our concepts of events, probabilistic relations, and causation. But such reconceptualization is essential to Shafer's enterprise precisely because of his view of empirical knowledge. Causation and scientific method need to be restated in probabilistic terms because the goal of empirical knowledge is to understand the world in probabilistic terms, which is how the world appears to nature and will ultimately appear to science. The structure of the knowable world is probabilistic.

2. Complementarity of the models

I suggest that the analytical endeavors of Kadane and Schum and Shafer should be viewed stereoscopically and the books should be read together. They are largely complementary to each other. Kadane and Schum attempt to model our evidence and our lines of inference from evidentiary propositions to conclusions of fact. Shafer attempts to model the world of events that those propositions describe. While Kadane and Schum model our inferences about events, Shafer models the causal structure that we attribute to those events. Kadane and Schum model our thinking about the world, and Shafer models the world about which we think.

Both analyses use a directed acyclic graph as the formal structure of their models. Where the two types of model diverge is in the interpretations given to the tree diagrams. The complementarity becomes evident in the probabilistic measures assigned to the arrows that connect nodes within the models. Both models use the traditional Pascalian probability calculus in order to define the probabilities that are assigned to these arrows. That is, probabilities in each model behave as traditional probabilities behave, satisfying the standard Kolmogorov axioms (Kadane and Schum, pp. 118–119; Shafer, pp. 68–69, Appendix D). But the two models employ two different types of probability assignments, which reflect the differences in the models. The "probabilistic linkages" of Kadane and Schum measure the "probative force" or "inferential force" of evidence for inferring the truth and falsehood of conclusions, while the "situational probabilities" of Shafer measure conditional probabilities of occurrence for nodes or events in the tree. I will briefly discuss the nature of each of these probability assignments.

Kadane and Schum primarily employ a Bayesian measure of inferential force, although they do discuss possible alternative measures. We can let the proposition "E" represent a witness's testimony or state the propositional significance we attach to some item of physical evidence. Moreover, let the proposition "H" stand for any hypothesis or possible conclusion, and let "H^C" represent its negation, not-H. Then the "likelihood ratio" for evidence E relative to the possible conclusions H and H^C is a ratio in which the numerator is the conditional probability that the proposition E would be true assuming the truth of the conclusion H, symbolized as "$P(E \mid H)$", and the denominator is the conditional probability of proposition E assuming the truth of H^C, written as "$P(E \mid H^C)$". The likelihood ratio quantifies the change, if any, from the "prior" odds of H to H^C *before* E is known, or "$\Omega(H{:}H^C)$", to the "posterior" odds of H to H^C *after* evidence E has been taken into account, or "$\Omega(H{:}H^C \mid E)$". Within certain constraints, likelihood ratios can be assigned to each arrow in the inference network, and can be used as a measure of the probative force of the evidence in making the inferential step represented by the arrow.

Shafer's trees, by contrast, are event trees, not trees constructed on propositional nodes. Arrows in Shafer's models are assigned probabilities of event occurrence. Moreover, those probabilities are always "probabilities in a situation" – that is, probabilities relative to an event node within the tree. The primitive probability operator in Shafer's formalism is "$\mathbf{P}_S(T)$", which is interpreted as the probability of event T in the situation S, or relative to node S (Shafer, p. 68). In many trees, such a situational probability can be interpreted in terms of the more familiar conditional probability, so that "$\mathbf{P}_S(T)$" can be taken to mean the same as "$P(T \mid S)$" (Shafer, 72–73, Appendix D). Thus, arrows in Shafer's model can be assigned probabilities that behave like classical conditional probabilities. An assigned situational probability predicts the occurrence of an event contingent upon the occurrence of some situation in the event tree. What we are trying to do in science is discover nature's situational probabilities of occurrence for events that can occur "downstream" from any set of events.

I have suggested that the two types of model should be viewed as being complementary to each other. Normally *both* kinds of model would be used to justify making any finding of fact. The exact relationship between the models can be expressed using the concept of warrant. First, what *warrants* a particular finding or conclusion is a line of reasoning from evidence that can be modeled with a Kadane and Schum evidence chart. Second, what *warrants* the assignment of a particular likelihood ratio to a particular arrow in the evidence chart is a generalization about how events normally occur. Finally, what *warrants* the generalization is a causal account of what types of events normally lead to other types of events. But this causal account of events is precisely what Shafer attempts to model. Put succinctly, any particular likelihood ratio in any particular evidence chart (Kadane and Schum) should be *warranted* by a probability tree that connects the relevant types of events (Shafer).

3. Three dimensions of complexity

With this general understanding of the formal models proposed by Kadane and Schum and by Shafer, and with this glimpse of their complementarity, we can anticipate three dimensions of complexity inherent in applications of their models. Using the models forces us to increase their complexity in each of these three dimensions.

First, there is what we might call "description complexity". In any particular application, even minimally adequate descriptions of our reasoning or of events in the world tend to be complicated. We typically reason in many steps and branches, and we conceive of events in the world as inherently complicated. Branches can always be decomposed into constituent branches, and nodes that might be considered "root" or "terminal" for one purpose can always branch to yet more nodes. As purely formal structures, trees are potentially infinite in the same way that the set of real numbers is potentially infinite. Interpreting those formal structures as describing either chains of reasoning or chains of events leads to particular, applied models that are actually complex.

This seems to be the case whether the model is that of Kadane and Schum or that of Shafer. Within a Kadane and Schum evidence model, we can "further decompose" any chain of inferences into constituent inferences, and we can always support our generalizations with ancillary evidence and further reasoning (Kadane and Schum, pp. 52–53, 87–89). Moreover, which propositions should count as basic evidence and which as "ultimate" conclusions can be determined only pragmatically, by a context external to the model. For example, in the case of Sacco and Vanzetti, such pragmatic closure is provided by the criminal law, the law of evidence, and the theories of the attorneys. But even within these confines, the Sacco and Vanzetti case study forcefully illustrates how quickly evidence trees become complex in practice, how likelihood ratios proliferate, and how difficult the computations become.

While the evidence in the Sacco and Vanzetti trial is extensive, it is nevertheless primarily the testimony of fact witnesses. It is far less complicated than the evidentiary records common in legal proceedings today, especially with the great increase in the use of expert testimony. Yet even Kadane and Schum, faced with the complexity of their own evidence charts for the Sacco and Vanzetti case, do not provide a Bayesian analysis at a detailed ("decomposed") level for the entire evidence chart, and they explain candidly that "[w]e could not even enforce on ourselves the labor of assessing likelihoods for this entire network" (Kadane and Schum, pp. 185–239, 264). After we work through the complexity of the analysis that Kadane and Schum so painstakingly undertook, it is hard to imagine how any expert witness could describe an inference network in any moderately complicated trial, and hope to use it to convince an otherwise skeptical trier of fact. This case study by Kadane and Schum makes clear the enormous complexity of an inference model if it is adequately descriptive, and their book forces us to wonder how the method could be useful in any serious legal factfinding.

Applying Shafer's model to describe events also promises to produce very dense and very extensive probability trees. By contrast with the Sacco and Vanzetti case, the examples that Shafer provides are all extremely simple. Nevertheless, it is not difficult to imagine the tremendous complexity if the model were ever used rigorously to describe a realistic set of events. For Shafer, nature's event tree probably begins at least as early as the Big Bang of the universe and explodes into the indefinite future, and Humean events (such as a rock falling or Mark dropping out of college) can be refined into subsidiary events with more detail (Shafer, pp. 275ff). Although Shafer does not supply us with a rich case study of his model as applied, we readily anticipate that serious applications of his event models would be at least as dense and extensive as those of Kadane and Schum.

Both books show us a second type or dimension of complexity, which we can call "validation complexity". Along the first dimension discussed above, increased complexity results from the fact that the tree models are used to describe actual chains of reasoning or chains of events. The second dimension is that the structural complexity of any particular, applied model necessarily increases whenever we attempt to validate it. If an applied model is challenged, the attempt to defend it makes the model even more complex. For example, if I happen to believe the alibi testimony that Sacco was not at the scene of the murder, and you challenge my reasoning, I generally resort to more reasoning, more generalizations, and more ancillary evidence. I provide more evidence and inferences about the behavior of the alibi witnesses, their lack of motives to lie, the plausibility of the alternative account of Sacco's whereabouts, and so forth. The same is true about my probability tree concerning the events involved. If you challenge my probability assignments about how likely it is that Sacco would resort to murder given other events in his past, in defense of my belief I would introduce new generalizations based on other patterns of events – such as how Sacco allegedly behaved at other times in other situations. In defending either kind of model, whether about inferences or about events, the method for defending any particular likelihood ratio or situational probability of occurrence is to provide a more complex tree. The structures of particular models, as applied, become necessarily more complex whenever the validity of the models is challenged and the models' proponents attempt to warrant the accuracy of their models.

There is yet a third dimension of complexity that is illustrated by the two books. We can call it "construct complexity". Not only are the particular trees that are used in particular cases necessarily and increasingly complex in their structures, but any explanation of the nature or usefulness of these types of modeling is also complex. Kadane and Schum are exceptionally lucid in explaining evidence chart construction and the benefits to be gained from constructing such charts, but it still requires a great deal of dedication to follow their explanations. Moreover, these authors clearly refuse to make any particular claim about how to use their models in a trial setting (Kadane and Schum, pp. 197–198, 258–265). (In fact, one of the great strengths throughout their book is the frank discussion of the limitations of their methods.) Shafer's analysis forces us to work through the syntax and semantics of event trees, probability trees, and causal conjectures, and to appreciate the richness of the related ontologies. As Shafer marches us away from our vague but familiar concepts of ordinary language, and replaces those concepts with more carefully defined but more abstract ones, we realize that explaining what we are up to takes superior human skill. Skill far beyond the capabilities of lawyers and legislators and administrators and judges, let alone ordinary jurors. "Surely", we think to ourselves as we read these books, "there must be some commensurate factfinding gain for all this effort!"

4. Transparent warrant in factfinding

This last point brings us back to the context of legal factfinding. It would be a serious mistake to dismiss the typical players in factfinding proceedings as simply unintelligent, or unwilling to take adequate time, or uninterested in discovering the truth. Although such factors often help to explain human behavior, there is something much more significant going on. Legal factfinding must be transparently warranted. Legal factfinding seeks conclusions about actual events in the world, but conclusions transparently based on evidence. The methods of Kadane and Schum *appear* to be useful to this task because their evidence charts model the inference networks linking our evidence to our conclusions. Shafer's methods *appear* useful because his probability trees model those relationships among events that we are trying to infer. But – and this is the significant point – if those models were applied in a conscientious way to *describe* any legally significant situation, they would be incredibly complex, even at the outset. That initial structural complexity would increase, probably exponentially, whenever the *validity* of those models came under attack – as the models certainly would in any legal proceeding. And finally, as the amount of time and effort needed to do the modeling increased and became clear, the model proponents would be asked to explain the usefulness of using such complex *constructs* at all.

Complexity, transparency and warrant are moments in a dynamic process. We know we must warrant the application of formal systems to the circumstances of the legal case. But in attempting to provide warrant, we necessarily create additional complexity. The increasing complexity decreases the transparency of both the applied models and the alleged advantages of using them. With this loss of transparency, we lose all hope of transparent warrant. In addition, once the nature of this dynamic progression is understood, moving through the progression appears inevitable once it begins. So why even *try* to apply such formal probability systems in the context of legal factfinding? To attempt to do so appears to be self-defeating.

I do not think that this is an insurmountable problem, simply a very serious and inherent one. It is a great contribution of these two excellent books that they bring us face-to-face with the enormity of this problem. And the problem to which I refer is not simply the difficulty of modeling a complex subject matter. To see the difficulty solely in such terms is to underestimate it. Formal modeling has always encountered increasing complexity in its target subject matter. If artificial intelligence modelers were daunted by complexity, progress would never be made. Doubts about the capabilities of artificial intelligence have generally been dispelled by results, by creating more advanced systems that in fact perform well on even complicated tasks. This familiar process will surely continue to occur in attempts to model inference networks and possible events. Indeed, the books by Kadane and Schum and by Shafer make significant contributions to this progress.

But mere complexity of subject matter is not the whole of the problem I am suggesting. The precise question is how formal systems, including artificial intelligence systems generally, can be of assistance in legal factfinding, when a principal test of their very usefulness is transparency of the warrant they help provide? Successful performance in the context of legal factfinding involves more than merely getting "an answer" to a problem. It involves getting that answer by a means that also provides warrant for that answer, and warrant that is reasonably transparent to participants in the factfinding process and to those potentially affected by its outcome. An artificial intelligence application in the context of legal factfinding cannot leave the ordinary world behind while it makes an opaque computation and returns an answer. Somehow the formal system as applied must take ordinary nonexperts with it through its computations, and provide them in the end with not only an answer, but also a *settled sense* of *why* it is the *right* answer. Providing such transparent warrant is an additional challenge for the use of artificial intelligence applications in a legal context – a challenge that, unfortunately, must be added to the usual difficulties of the modeling enterprise.