Applied Artificial Intelligence, 18:231–249, 2004

ISSN: 0883-9514 print/1087-6545 online
DOI: 10.1080/08839510490279861

Taylor & Francis
Taylor & Francis Group

CAUSAL INFERENCE IN LEGAL DECISION MAKING: EXPLANATORY COHERENCE VS. BAYESIAN NETWORKS

PAUL THAGARD
Philosophy Department, University of Waterloo, Waterloo, Ontario, Canada

Reasoning by jurors concerning whether an accused person should be convicted of committing a crime is a kind of casual inference. Jurors need to decide whether the evidence in the case was caused by the accused's criminal action or by some other cause. This paper compares two computational models of casual inference: explanatory coherence and Bayesian networks. Both models can be applied to legal episodes such as the von Bülow trials. There are psychological and computational reasons for preferring the explanatory coherence account of legal inference.

In December, 1980, Martha von Bülow, a very wealthy heiress, lapsed into a coma that still continues. In 1982, a jury found her husband, Claus von Bülow, guilty of two counts of assault with intent to murder. But an appeal granted him a new trial, and in 1985 he was acquitted on both counts. What was the nature of the inferences that led the first jury to find Claus von Bülow guilty and the second jury to find him not guilty?

This paper presents an analysis of jury decision making as a kind of causal inference. Members of the jury had to make inferences about the causes of Martha von Bülow's coma, in particular about whether it had been the result of an attempt by her husband to kill her. Reasoning about the testimony of witnesses is also causal, in that the jury had to infer whether some witnesses said what they did because they were in a position to know it was true, or because they were lying. In general, legal reasoning in trials such as those of Claus von Bülow's can be characterized as inference to the best overall causal story.

I am grateful to Ray Grondin for his undergraduate thesis on judicial inference, and to the Natural Sciences and Engineering Council of Canada for research support.

Address correspondence to Paul Thagard, Philosophy Department, University of Waterloo, 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada. E-mail: pthagard@uwaterloo.ca

There are currently available two computational models of this kind of causal inference, one based on Bayesian networks and the other based on explanatory coherence. I will argue that the explanatory-coherence account is superior both as a descriptive account of how jurors do reason and of how they should reason. After describing the causal structure of the von Bülow case, I present computational models of it using both explanatory coherence and Bayesian networks. In principle, both kinds of models can explain why the first jury found Claus von Bülow guilty and the second one found him innocent. However, the Bayesian account has serious problems of interpretation and implementation that make it unsatisfactory as an account of jury reasoning. Reflections on the nature of probability and causality show the superiority of the explanatory coherence account of causal inference in legal reasoning.

Abductive inference is a form that goes from data describing something to a hypothesis that best explains the data (Josephson and Josephson 1994). Explanatory coherence and Bayesian networks are two rich ways of specifying how abductive inference works. But the definition of abductive inference just given assumes that the relation between data and hypotheses is primarily explanation rather than conditional probability. To remain neutral for the moment between probabilistic and explanation-based approaches to abductive inference, I will begin by describing legal decisions as a kind of causal inference from observational evidence to hypotheses that propose causes of what was observed.

CAUSAL INFERENCE IN THE VON BÜLOW CASE

Claus von Bülow was tried twice for intent to murder his wife Martha, whose nickname was Sunny. My account of the trials is based primarily on the book by Harvard University law professor, Alan Dershowitz (1986), who was one of Claus von Bülow's lawyers for the appeal and second trial (see also Gribben 2001). The primary issue here is: Why did the jury in the first trial find him guilty, and why did the jury in the second trial find him not guilty?

According to Dershowitz (1986, p. 37), the prosecution's case in the first trial "was based heavily on hard scientific evidence, eyewitness testimony and compelling motives." The prosecution argued that Sunny's coma was the result of her being injected with insulin by her husband Claus. The most important witnesses were Sunny's maid, Maria Schrallhammer, and her son from a previous marriage, Alex von Auersperg. Maria testified that she had found a black bag of Claus's containing insulin in the month before Sunny went into a coma. Alex testified that after the coma he found the bag in Claus's closet, this time with three hypodermic needles, including one that had been used. Scientific tests found a residue of insulin on this needle.

Moreover, Sunny's blood after she was taken to hospital displayed a high insulin level, and it is well known that excess insulin can induce a coma.

The testimony by Maria, Alex, and scientific experts thus all supported the prosecution's causal story that Sunny's coma was the result of an insulin injection by Claus. The central causal hypothesis in the case is that Claus injected Sunny with insulin. No one directly observed this happening, so its plausibility depends largely on the indirect evidence for it: Sunny's coma, Claus having insulin, and Claus having a needle with insulin on it. Implicit in the acceptance of testimony is a causal judgment that witnesses say what they do because they believe it, and they believe it in part because it is true (Thagard forthcoming-a). Other causes are possible too, such as that the witness is mistaken or lying. For example, in the case of Maria, the jury inferred that she *said* she found the insulin in Claus's black bag because she *did* really find the insulin in the bag. In the first trial, the defense lawyers did not succeed in impugning the testimony of Maria or Alex, in contrast to the second trial where evidence was presented that they might not be telling the truth.

In the first trial, the prosecution was able to present evidence that gave Claus a strong motive to get rid of Sunny. His mistress, Alexandra Isles, testified that she had demanded that Claus divorce Sunny. Sunny's banker testified that Claus stood to gain a large inheritance if Sunny died, but would receive little if he divorced her. Thus the plausibility of the hypothesis that Claus tried to kill Sunny rested on there being a potential cause of his attempt, namely his romantic and pecuniary motive, as well as on the hypothetical effects of the attempt, including the needle with insulin on it and Sunny's insulin-based coma.

The defense tried to propose an alternative causal story of what produced Sunny's coma. They had a witness, Joy O'Neill, who said that she had frequently given Sunny exercise instruction. O'Neill said Sunny told her that insulin injection was a good way to avoid gaining weight. The defense tried to use this report to support their hypothesis that Sunny's coma was caused by self-injection of insulin. O'Neill's testimony was greatly weakened, however, when records of the exercise studio showed that O'Neill had taught Sunny much less than she had claimed, and had not taught her at all during the year that O'Neill claimed to have been told about insulin use. Given all the evidence that supported the prosecution's contention that Sunny was injected with insulin by Claus, it is not surprising that the jury found him guilty.

The second trial was very different from the first. Alan Dershowitz's appeal succeeded in getting defense access to notes collected by a private investigator hired by Alex von Auersperg. These notes showed that Maria had not mentioned finding insulin in Claus's bag until after Sunny's coma had been identified as insulin related. The prosecution thus suggested that Maria's testimony was caused by her dislike of Claus rather than by her having grounds to believe that there was insulin in Claus's bag. Moreover, Alex's

testimony was undermined by the revelation that a detective who had been with him when he found Claus's bag had not seen any needles in the bag. In addition, the defense called many scientific experts who challenged the reports in the first trial that Sunny's coma was insulin-induced and that the needle found with insulin on it had acquired the insulin during an injection. Finally, the alleged motive for Claus's attempted murder was undermined when Sunny's banker was not allowed to testify about how much Claus stood to inherit.

Having undermined the hypothesis that Sunny's coma was insulin-induced, the defense did not have to argue that she had injected herself. Rather, they presented a different story in which Sunny's many health problems and strange behaviors (ingesting huge amounts of aspirin, taking a variety of drugs, and eating ice cream sundaes even though she had blood sugar problems) could have been responsible for her coma. Thus the defense's causal story was much stronger than the self-injection story in the first trial, and the prosecution's story was much weaker. Accordingly, the jury in the second trial reached the verdict that Claus von Bülow was not guilty.

I hope this very brief summary of the two trials suffices to show that jurors' decisions were based on causal inferences. Should they conclude that an insulin injection by Claus had put Sunny in a coma, or should they judge that other causes could not be ruled out beyond a reasonable doubt? Lacking in my review so far, as well as in Dershowitz's (1986) much more detailed history, is any account of the inferential processes by means of which the jurors integrated the various pieces of information in the two trials in order to reach their verdicts. I will now show how the theory of explanatory coherence can provide such an account.

EXPLANATORY COHERENCE

The theory of explanatory coherence and the computational model ECHO have been applied to a great many examples of abductive inference in science, law, and everyday life (see, for example, Thagard 1989; 1992; 2000). The theory of explanatory coherence consists of the following principles:

Principle El. Symmetry. Explanatory coherence is a symmetric relation, unlike, say, conditional probability. That is, two propositions, p and q, cohere each other equally.

Principle E2. Explanation. (a) A hypothesis coheres with what it explains, which can either be evidence or another hypothesis; (b) hypotheses that together explain some other proposition cohere with each other; and (c) the more hypotheses it takes to explain something, the lower the degree of coherence.

Principle E3. Analogy. Similar hypotheses that explain similar pieces of evidence cohere.

Principle E4. Data priority. Propositions that describe the results of observations have a degree of acceptability on their own.

Principle E5. Contradiction. Contradictory propositions are incoherent with each other.

Principle E6. Competition. If *P* and *Q* both explain a proposition, and if *P* and *Q* are not explanatorily connected, then *P* and *Q* are incoherent with each other. (*P* and *Q* are explanatorily connected if one explains the other or if together they explain something.)

Principle E7. Acceptance. The acceptability of a proposition in a system of propositions depends on its coherence with them.

These principles do not fully specify how to determine coherence-based acceptance, but algorithms are available that can compute acceptance and rejection of propositions on the basis of coherence relations. The most psychologically natural algorithms use artificial neural networks that represent propositions by artificial neurons or *units* and represent coherence and incoherence relations by excitatory and inhibitory links between the units that represent the propositions. Acceptance or rejection of a proposition is represented by the degree of activation of the unit. The program ECHO spreads activation among all units in a network until some units are activated and others are inactivated, in a way that maximizes the coherence of all the propositions represented by the units.

Thagard (2000, pp. 30–31) describes a general algorithm for using an artificial neural network to solve constraint satisfaction problems such as explanatory coherence. For application of coherence as constraint satisfaction to causal inference in the law, think of a set of elements E as the set of propositions that represent hypotheses and evidence, and positive constraints C+ as the coherence relations established by explanation relations. Negative constraints C− are based on relations of contradiction or incompatibility between propositions as established by principles E5 and E6 above. We can then use the following algorithm to decide what causal hypotheses to accept or reject:

1. For every element e_i of E, construct a unit u_i which is a node in a network of units U. Such networks are very roughly analogous to networks of neurons.
2. For every positive constraint in C+ on elements e_i and e_j, construct a symmetric excitatory link between the corresponding units u_i and u_j. Elements whose acceptance is favored because they represent observed evidence can be positively linked to a special unit whose activation is clamped at the maximum value.

3. For every negative constraint in C− on elements e_i and e_j, construct a symmetric inhibitory link between the corresponding units u_i and u_j.
4. Assign each unit u_i an equal initial activation (say, .01), then update the activation of all the units in parallel. The updated activation of a unit is calculated on the basis of its current activation, the weights on links to other units, and the activation of the units to which it is linked. A number of equations are available for specifying how this updating is done (McClelland and Rumelhart 1989). For example, on each cycle the activation of a unit *j*, a_j, can be updated according to the following equation:

$$a_j(t+1) = a_j(t)(1-d) + net_j(max - a_j(t)) \text{ if } net_j > 0,$$
$$\text{otherwise } net_j(a_j(t) - min).$$

   Here *d* is a decay parameter (say, .05) that decrements each unit at every cycle, *min* is a minimum activation (−1), and *max* is maximum activation (1). Based on the weight w_{ij} between each unit *i* and *j*, we can calculate net_j, the net input to a unit, by:

$$net_j = \Sigma_i w_{ij} a_i(t).$$

   Although all links in coherence networks are symmetrical, the flow of activation is not, because a special unit with activation clamped at the maximum value spreads activation to favored units linked to it, such as units representing evidence in the explanatory coherence model ECHO. Typically, activation is constrained to remain between a minimum (e.g., −1) and a maximum (e.g, 1).
5. Continue the updating of activation until all units have settled—achieved unchanging activation values. If a unit u_i has final activation above a specified threshold (e.g., 0), then the element e_i represented by u_i is deemed to be accepted. Otherwise, e_i is rejected.

This algorithm is psychologically natural in that it views inference as analogous to neurological processes in which multiple neurons interact in parallel. But other algorithms are available for solving constraint satisfaction problems, such as the following greedy algorithm (Thagard 2000, p. 35; compare Selman et al. 1992):

1. Randomly assign the elements of E into A (accepted) or R (rejected).
2. For each element e in E, calculate the gain (or loss) in the weight of satisfied constraints that would result from flipping e, i.e., moving it from A to R if it is in A, or moving it from R to A otherwise.
3. Produce a new solution by flipping the element that most increases coherence, i.e., move it from A to R or from R to A. In case of ties, choose randomly.

4. Repeat 2 and 3 until either a maximum number of tries have taken place or until there is no flip that increases coherence.

This algorithm usually produces the same acceptances and rejections as the connectionist algorithm; exceptions arise from the random character of the initial assignment in step 1 and from the greedy algorithm breaking ties randomly. LISP code for ECHO is available on my Web site (Thagard 2002).

APPLICATION TO THE VON BÜLOW CASE

To apply the theory of explanatory coherence and the computational model ECHO to the von Bülow case, it is necessary to express the causal relations described in the section "causal inference in the von Bülow case" as explanations. ECHO takes input such as (explains (H1 H2)E1), which signifies that hypotheses H1 and H2 together explain evidence E1. Appendices A and B list all the input given to ECHO for the first and second von Bülow trials, respectively. The structure of the constraint network produced by ECHO is most easily understood graphically, as shown in Figure 1 for the first trial. The relation *explains* is asymmetrical, but ECHO establishes a symmetrical link between a hypothesis and what it explains. Quine and Ullian (1970, p. 79) argued: "We see therefore that there can be mutual reinforcement between an explanation and what it explains. Not only does a supposed truth gain credibility if we can think of something that would explain it, but also conversely: an explanation gains credibility if it accounts for something we suppose to be true."

Note how ECHO naturally encodes the two competing causal stories about why Sunny went into a coma. The theory of explanatory coherence and the computational model of ECHO are highly compatible with the predominant psychological theory of jury decisions, according to which jurors choose between competing stories of what happens (Pennington and Hastie 1992; 1993; see also Byrne 1995).

What is the relation between the concept *explains* that ECHO takes as primitive and the concept *cause* that I used to describe the von Bülow case in the earlier section? In the philosophy of science, there are several competing conceptions of explanation (see, e.g., Salmon 1989 and Thagard 1992, ch. 5). The one I prefer ties explanation directly to cause: A proposition A is part of the explanation of a proposition B if the entities and their properties described by A are part of a causal process that produces the properties of the entities described by B. For example, I interpret the claim that Claus's injecting Sunny with insulin explains her coma as saying that there was a causal process that started with Claus, the needle, insulin, and the act of injection, and ended with Sunny unconscious. We need not know all the steps in the relevant causal process in order to have an explanation, but we can at least

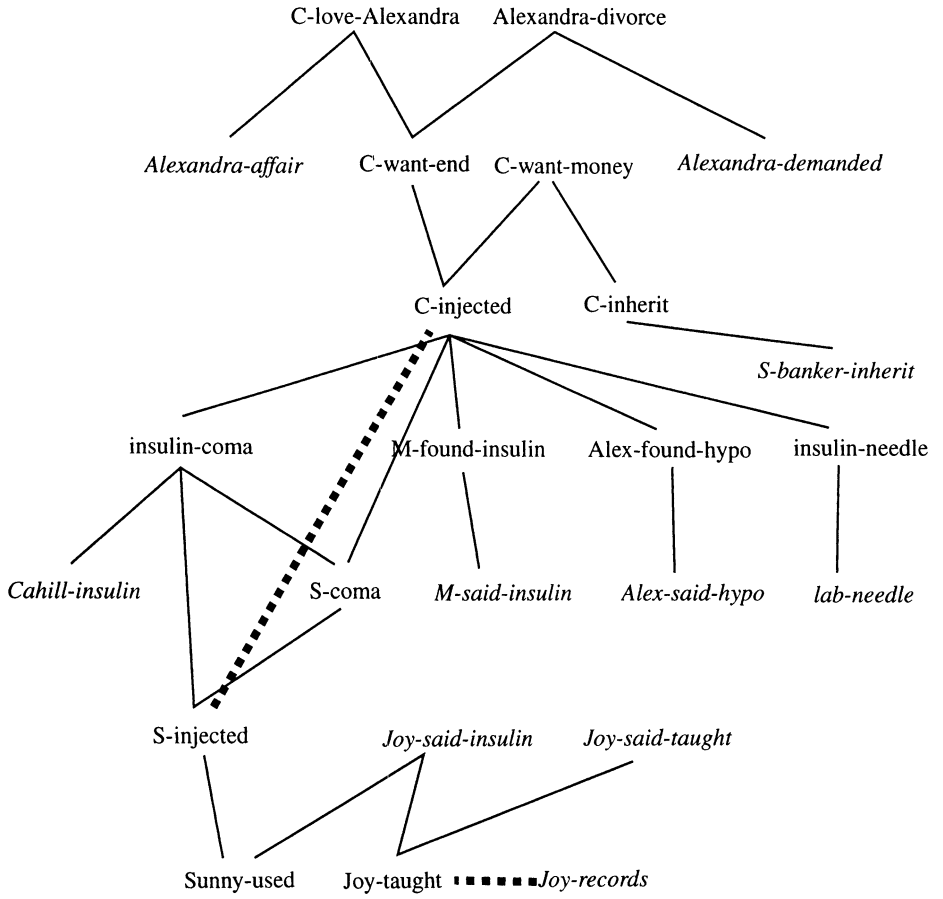

FIGURE 1. Graphical structure of the constraint network created by ECHO for the input presented in Appendix A. Solid lines indicate coherence relations established by explanations, while dotted lines indicate incoherence relations established by contradiction and competition. Data nodes are indicated by italics.

sketch some of the mechanisms such as injection and the body's reaction to insulin that lead from the explanatory cause to the explained effect.

For the first trial, ECHO ends up accepting the hypothesis that Claus injected Sunny with insulin, but for the second trial ECHO ends up rejecting it. This result occurs regardless of whether the neural network or greedy algorithms are used to maximize constrain satisfaction. To simulate decision making in the first trial, the connectionist algorithm requires 188 cycles of updating activations before the network has settled, and the greedy algorithm requires around 16 flips to reach the same partition of propositions into accepted and rejected. The connectionist algorithms can easily handle much larger networks. My student Ray Grondin developed a simulation of the 1684 trial of Laurence Braddon, analyzed by Wigmore (1937), which

involved more than 200 propositions and still settled in a few minutes on a laptop computer.

Appendix B shows the input to ECHO used to simulate the second trial. Notice that it undermines the testimony of Maria and Alex by providing alternative explanations of why they said what they did about insulin and hypodermic needles in Claus's bag. Most important, it includes new evidence in the form of expert testimony that Sunny's coma was not produced by insulin rejection. Moreover, the defense's explanation of what put Sunny into a coma is better supported than the self-injection explanation in the first trial, and the prosecution's description of Claus's motives is weaker. All of these factors contribute to ECHO's rejection of the hypothesis that Claus injected Sunny.

How subjective is the analysis of two trials presented in the appendices? It may seem that ECHO simulations require many numerical values such as excitation, inhibition, and decay that depend on arbitrary decisions by the programmer. In fact, however, I use the same numerical values (e.g., 04 for excitation, −.06 for inhibition) in all ECHO runs, and sensitivity analyses have shown that the actual values do not much matter as long as excitation is greater than inhibition. More problematic is specification of the "explains" relations which requires the programmer to understand the causal structure of the case. But the same understanding is required for simulations using Bayesian networks discussed below. Marking a proposition as "data" is not arbitrary: In the legal context, the data are the utterances made by witnesses that are observed by everyone in the courtroom.

Thus the program ECHO has successfully modeled the decisions of the juries in both the first and second trial. Along with the theory of explanatory coherence, it shows how different kinds of causal factors can be integrated into a single judgment about whether an accused is innocent or guilty. Previous application of ECHO to legal cases can be found elsewhere (Thagard 1989; 2003). The latter paper concluded that explanatory coherence was *not* the best explanation of the judgment by the jury in the 1995 O. J. Simpson trial, arguing that the judgment was partly a matter of *emotional* coherence deriving from juror bias in favor of Simpson and against the Los Angeles Police Department. I have found no suggestions that the von Bülow juries were biased either for or against him, so the explanatory coherence account appears adequate for both of his trials.

BAYESIAN NETWORKS

There is, however, a distinguished alternative account of legal inference, using Bayesian networks. Kadane and Schum (1996) argue that legal trials can be understood using probability theory and they present a comprehensive and detailed analysis of the famous 1921 trial of the accused

anarchists, Sacco and Vanzetti. In the past decade and a half, Bayesian networks have become increasingly influential in artificial intelligence (Pearl 1988; 2000). It is necessary to consider, therefore, whether this approach provides a plausible alternative account of juror inference in the von Bülow trials.

Bayesian networks consist of a network of nodes, some pairs of which are joined by arrows. The arrows indicate relations of probabilistic dependence: If A - > B, then the probability of B depends in part on the probability of A. In addition, the arrows can be interpreted as causal relations: If A - > B, then A causally influences B. Given this second interpretation, it is natural to analyze legal trials as Bayesian networks. Figure 2 shows a Bayesian network built using the programming tool JavaBayes (Cozman 2001). This network was constructed using nodes for the propositions in the ECHO analysis of the first von Bülow trial (see Appendix A and Figure 1). The major difference between the Bayesian network in Figure 2 and the coherence network in Figure 1 is that the connections in the Bayesian network are unidirectional.

In addition to constructing the causal network shown in Figure 2, the Bayesian analysis of the first von Bülow trial and others requires specifying many conditional probabilities. For each node that has *n* arrows coming into it, it is necessary to specify 2^{n+1} conditional probabilities. For example, the node *Alexandra-affair* has one arrow coming down to it from *C-love-Alexandra,* so the Bayesian simulation requires the conditional probabilities:

P(Alexandra-affair is true/C-love-Alexandra is true)
P(Alexandra-affair is false/C-love-Alexandra is true)
P(Alexandra-affair is false/C-love-Alexandra is false)
P(Alexandra-affair is false/C-love-Alexandra is false)

Thus, if jurors are Bayesian networks, they would need to have some estimate of the probability of what Alexandra says about Claus given that she says it. In constructing the JavaBayes network, I had no idea what this probability would be, but I figured that in general witnesses are reliable so I guessed that

P(Alexandra-affair is true/C-love-Alexandra is true) = .7
P(Alexandra-affair is false/C-love-Alexandra is true) = .3.

I had no idea what the conditional probabilities might be if C-love-Alexandra is false, so I just left them at the default values of .5.

Even more problematic was coming up with conditional probabilities in cases where there are two arrows coming into a node. For example, the node *C-injected* required eight conditional probabilities such as:

P(C-injected is true / C-want-end is true & C-want-money is true).

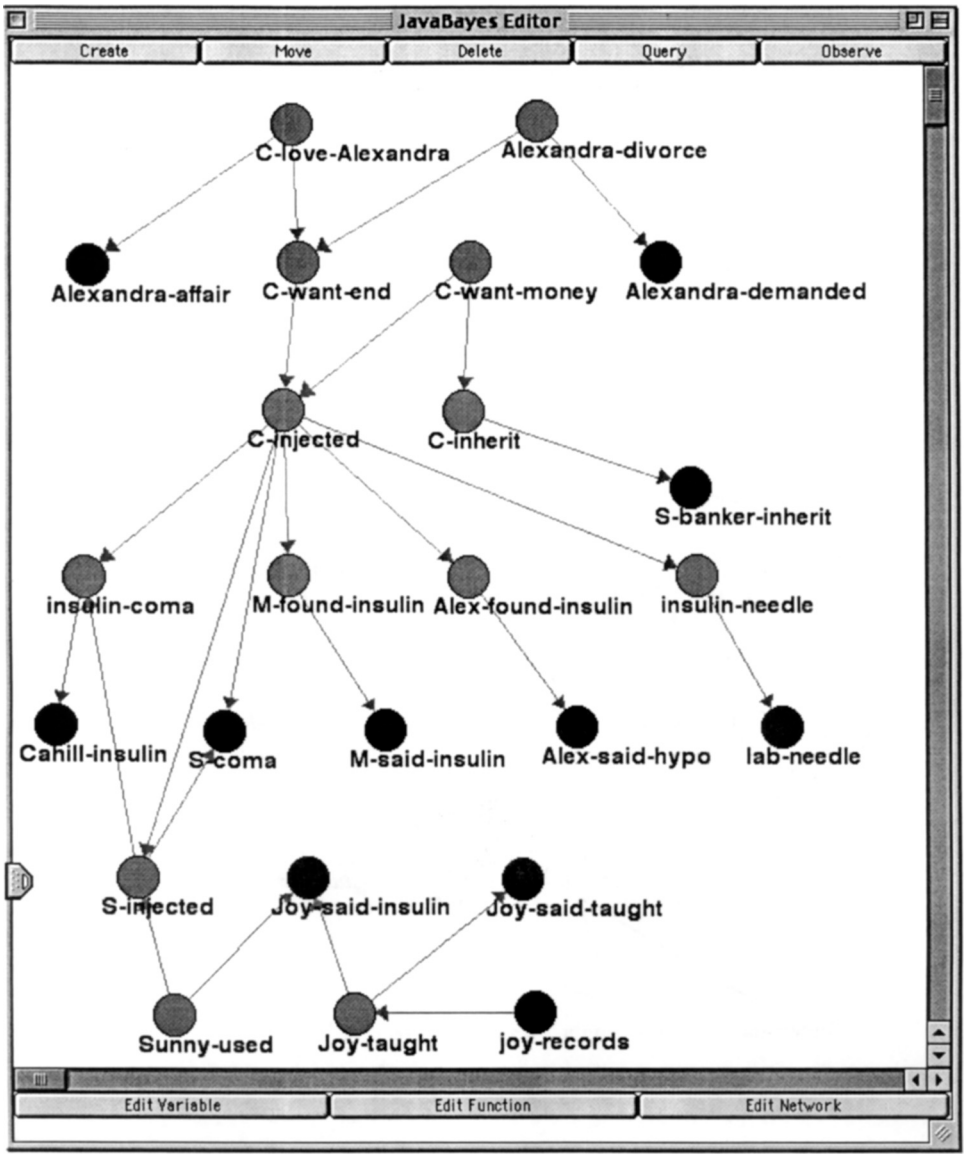

FIGURE 2. Bayesian network analysis of the first von Bülow trial produced using JavaBayes (Cozman 2001). Arrows indicate conditional probability, for example, that Alexandra-divorce is conditional on Alexandra-demanded. Dark nodes are observed to be true.

Given the causal link, it was plausible to give this my default high-probability value of .7, but for most of the Boolean combinations, I had no idea what the conditional probabilities might be and left them at the default value of .5.

Nevertheless, the probability values I inserted were sufficient to simulate jury behavior in the first von Bülow trial. JavaBayes uses the provided conditional probabilities and the information that some nodes were marked as observed to be true to calculate the posterior probability of each nodes. My simulation resulted in *C-injected* having a probability of greater than .5 and in *S-injected* having a probability of less than .5. Thus it seems that JavaBayes, like the jury in the first trial, found von Bülow guilty. I have not done a JavaBayes simulation of the second trial, but I expect that it could be made to find von Bülow innocent, perhaps with some adjustment of conditional probabilities. I have assumed that "guilty beyond a reasonable doubt" means something like "probability above a specified value such as .5"; reasonable doubt is a complex notion that I discuss elsewhere (Thagard forthcoming-b).

However, I have serious doubts about the Bayesian explanation of juror reasoning in these two trials. These doubts derive from what I shall call the *interpretation problem* and the *implementation problem.* The interpretation problem is that there is no plausible meaning for the probabilities used in the Bayesian simulation. The network shown in Figure 2 is unproblematic if the arrows are interpreted as causal relations. The coherence relations in Figure 1 are also based on causal relations that underpin explanation relations. But Bayesian networks require also that the arrows have a probabilistic interpretation so that conditional probabilities can be specified. Otherwise, the algorithms for calculating probabilities used by JavaBayes and similar programs have no application. I will now argue that there is no satisfactory interpretation of the probabilities that would be needed for legal applications.

As Hacking (2001) reviews, there are two kinds of interpretations of probability: frequency-type and belief-type. Frequency-type interpretations apply best to cases such as games of chance and large databases. For example, we can say that the probability of rain in Waterloo on a July day is x because the frequency of days with rain over the past 50 years of collecting records, that is the ratio of days with rain to all days in July, is x. It is obvious that frequency-type probabilities are irrelevant to the von Bülow trial. Nobody, including members of the jury, has frequency information either for single hypotheses such as *C-injected* or for conditional probabilities such as P (C-injected/C-want-end).

Proponents of Bayesian networks such as Pearl (2000, p. 2) usually endorse belief-type probabilities, according to which probabilities encode degrees of belief about events in the world. At first glance, this seems like a reasonable way to think about legal inference. Jurors will find an accused guilty if they believe the prosecution's story to a sufficient degree, i.e., with high probability. But there is abundant psychological evidence that people's degrees of belief do not conform to the calculus of probability (e.g.,

Kahneman et al. 1982; Gilovich et al. 2002). Rather than think of degrees of belief in nonfrequency matters as like probabilities in ranging from 0 to 1, it is more psychologically plausible to view them as a small number of qualitative states, perhaps the following: strongly believe – weakly believe – no belief – weakly doubt – strongly doubt. The mathematical range of probabilities is excellent for expressing frequencies, but does not map well at all onto degrees of belief.

Explanatory coherence theory does not attach any special significance to the numerical activations arrived at by different nodes in the neural networks that it uses to maximize coherence: The crucial issue is whether a proposition is accepted or not. In contrast, the Bayesian model requires an interpretation of probabilities as either frequencies or degrees of belief, neither of which is plausible in the context of legal decision making.

By the *implementation problem* for Bayesian networks I mean the difficulty of coming up with all the conditional probabilities that the analysis requires. My simulation of the first von Bülow trial required the specification of 96 conditional probabilities, and the numbers I came up with are largely arbitrary. The problem was even more serious for the much larger JavaBayes simulation that Ray Grondin did of the Braddon trial, which required many hundreds of conditional probabilities that he simply had to make up.

If the actual numbers do not much matter, as suggested by my use of a few default values for conditional probabilities in the JavaBayes simulation of the von Bülow trial, then the apparatus of probability theory is largely superfluous. All that really matters is causal structure, and explanatory coherence theory captures that without dealing at all with probabilities. On the other hand, an honest Bayesian has to be willing to ascribe to jurors and other people a great many numerical probabilities, and there is no principled way to do this. We might ask jurors to compare their degrees of belief in a proposition to outcomes in games of chance where probabilities are well defined, but there is no clear psychological mapping between beliefs such as *Claus injected Sunny* and rolls of a dice.

CONCLUSION

Because of the interpretation and implementation problems, I judge the Bayesian network account of juror inference to be less plausible than the explanatory coherence account. Both approaches capture the insight that legal decisions such as whether an accused is guilty depend on causal inference. But the explanatory coherence account better captures the insight of psychologists such as Pennington and Hastie (1993) and legal scholars such as Dershowitz (1986) and Allen (1997) that jurors decide by choosing between competing stories about what went on in an alleged

crime. Such stories are full of causal structure, concerning both what happened in the crime situation and what happened in court when witnesses made their testimonies. Legal inference is inference to the most plausible causal story, but the psychological mechanism by which jurors evaluate causal stories seems based, not on Bayesian probability calculations, but on explanatory coherence.

REFERENCES

Allen, R. J. 1997. Rationality, algorithms, and juridical proof: A preliminary inquiry. *International Journal of Evidence and Proof* 1:253–360.

Byrne, M. D. 1995. The convergence of explanatory coherence and the story model: A case study in juror decision. In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, eds. J. D. Moore and J. F. Lehman, pages 539–543. Erlbaum.

Cozman, F. J. 2001. JavaBayes: Bayesian networks in Java. http://www-2.cs.cmu.edu/ ~ javanbayes/ (viewed August 2002).

Dershowitz, A. M. 1986. *Reversal of Fortune: Inside the von Bülow Case.* New York: Random House.

Gilovich, T., D. Griffin, and D. Kahneman eds. 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment.* Cambridge: Cambridge University Press.

Gribben, M. 2001. The Claus Von Bulow Case. *http://www.crimelibrary.com/classics6/bulow/*(consulted July 30, 2002).

Hacking, I. 2001. *An Introduction to Probability and Inductive Logic.* Cambridge: Cambridge University Press.

Josephson, J. R., and S. G. Josephson eds. 1994. *Abductive Inference: Computation, Philosophy, Technology.* Cambridge: Cambridge University Press.

Kadane, J. B., and D. A. Schum. 1996. *A Probabilistic Analysis of the Sacco and Vanzetti Evidence.* New York: John Wiley and Sons.

Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment Under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press.

McClelland, J. L., and D. E. Rumelhart. 1989. *Explorations in Parallel Distributed Processing.* Cambridge, MA: The MIT Press.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems.* San Mateo: Morgan Kaufman.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Pennington, N., and R. Hastie. 1992. Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology* 51:189–206.

Pennington, N., and R. Hastie. 1993. Reasoning in explanation-based decision making. *Cognition* 49:125–163.

Quine, W. V. O., and J. Ullian. 1970. *The Web of Belief.* New York: Random House.

Salmon, W. C. 1989. Four decades of scientific explanation. In *Scientific Explanation (Minnesota Studies in the History of Science., vol. XIII)* eds. P. Kitcher, and W. C. Salmon, 3-219. Minneapolis: University of Minnesota Press.

Selman, B., H. Levesque, and D. Mitchell. 1992. A new method for solving hard satisfiability problems. In *Proceedings of the Tenth National Conference on Artificial Intelligence,* pages 440–446. AAAI Press.

Thagard, P. 1989. Explanatory coherence. *Behavioral and Brain Sciences* 12:435–467.

Thagard, P. 1992. *Conceptual Revolutions.* Princeton: Princeton University Press.

Thagard, P. 2000. *Coherence in Thought and Action.* Cambridge, MA: The MIT Press.

Thagard, P. 2002. Computational epistemology laboratory. *http://cogsci.uwaterloo.ca* (viewed August 2002).

Thagard, P. (forthcoming-a). Testimony, credibility, and explanatory coherence.

Thagard, P. (forthcoming-b) What is doubt and when is it reasonable? In M. Ezcurdia, R. Stainton & C. Viger (eds.), New Essays in the Philosophy of Language and Mind. *Canadian Journal of Philosophy*, Supplementary Volume. Calgary: University of Calgary Press.

Thagard, P. 2003. Why wasn't O. J. convicted? Emotional coherence in legal inference. *Cognition and Emotion*, 17:361–383.

Wigmore, J. H. 1937. *The Science of Judicial Proof as Given by Logic, Psychology, and General Experience and Illustrated in Judicial Trials,* 3rd edition. Boston: Little Brown.

APPENDIX A

Input for ECHO simulation of Trial I

;Evidence:
(proposition 'S-coma ''Sunny went into a coma.'')
(proposition 'Maria-said-insulin ''Maria said she found insulin in Claus's bag.'')
(proposition 'Alex-said-hypo ''Alex said he found used hypodermic in Claus's bag.'')
(proposition 'Cahill-insulin ''Cahil said insulin put Sunny in coma.'')
(proposition 'lab-insulin ''Lab reported insulin on used hypodermic.'')
(proposition 'S-banker-inherit ''Sunny's banker said Claus would inherit $14 million.'')
(proposition 'Alexandra-affair ''Alexandra said she was having an affair with Claus.'')
(proposition 'Alexandra-demanded ''Alexandra said she demanded Claus divorce Sunny.'')
(proposition 'Joy-said-insulin ''Joy said Sunny recommended insulin.'')
(proposition 'Joy-said-taught ''Joy said she taught Sunny many times.'')
(proposition 'Joy-records ''Records showed Joy hardly taught Sunny.'')

;Prosecution hypotheses:
(proposition 'C-loved-Alexandra ''Claus loved Alexandra.''); p. xxi
(proposition 'Alexandra-divorce ''Alexandra demanded Claus divorce Sunny'')
(proposition 'C-want-end ''Claus wanted to end his marriage to Sunny.'')
(proposition 'C-want-money ''Claus wanted to inherit money.'')
(proposition 'C-injected ''Claus injected Sunny with insulin.'')
(proposition 'M-found-insulin ''Maria had earlier found insulin in Claus's bag.'')
(proposition 'Alex-found-hypo ''Alex found hypodermic needle in Claus's bag.'')
(proposition 'insulin-needle ''Insulin found hypodermic needle.'')
(proposition 'insulin-coma ''Insulin put Sunny into a coma.'')

(proposition 'C-inherit ''Claus would inherit $14 million'')
(proposition 'Joy-lied ''Joy lied that Sunny recommended insulin.'')

;Defense hypothees
(proposition 'S-injected ''Sunny injected herself with insulin.'')
(proposition 'Sunny-used ''Sunny used insulin.'')
(proposition 'Joy-taught ''Joy taught Sunny many times.'')

;Contradictions
(contradict 'Joy-taught 'Joy-records)

;Prosecution explanations:
(explain '(insulin-coma)'S-coma)
(explain 'insulin-coma) 'Cahill-insulin)
(explain '(C-injected) 'insulin-coma)
(explain '(C-injected) 'S-coma)
(explain '(C-want-end) 'M-found-insulin)
(explain '(C-injected) 'insulin-needle)
(explain '(C-injected) 'Alex-found-hypo)
(explain '(Alex-found-hypo) 'Alex-said-hypo)
(explain '(M-found-insulin) 'M-said-insulin)
(explain '(insulin-needle) 'lab-insulin)

;motive
(explain '(C-love-Alexandra Alexandra-divorce) 'C-want-end)
(explain '(C-love-Alexandra) 'Alexandra-affair)
(explain '(Alexandra-divorce) 'Alexandra-demanded)
(explain '(C-want-end) 'C-treated-bad)
(explain '(C-want-end C-want-money) 'C-injected)
(explain '(C-want-money) 'C-inherit)
(explain '(C-inherit) 'S-banker-inherit)

;Defense explanations:
(explain '(S-injected) 'insulin-coma)
(explain '(S-injected) 'S-coma)
(explain '(Sunny-used) 'S-injected)
(explain '(Sunny-used Joy-taught) 'Joy-said-insulin)
(explain '(Joy-taught) 'Joy-said-taught)

;Data
(data '(S-coma Maria-said-insulin Maria-said-bad Alex-said-hypo Cahill-insulin lab-insulin S-banker-inherit Alexandra-affair Alexandra-demanded Joy-said-insulin Joy-said-taught Joy-records))

APPENDIX B

Input for ECHO simulation of Trial 2

;Evidence:
(proposition 'S-coma ''Sunny went into a coma.'')
(proposition 'Maria-said-insulin ''Maria said she found insulin in Claus's bag.'')
(proposition 'Maria-notes ''Maria did not mention insulin in Kuh's notes.'')
(proposition 'Alex-said-hypo ''Alex said he found used hypodermic in Claus's bag.'')
(proposition. 'Detective-vs-Alex ''Detective said Alex did not find needle.''); 107
(proposition 'Cahill-insulin ''Cahil said insulin put Sunny in coma.'')
(proposition 'lab-insulin ''Lab reported insulin on used hypodermic.'')
(proposition 'S-banker-inherit ''Sunny's banker said Claus would inherit $14 million.'')
(proposition 'Alexandra-affair ''Alexandra said she was having an affair with Claus.'')
(proposition 'Alexandra-demanded ''Alexandra said she demanded Claus divorce Sunny.'')
(proposition 'Cortivo-said-needle-not-injected ''Cortivo said needle not injected.'');202
(proposition 'Rubenstein-said-not-insulin ''Rubenstein said tests did not show high insulin.''); 202
(proposition 'Galitis-said-not-insulin-coma ''Dr. Galitis said it was not an insulin coma.'')

;Prosecution hypotheses:
(proposition 'C-loved-Alexandra ''Claus loved Alexandra.''); p. xxi
(proposition 'Alexandra-divorce ''Alexandra demanded Claus divorce Sunny'')
(proposition 'C-want-end ''Claus wanted to end his marriage to Sunny.'')
(proposition 'C-want-money ''Claus wanted to inherit money.'')
(proposition 'C-injected ''Claus injected Sunny with insulin.'')
(proposition 'M-found-insulin ''Maria had earlier found insulin in Claus's bag.'')
(proposition 'Alex-found-hypo ''Alex found hypodermic needle in Claus's bag.'')
(proposition 'insulin-needle ''Insulin found on hypodermic needle.'')
(proposition 'insulin-coma ''Insulin put Sunny into a coma.'')

;Defense hypothees
(proposition 'S-coma-noninsulin ''Sunny went into a coma for non-insulin reasons.'')
(proposition 'Sunny-health ''Sunny had many health problems.'')
(proposition 'Sunny-behavior ''Sunny had many weird health behaviors.'')
(proposition 'M-lied ''Maria lied about Claus.'')
(proposition 'A-lied ''Alex lied about hypodermic.'')
(proposition 'needle-not-injected ''The needle found in Claus's bag was not injected.'')
(proposition 'not-insulin-coma ''Sunny's coma was not insulin induced.'')

;Prosecution explanations:
(explain '(insulin-coma)' 'S-coma)
;(explain '(insulin-coma)' 'Cahill-insulin)
(explain '(C-injected) 'S-coma)
(explain '(C-injected) 'insulin-coma)
(explain '(C-want-end) 'M-found-insulin)
(explain '(C-injected) 'insulin-needle)
(explain '(C-injected) Alex-found-hypo)'
(explain '(Alex-found-hypo) 'Alex-said-hypo)
(explain '(M-found-insulin) 'M-said-insulin)
(explain '(insulin-needle) 'lab-insulin)

;motive
(explain '(C-love-Alexandra Alexandra-divorce) 'C-want-end)
(explain '(C-love-Alexandra) 'Alexandra-affair)
(explain '(Alexandra-divorce) 'Alexandra-demanded)
(explain '(C-want-end C-want-money) 'C-injected)
;(explain '(C-want-money) 'C-inherit
;(explain '(C-inherit) 'S-banker-inherit); ruled out, p. 200

;Defense explanations:
(explain '(not-insulin-coma) 'S-coma)
(explain '(Sunny-health Sunny-behavior) 'S-coma-noninsulin)
(explain '(M-lied) 'Maria-said-insulin)
(explain '(M-lied) 'Maria-notes)
(explain '(A-lied) 'Alex-said-hypo)
(explain '(A-lied) 'Detective-vs-Alex)
(explain '(needle-not-injected) 'Cortivo-said-needle-not-injected)
(contradict 'needle-not-injected 'C-injected)
(explain '(not-insulin-coma) 'Rubenstein-said-not-insulin)
(contradict 'not-insulin-coma 'insulin-coma)
(explain '(not-insulin-coma) 'Galitis-said-not-insulin-coma)

```
;Data
(data '(S-coma Maria-said-insulin Maria-said-bad Alex-said-hypo
Cahill-insulin lab-insulin S-banker-inherit Alexandra-affair Alexandra-
demanded Maria-notes Detective-vs-Alex Rubenstein-said-not-insulin
Cortivo-said-needle-not-injected Galitis-said-not-insulin-coma Sunny-
health Sunny-behavior))
```