Why Wasn't O. J. Convicted?
Emotional Coherence in Legal Inference

Paul Thagard
Philosophy Department
University of Waterloo
http://cogsci.uwaterloo.ca
pthagard@uwaterloo.ca
October, 2001

In 1995, O. J. Simpson was tried for the murder of his ex-wife, Nicole Brown Simpson, and her friend, Ron Goldman, both of whom had been found with multiple knife wounds. To the surprise of many, the jury found Simpson not guilty of the crime, and many explanations have been given for the verdict, ranging from emotional bias on the part of the jury to incompetence on the part of the prosecution. Of course, there is also the possibility that, given the evidence presented to them, the jury rationally made the decision that Simpson was not guilty beyond a reasonable doubt.

This paper evaluates four competing psychological explanations for why the jury reached the verdict they did:

1. Explanatory coherence. The jury found O. J. Simpson not guilty because they did not find it plausible that he had committed the crime, where plausibility is determined by explanatory coherence.

2. Probability theory. The jury found O. J. Simpson not guilty because they thought that it was not sufficiently probable that he had committed the crime, where probability is calculated by means of Bayes's theorem.

3. Wishful thinking. The jury found O. J. Simpson not guilty because they were emotionally biased toward him and wanted to find him not guilty.

4. Emotional coherence. The jury found O. J. Simpson not guilty because of an interaction between emotional bias and explanatory coherence.

I will describe computational models that provide detailed simulations of juror reasoning for explanatory and emotional coherence, and argue that the latter account is the most plausible. Application to the Simpson case requires expansion of my previous theory of

emotional coherence to introduce emotional biasing of judgments of explanatory coherence.

Social psychologists distinguish between "hot" and "cold" cognition, which differ in that the former involves motivations and emotions (Abelson, 1963; Kunda, 1999). The first two explanations above involve cold cognition, the third based on wishful thinking involves only hot cognition, but my preferred emotional-coherence explanation shows how hot and cold cognition can be tightly integrated.

Explanatory Coherence

At first glance, the evidence that O. J. Simpson was the murderer of his ex-wife was overwhelming. Shortly after the time that the murder took place, he caught a plane to Chicago carrying a bag that disappeared, perhaps because it contained the murder weapon and bloody clothes. Police who came to Simpson's house found drops of blood in his car that matched his own blood and that of Ron Goldman. In Simpson's back yard, police found a bloody glove that was of a pair with one found at the scene of the crime, and they found a bloody sock in his bedroom. Simpson had a cut on his hand that might have been caused by a struggle with the victims who tried to defend themselves. Simpson's blood was found on a gate near the crime scene. Moreover, there was a plausible motive for the murder, in that Simpson had been physically abusive to his wife while they were married and was reported to have been jealous of other men who saw Nicole after their divorce.

Based on all this evidence, many people judged that Simpson was guilty. One way of understanding this judgment is in terms of the theory of explanatory coherence, which I developed to explain how scientists evaluate competing theories but which has also been applied to legal and other kinds of reasoning (Thagard, 1989, 1992, 1999, 2000). On this theory, a hypothesis such as the claim that Simpson killed Nicole is accepted if doing so maximizes the overall coherence among pieces of evidence and the conflicting hypotheses that compete to explain the evidence. The theory of explanatory

coherence can be summarized in the following principles, discussed at length elsewhere (Thagard, 1992, 2000).

Principle E1. Symmetry. Explanatory coherence is a symmetric relation, unlike, say, conditional probability. That is, two propositions *p* and *q* cohere with 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.

The theory of explanatory coherence is implemented in a computational model, ECHO, that shows precisely how coherence can be calculated. Hypotheses and evidence are represented by units, which are highly simplified artificial neurons that can have excitatory and inhibitory links with each other. When two propositions cohere, as when a hypothesis explains a piece of evidence, then there is an excitatory link between the two units that represent them. When two propositions are incoherent with each other, either because they are contradictory or because they compete to explain some of the evidence, then there is an inhibitory link between them. Standard algorithms are available for

spreading activation among the units until they reach a stable state in which some units have positive activation, representing the acceptance of the propositions they represent, and other units have negative activation, representing the rejection of the propositions they represent. Thus algorithms for artificial neural networks can be used to maximize explanatory coherence, as can other kinds of algorithms (Thagard and Verbeurgt, 1998; Thagard, 2000).

Figure 1 shows the structure of an explanatory-coherence account of why O. J. Simpson might be judged guilty. The hypothesis that he was the killer explains why Nicole Simpson and Ron Brown are dead, why Simpson's blood was found on a gate at the crime scene, why there was blood in his car, why a bloody glove was found in his yard, and why his sock had blood on it. Moreover, there is an explanation of why Simpson killed Nicole based on his past history of abuse and jealousy. In the computational model ECHO, the principle of data priority, E4, is implemented by spreading activation directly to units representing evidence, from which activation spreads to the unit representing the hypothesis that Simpson was the murderer. Given the inputs shown in figure 1, ECHO activates this unit and finds the accused guilty.

In the criminal trial, Simpson was represented by a stellar team of 14 lawyers. who needed to convince the jury that there was reasonable doubt whether Simpson was guilty. They realized that they needed to provide alternative explanations of the apparently damning evidence that implicated Simpson as the murderer. According to Schiller and Willwerth (1997, p. 417), the defense lawyers were familiar with the story model of juror decision making (Pennington and Hastie, 1992, 1993). On this model jurors reach their decisions based on whether the prosecution or the defense presents a more compelling story about the events of the case. One of Simpson's main attorneys, Johnnie Cochran wrote (1997, pp. 236-237):

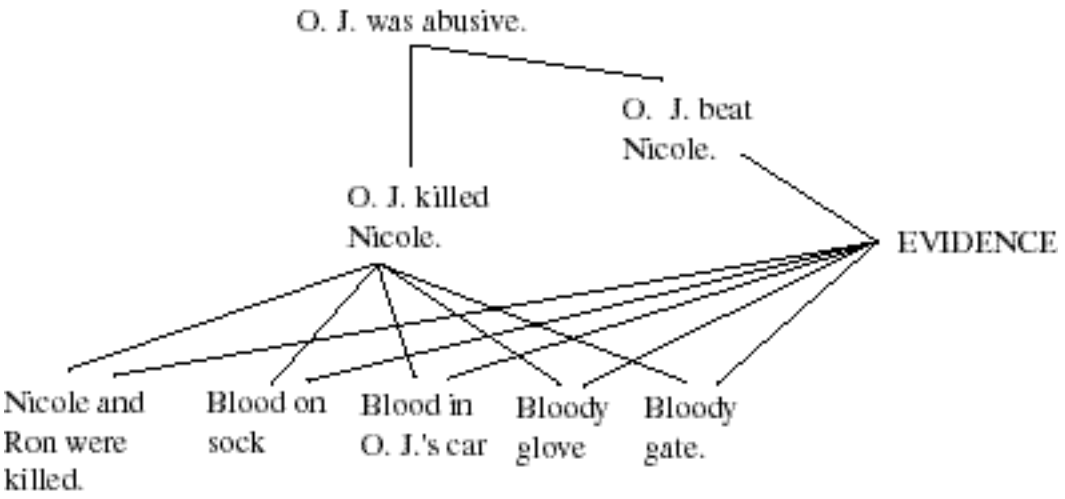

Figure 1. Part of the evidence supporting the hypothesis that O. J. Simpson killed his ex-wife. Solid lines indicate coherence relations.

> Whatever the commentators may say, a trial is not really a struggle between opposing lawyers but between opposing stories. … What juries require is a story into whose outline they can plug the testimony and evidence with which they are relentlessly bombarded.

As Byrne (1995) has argued, the story model of juror reasoning can be viewed as an instantiation of the theory of explanatory coherence, which provides a fuller and more rigorous account of what it is for one story to be more plausible than another. In accord with the theory of explanatory coherence, the defense lawyers set out to generate and support hypotheses that explained the deaths and other evidence using hypotheses that would compete with the hypothesis of Simpson's guilt.

The first task of the defense lawyers was to generate an alternative explanation of who killed Nicole Simpson and Ron Goldman. Based on Nicole's known history of

cocaine use, they hypothesized that she was killed by drug dealers, and argued that a more thorough police investigation right after the murders would have turned up evidence that supported this explanation. In order to explain the circumstantial evidence linking O. J. Simpson to the crime scene, including the bloody car, glove, and sock, the defense contended that the items had been planted by Los Angeles Police Department officers determined to frame Simpson for the crime. With the help of a strong team of forensic experts, the lawyers were able to identify irregularities in the conduct of the investigation by LAPD detectives and forensic specialists. For example, one of the detectives, Philip Vannatter, had carried a sample of Simpson's around with him for hours; and some of the blood taken from Simpson was unaccounted for. After much digging, the defense team found evidence that Mark Fuhrman, the detective who had allegedly found the bloody glove in Simpson's yard, was a raving racist who, contrary to his claim on the stand, frequently used the word "nigger" and had bragged in the past about framing blacks, especially ones involved with white women.

Figure 2 shows part of an explanatory coherence analysis of the case made by the defense. The hypothesis that Nicole and Goldman were killed by drug dealers competes with the hypothesis that O. J. Simpson was the killer. Unfortunately for the defense, they were unsuccessful in finding any substantial evidence for this hypothesis. But they were very effective in offering alternative explanations of the blood evidence using the hypothesis that the LAPD had planted evidence. The glove that Simpson had supposedly used in the murder did not appear to fit his hand when he tried to put it on in court. Blood had not been noticed on the sock until weeks after it had been held by the police. The blood on the glove and sock showed traces of EDTA, a chemical used as an anti-coagulant in samples taken from O. J. Simpson and Ron Goldman. Fuhrman and other detectives had ample opportunity to plant the evidence that implicated Simpson, and Fuhrman had a racist motivation to do so.

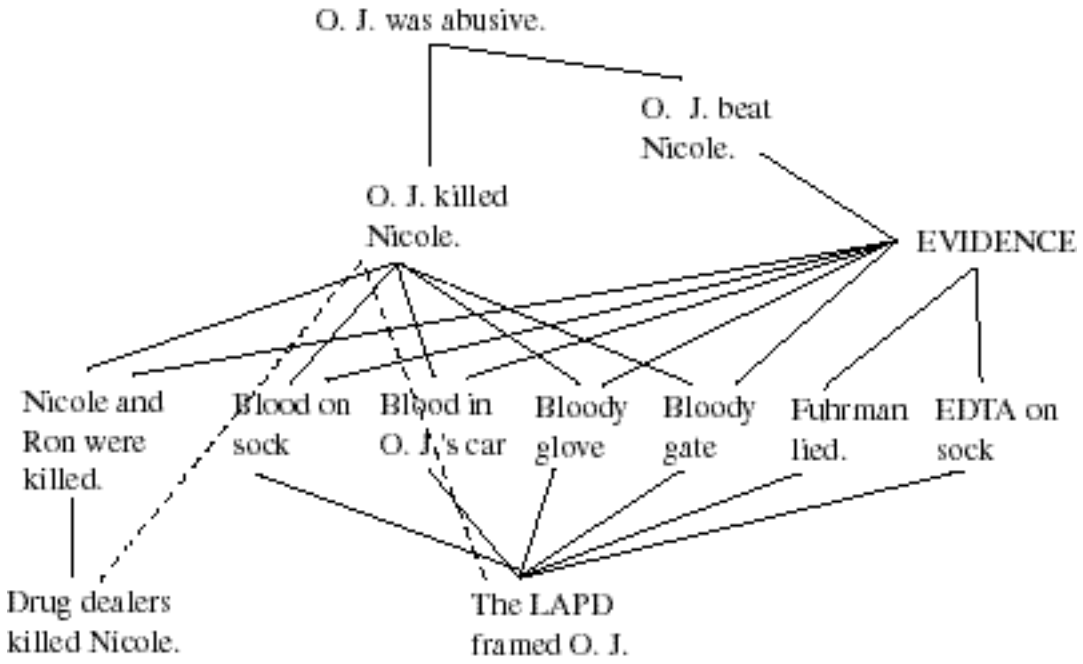

Figure 2. Expanded explanatory coherence analysis of the competing stories in the Simpson trial. Solid lines indicate coherence relations, while dotted lines indicated incoherence relations between competing hypotheses.

The complex of hypotheses and evidence shown in figure 2 provides a possible cold-cognitive explanation of why the jurors found Simpson not guilty. Perhaps, given the evidence and all the competing hypotheses, they found greater explanatory coherence in the story that Simpson had not been the killer. However, when the program ECHO is given input that corresponds to the evidence, hypotheses, and explanations in figure 2, it accepts the proposition that Simpson was the killer and rejects the alternative hypothesis that the murder was committed by drug dealers. Interestingly, ECHO also accepts the hypothesis urged by the defense that the LAPD tried to frame Simpson. Frankly, the conclusion that Simpson was guilty AND he was framed strikes me as quite reasonable.

The jury did not see additional evidence that is best explained by the hypothesis that Simpson was the murderer. For procedural reasons, evidence was not admitted

concerning the finding of unusual fibers from Simpson's car at the crime scene. Months after the trial, photographs were found that showed that Simpson had owned a pair of size 12 Bruno Magli loafers of the sort that had left bloody foot prints at the scene of the crime. But even without this additional evidence, ECHO's assessment of explanatory coherence accepts the hypothesis that Simpson was the killer. ECHO, unlike the jury, finds Simpson guilty. Hence given the evidence and explanations shown in figure 2, explanatory coherence fails to account for why the jury did not convict him.

It is of course possible that the jurors were mentally working with a different explanatory network, not represented by figure 2, in which the hypothesis of Simpson's innocence fit with the most coherent story. Moreover, it is also possible that the jurors did think that the evidence supported his guilt, but not beyond a reasonable doubt. According to the legal scholar Alan Dershowitz (1997), who was also a member of the Simpson defense team, the incompetence of the police and prosecution made room for the jury to conclude that there was reasonable doubt about Simpson's guilt. Two additional Simpson lawyers, Cochran (1997) and Shapiro (1996), also suggested that it was reasonable for the jury to doubt the prosecution's case. Three of the jurors describe their conclusions as based on reasonable doubt (Cooley, Bess, and Rubin-Jackson, 1995).

From the perspective of the theory of explanatory coherence, reasonable doubt might be viewed as an additional constraint on the maximization of coherence, requiring that hypotheses concerning guilt must be substantially more plausible than ones concerning innocence. In ECHO, presumption of innocence can be modeled by treating hypotheses concerning guilt as the opposite of data, so that their activation is suppressed in order to require that the hypotheses they represent achieve only when coherence overwhelmingly requires it. In fact, simulation of the network in figure 2 with inhibition of the unit representing Simpson being the killer can reject that hypothesis, but only if the inhibition is over .065, which is stronger than the default .05 excitation value for data. I am inclined, therefore, to conclude that the jurors' decisions in favor of Simpson was not

based solely on explanatory coherence and reasonable doubt, and later I will present evidence that their decisions were in part emotional. First, however, it is necessary to consider an alternative cold-cognitive explanation of the jury decision, based on probability theory rather than on explanatory coherence. Later in the paper I consider an emotional-coherence interpretation of reasonable doubt.

Probability Theory

Perhaps the jury in the Simpson trial inferred that the probability that he committed the crime given the evidence was insufficient for conviction. The conditional probability that Simpson was guilty given the evidence, P(guilty/evidence), can in principle be calculated by Bayes's theorem, which says that the posterior probability of a hypothesis given the evidence, P(H/E), is a function of the prior probability of the hypothesis, P(H), the likelihood of the evidence given the hypothesis, P(E/H), and the probability of the evidence: P(H/E) = P(H)*P(E/H) / P(E). To calculate P(guilty/evidence), we need to know the prior probability that Simpson was guilty, the probability of the evidence in the trial given the hypothesis that Simpson committed the murder, and the probability of the evidence. Some legal scholars (e.g. Lempert, 1986) contend that jurors do and should use probabilistic reasoning of this kind.

It is obvious that these probabilities are hard to come by. If probability is interpreted objectively as involving frequencies in a population, then the relevant probabilities are undefined, since we have no idea about the relative frequencies needed to attach a probability to propositions such as that Simpson committed the murder. Hence advocates of the application of probability theory to complex inference, often called Bayesians, have to rely on a subjective conception of probability as degree of belief. But this interpretation of probability is also problematic for legal applications, since there is no support for the view that the degrees of belief of the jurors conform to the principles of the calculus of probabilities. Indeed, there is considerable

psychological evidence that people's degrees of belief violate the laws of probability theory (Kahneman, Slovic, and Tversky, 1982; Tversky and Koehler, 1994).

Even supposing that it were reasonable to measure degrees of belief by probabilities, there is the great practical problem of attaching meaningful probabilities to the various propositions involved in a judgment. In addition to its explanatory coherence interpretation, figure 2 can be given an interpretation in terms of conditional probabilities based on causal relations. Much sophisticated work in artificial intelligence has concerned the calculation of probabilities in probabilistic causal networks, which are usually called Bayesian networks (Pearl, 1988; Neapolitain, 1990). For calculation of probabilities such as P(O.J. killed Nicole), such networks need a wealth of conditional probabilities such as P(blood in O. J.'s car/O. J. killed Nicole) and P(blood in O. J.'s car/O. J. did not kill Nicole). The jurors in the Simpson trial had no idea what values might be attached to such probabilities, and neither does any expert one might consult. Probability theory is an extraordinarily valuable tool for dealing with frequencies in populations, but its application to the psychology of causal reasoning is rather fanciful.

Moreover, Ronald Allen (1991,1994) has pointed out many respects in which probability theory and Bayesian inference do not fit well with legal practice. For example, if two hypotheses H1 and H2 are independent, then P(H1 & H2) is always less than or equal to P(H1) and less than or equal to P(H2). In a trial in which the case for the prosecution involves many propositions that must be jointly evaluated, the probability of the conjunction of these hypotheses will typically drop below .5, so that it would seem that a probabilistically sophisticated jury would never have good reason to convict anyone. In addition, no one has been able within a probabilistic framework to give a plausible interpretation of reasonable doubt, which is a corner stone of criminal law in the U. S. and elsewhere. Does "beyond a reasonable doubt" mean that the probability that a person committed a crime must be greater than .6 rather than .5 for conviction, or does it mean that the ratio of the probability of guilt to the probability of innocence must be well

over 1, or what? I gave an explanatory coherence account of reasonable doubt above, but will give a fuller account below as part of the application of emotional coherence to legal inference.

A full probabilistic analysis of the 12-node network shown in figure 2 would require 2^{12}=4096 probabilities, but Bayesian networks simplify the calculation of probabilities by assuming that each variable representing the truth or falsity of a proposition is independent of most other variables. Figure 3 shows a Bayesian network constructed with JavaBayes, a very sophisticated and convenient Bayesian simulator due to Cozman (2001). The network in figure 3 has the same structure as the coherence network in figure 2, except that the links are unidirectional; they represent conditional probability rather than coherence. Each node represents a variable that has two possible values, TRUE and FALSE. The arrows typically indicate causal relations: for example, Simpson's abusive nature was a possible cause of his killing Nicole and not vice versa. The exception to this interpretation are the link that I put in between *OJ-killed-Nicole* and the alternative explanations, *drug-killing*. This is not a causal relation, but some probabilistic connection must be represented to indicate the presumed incompatibility of these hypotheses. I drew the links *from OJ-killed-Nicole* rather than to it because doing so reduced the number of required conditional probabilities, for reasons I will now explain.

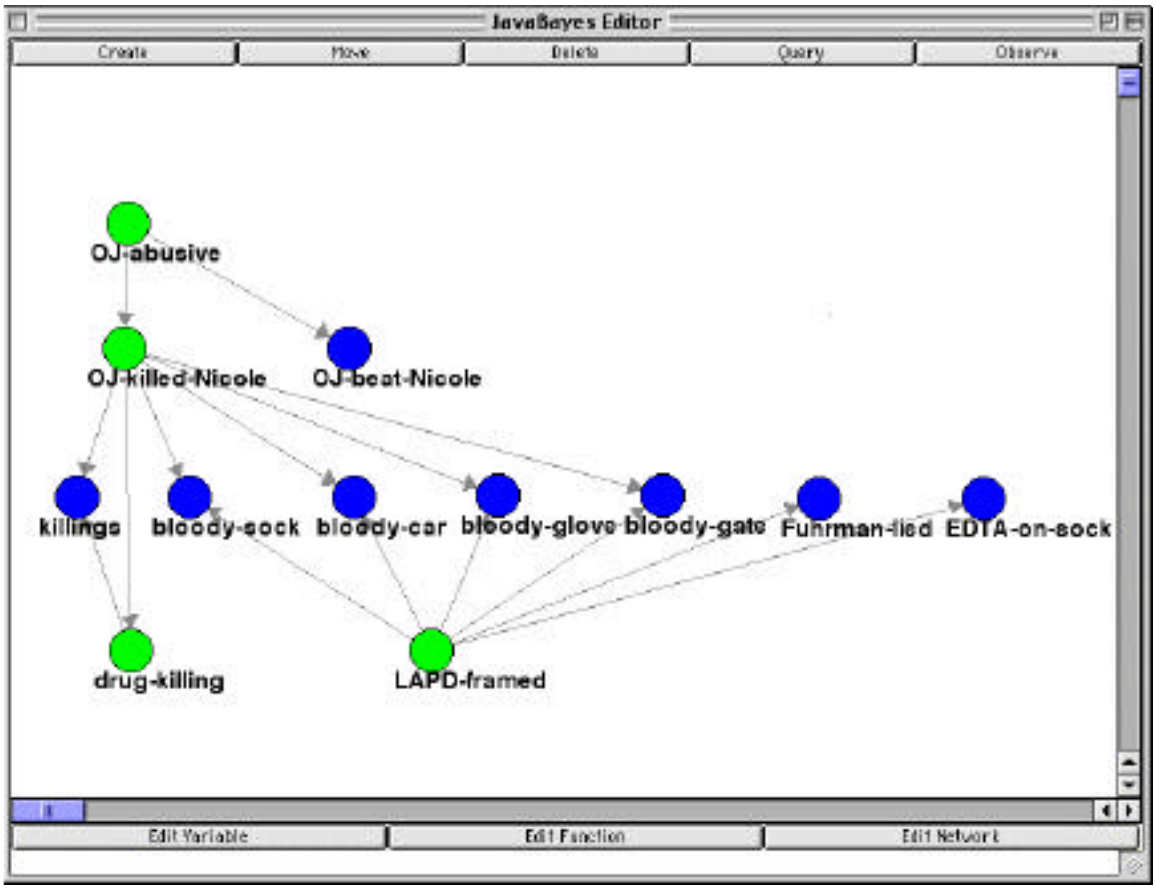

Figure 3. Bayesian network produced with the program JavaBayes (Cozman, 2001). Dark nodes represent observed variables, while light nodes represent explanatory variables. Links indicate conditional probabilities.

I did not, however, draw a link between *OJ-killed-Nicole* and *LAPD-framed*, which seem to me to be probabilistically independent. In ECHO, the framing hypothesis is treated as incoherent with *OJ-killed-Nicole*, not because they are logically incompatible, but because they offer competing explanations of evidence such as the bloody sock. If there is to be a connection between the two hypotheses, JavaBayes requires numbers for P(*OJ-framed/OJ-killed-Nicole*), P(*not-OJ-framed/OJ-killed-Nicole*), P(*OJ-framed/not-OJ-killed-Nicole*), P(*not-OJ-framed/not-OJ-killed-Nicole*). I do not know what these probabilities should be, logically or psychologically.

After a network is created with JavaBayes, it is necessary to insert probability values that feed into the calculation of the probability of the explanatory variables, in this case *OJ-killed-Nicole, drug-killing* and *LAPD-framed***.** Some variables are marked as having observed values: all the dark nodes in figure 3 represent variables with observed values of TRUE, so that P**(***killings***)**=1. For each node that has *n* arrows going into it, it is necessary to specify $2*2^n$ conditional probabilities. Specifying the 60 probabilities for the network in figure 3 is much easier than doing the full joint distribution, but it is still a daunting task.

Consider first the conditional probabilities for *OJ-killed-Nicole***.** JavaBayes requires numbers for P(*OJ-killed-Nicole/OJ-abusive***),** P(*OJ-killed-Nicole/not-OJ-abusive*), P(*not-OJ-killed-Nicole/OJ-abusive*), and P(*not-OJ-killed-Nicole/not-OJ-abusive*). After some reflection on the frequency of marital abuse and murder, I plugged in the following numbers: .2, .01, .8, .99. I would be hard pressed to defend these numbers either logically or psychologically, although they do capture the intuition that Simpson was more likely to have killed Nicole if he was abusive than if he was not. The specifications for the *killings* node is easier even though 8 conditional probabilities are needed, because obviously the probability of *killings* given either *OJ-killed-Nicole* or *drug-killings* is 1. But specifying the conditional probabilities for the *bloody-sock* node is extremely difficult: you might be able to come up with a high number for P(*bloody-sock/OJ-killed-Nicole & LAPD-framed*), but what number would you estimate for such quantities as P(*not-bloody-sock/not-OJ-killed-Nicole & LAPD-framed*)? After much reflection I inserted, for this variable and for others, some numbers that were guesses at best.

Once all the conditional probabilities had been inserted, I had JavaBayes query the different variables to calculate posterior probabilities given the evidence. The results were not unreasonable: P(*OJ-killed-Nicole*/evidence)=.72, P(*drug-killings*/evidence)=.29, and P(*LAPD-framed*/evidence)=.99. Like ECHO, JavaBayes concluded that Simpson

was guilty *and* he was framed. The last probability strikes me as much too high, but it is not clear how to alter the network and probability values in order to change it.

Since JavaBayes with the input shown in figure 3 reached a different conclusion from the jury, we can conclude that this network is not a good model of the jurors' thinking. Of course, it is possible that they had a different set of variables and conditional probabilities than the ones I provided. Constructing the causal network was no more difficult than constructing the explanatory network for ECHO, but coming up with values for the 60 conditional probabilities was very hard, even for someone who knows more about probability than most jurors. Thus explanatory coherence provides a much simpler and psychologically natural cold-cognitive account of how jurors reach decisions. Neither account, however, appears adequate to explain the jurors' thinking.

In sum, there are psychological and legal reasons for doubting the applicability of a Bayesian analysis to the jury's decision in the Simpson trial, even if it was a rational decision based only on the plausibility of the various hypotheses given the evidence. Even more obviously, probability theory can not account for the role that hot cognition involving emotions and motivation may have played in the Simpson trial.

Wishful Thinking

There is substantial evidence that emotional bias on the part of the jurors may have contributed to their decision to acquit O. J. Simpson. His lawyers hired a jury consultant who conducted a poll in which she found that 20 per cent of the sample believed Simpson innocent, and 50 per cent did not *want* to believe Simpson was guilty (Schiller and Willwerth, 1997, p. 220). The consultant then worked intensively with 75 people and found that black, middle-aged women were Simpson's most aggressive champions (p. 243). This was contrary to the expectations of the defense lawyers, who had thought that black women would resent O. J. Simpson for marrying a white woman, but found instead that virtually every middle-aged African-American woman in the focus group supported Simpson and resented the murder victim (p. 244)! Accordingly, the

defense team set out to get as many black, middle-aged women on the jury as possible. Further polling found that only 3 per cent of 200 African-Americans assumed that Simpson was guilty (p. 251), and 44 per cent said that Los Angeles police had treated them unfairly at least once. Most strikingly, 49 per cent of divorced black women wanted to see Simpson acquitted (p. 251). The defense was elated when juror selection produced a jury that included 8 blacks, most of them middle-aged women. In accord with their strategy to impress the female African-American women, the defense began its case with testimony with Simpson's daughter and mother. News polls also found that African-Americans, especially women, were inclined to believe that Simpson was innocent. Bugliosi (1997, p. 74) reports that a *Los Angeles Times* poll of blacks in Los Angeles county found that 75 per cent of them believed Simpson was framed. Psychological experiments have also found that blacks were more likely than whites to view Simpson as innocent (Newman et al., 1997).

It is possible that these differences between white and black attitudes were in part based on differences in personal experience: black residents of Los Angeles had probably observed or heard of more cases of people being framed by the police than white residents were aware of. But there is also reason to believe that some members of the jury in the Simpson trial were emotionally biased toward finding him not guilty. At the most extreme, one might propose that their verdict in the face of all the evidence linking Simpson with the crime was a matter of wishful thinking: the jurors found Simpson not guilty because they wanted to. Explanatory coherence, probability theory, and other cold cognitive factors had nothing to do with the jurors' decisions, which was based on their emotional attachment to Simpson and their motivation to acquit him.

It is implausible, however, to suppose that the jurors' decisions were merely a matter of wishful thinking. The defense certainly did not rely only on the fact that many of the jurors were probably predisposed toward Simpson; rather, his lawyers labored intensively to show that the LAPD was incompetent in collecting and protecting evidence

and that officers such as Fuhrman had the motive and opportunity to frame Simpson. Some members of the jury may have been emotionally inclined to acquit Simpson, but they would not have done so if the evidence had been overwhelmingly against him. One of the jurors reported after the trial that if she had been aware of some of the evidence that was not presented at the trial, then she would have voted to convict Simpson (Bugliosi, 1997, p. 143; see also Cooley, Bess, and Rubin-Jackson, 1995, p. 198). If there had been stronger evidence against Simpson, and if the case against the LAPD had not been so strong, then the jurors may well have found Simpson guilty despite their emotional attachments. One of the jurors, Carrie Bess, said on television (Bugliosi, 1997, p. 301): "I'm sorry, O. J. would have had to go if the prosecution had presented the case differently, without the doubt. As a black woman, it would have hurt me. But as a human being, I would have to do what I had to do."

This assessment is consistent with psychological research on motivated inference. Kunda (1999, p. 224) summarizes the results of psychological experiments as follows:

> Motivation can color our judgments, but we are not at liberty to conclude whatever we want to conclude simply because we want to. Even when we are motivated to arrive at a particular conclusion, we are also motivated to be rational and to construct a justification for our desired conclusion that would persuade a dispassionate observer. We will draw our desired conclusion only if we can come up with enough evidence to support it. But despite our best efforts to be objective and rational, motivation may nevertheless color our judgment because the process of justification construction can itself be biased by our goals.

This conclusion is supported by numerous experimental findings that people have reality constraints that keep them from believing whatever they want to believe (Kunda, 1990; Sanitioso, Kunda, and Fong, 1990). Thus the jurors in the Simpson trial may have started with an emotional bias to acquit him, but that motivation was probably not sufficient in

itself. Rather, there had to be interactions between the jurors' emotional attitudes and the evidence and explanations presented by the prosecution and the defense. Such interactions can be explained by the theory of emotional coherence.

Emotional Coherence

When people make judgments, they not only come to conclusions about what to believe, they also make emotional assessments. For example, the decision to trust people is partly based on purely cognitive inferences about their plans and personalities, but also involves adopting emotional attitudes toward them (Thagard 2000, ch. 6). The theory of emotional coherence serves to explain how people's inferences about what to believe are integrated with the production of feelings about people, things, and situations. On this theory, mental representations such as propositions and concepts have, in addition to the cognitive status of being accepted or rejected, an emotional status called a *valence*, which can be positive or negative depending on one's emotional attitude toward the representation. For example, just as one can accept or reject the hypothesis that Simpson was the murderer, one can attach a positive or negative valence to it depending on whether one thinks this is good or bad.

The computational model HOTCO implements the theory of emotional coherence by expanding ECHO to allow the units that stand for propositions to have valences as well as activations. In the original version of HOTCO (Thagard 2000), the valence of a unit was calculated on the basis of the activations and valences of all the units connected to it. Hence valences could be affected by activations and emotions, but not vice versa: HOTCO enabled cognitive inferences such as ones based on explanatory coherence to influence emotional judgments, but did not allow emotional judgments to bias cognitive inferences. HOTCO and the overly rational theory of emotional coherence that it embodied could explain a fairly wide range of cognitive-emotional judgments involving trust and other psychological phenomena, but were inadequate to explain the emotional biasing of inference that seems to have taken place in the Simpson trial.

Accordingly, I have altered HOTCO to allow a kind of biasing of activations by valences. Consider, for example, the proposition that *O. J. is good.* This proposition can be viewed as having an activation that represents its degree of acceptance or rejection, but it can also be viewed as having a valence that corresponds to a person's emotional attitude toward Simpson. The predicate "good" involves both a statement of fact and an evaluation. As such, it is natural for the valence of *O. J. is good* to affect its activation. Technical details concerning explanatory and emotional coherence are provided in an appendix.

Now we have a natural way to simulate the emotional bias of the jurors in the Simpson case. Figure 4 shows figure 2 with the addition of units corresponding to *O. J. is good* and *LAPD is good*. Depending on a person's emotional bias, these units may have positive or negative valence associated with them. It would seem, for example, that many of the black jurors had a positive emotional attitude toward Simpson, and a negative one toward the Los Angeles Police. Hence in figure 4, *O. J. is good* has a positive link to the valence unit that spreads valences, while *LAPD is good* has a negative link. The activation of these units is a function not only of the activation input to them but also of the valence input to them that they receive from the valence unit. Hence *O. J is good* tends to become active while *LAPD is good* tends to be deactivated. Then these units can influence the activations of the key hypotheses in the network, that O. J. was the killer and that the LAPD framed him. There is naturally a negative link between *O. J. is good* and *O. J. killed Nicole,* so that the positive evaluation of Simpson tends to suppress acceptance of the hypothesis that he killed his ex-wife. Similarly, a negative evaluation of the LAPD tends to support the hypothesis that Simpson was framed. When HOTCO 2 is run on the network shown in figure 4 with the default .05 valence link to *O. J. is good* and a -.05 valence link to *LAPD is good*, it rejects the conclusion that O. J. killed Nicole, just as the jury did.

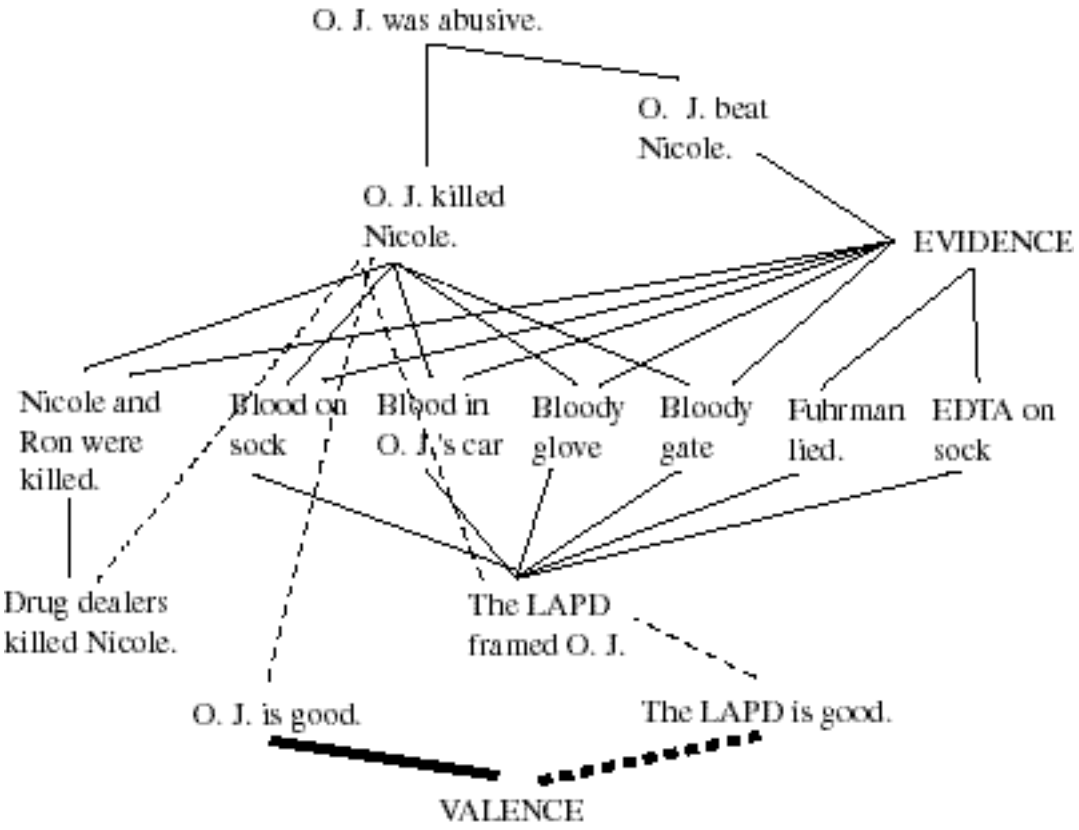

Figure 4. Emotional coherence analysis of the Simpson case. The thick lines are valence links. As with the other links, solid lines are excitatory and dashed lines are inhibitory.

But emotional coherence is not just wishful thinking, because it assumes that an inference is based in part on cognitive considerations, not just emotional bias. If the simulation just described is altered by deleting the defense's explanations of the evidence using the hypothesis that the LAPD framed Simpson, then HOTCO 2 finds Simpson guilty. If explanatory coherence supports a conclusion very strongly, then an emotional bias against the conclusion can be overcome. This fits well with the Simpson jurors contentions that if the evidence had been stronger they would have found Simpson guilty. Valences affect activations, but do not wholly determine them. Emotional bias requires coherence between emotional attitudes and evidence, not just wishful thinking.

It therefore seems that the most plausible answer to the question *Why wasn't O. J. convicted?* is that the jurors made their decisions based on emotional coherence, which combined an emotional bias with an assessment of competing explanations of the evidence. Given the flawed case presented by the prosecution and the ingenuity of the defense lawyers in generating alternative explanations, it was natural for the jurors to go with their emotional biases and find Simpson not guilty. A stronger case might have overcome the jurors' predisposition to acquit Simpson.

It might seem that emotional matters are totally inappropriate for use in deciding guilt or innocence, and I will argue in the conclusion that the kind of emotional biasing I have just described should generally *not* be part of legal decision making. But it is accepted in criminal proceedings that an accused should be convicted only if he or she is shown to be guilty beyond a reasonable doubt, which seems to me more a matter of value than of fact. Legal practice deems that acquitting a guilty person is not as bad as convicting an innocent one. This is a matter of fairness rather than fact or probability. The purpose of the law is not only to ascertain truth, but also to achieve fairness. In HOTCO 2, reasonable doubt is implemented by having a unit for *Acquit the innocent* which has positive valence and activation. It then inhibits the acceptance of any hypothesis concerning the guilt of an accused, such as that Simpson killed Nicole. When HOTCO 2 is run with an *Acquit the innocent* unit inhibiting the unit that represents guilt, it finds Simpson innocent with less pro-Simpson bias than is otherwise required to produce a not guilty decision. If this account of reasonable doubt is correct, then judgments of guilt and innocence legitimately involve emotional as well as explanatory coherence.

The involvement of emotional coherence in jury decision making also explains another aspect of legal practice that would be puzzling if juries used only cold cognition. According to Just (1998), one way in which the common law attempts to protect accused persons against irrational jury deliberations is the exclusion of evidence which has

prejudicial effect outweighing its probative value. Evidence can be prejudicial if it is of a kind to which a jury is likely to attach more importance than is deserved, or if it is likely to raise within a jury an emotional reaction to an accused that will distort calm and rational deliberation. In terms of the HOTCO 2 model, evidence is prejudicial if it attaches a negative valence to the accused in a way that would encourage acceptance of the hypothesis that the accused is guilty.

I have argued that the most plausible available explanation of Simpson's acquittal was that the mental processes of the jury involved emotional as well as explanatory coherence. What about the decision made by the jury in the civil trial initiated by the parents of Nicole brown Simpson and Ron Goldman? The jury in this trial found Simpson to be responsible and assessed him millions of dollars in damages (Petrocelli, 1998). There are several differences between the civil trial and the criminal trial that can help to explain the different outcomes. First, in a civil trial, there is no burden of proof beyond a reasonable doubt, so the jurors needed only to decide that the preponderance of evidence supported the hypothesis of Simpson's innocence. Second, the lawyers who made the case for Simpson's guilty avoided many of the mistakes made by the prosecution in the criminal trial, such as having the demonstrably racist detective Fuhrman called as a witness. Third, additional evidence had come to light by the time of the civil trial, particularly the pictures showing Simpson wearing Bruno Magli shoes. Fourth, the civil trial was conducted in Santa Monica and drew on a different population of jurors from those in the criminal trial, which was conducted in downtown Los Angeles. The lawyer for the families of Nicole Simpson and Ron Goldman was acutely aware of the pro-Simpson bias of black women, and managed to get a jury of mostly white males, with only one black women, (Petrocelli, 1998, p. 376). I conjecture therefore, that the jurors in the civil trial reached their conclusions because they had different emotional biases from those of the jurors in the criminal trials, as well as

because case for Simpson's guilt had greater explanatory coherence and no burden of reasonable doubt to overcome.

Psychological Evidence for Emotional Coherence

I have argued that the emotional coherence account of juror decision making is more plausible than purely cold or hot accounts, but have presented no direct evidence that the mental processes of jurors involve emotional coherence. But the results of two recent psychological studies support the hypothesis that people's inferences involve both cognitive and emotional constraint satisfaction as implemented in the HOTCO 2 model.

Westen and Feit (forthcoming) conducted three studies in 1998 during the scandals concerning U. S. President Clinton. All three studies found that people's beliefs about Clinton's guilt or innocence bore minimal relation to their knowledge of relevant data, but were strongly predicted by their feelings about Democrats, Republicans, Clinton, high-status philandering males, feminism, and infidelity. Westen and Feit argue that people's inferences about the scandal involved a combination of cognitive constraints (data) and affective constraints (feelings, emotion-laden attitudes, and motives). Their views are clearly consistent with the theory of emotional coherence described above, and HOTCO 2 can be used to simulate the inferences made by people in their studies.

Figure 5 shows the structure of a highly simplified HOTCO 2 simulation of central aspects of the first study of Westen and Feit (forthcoming), which concerned the allegations made by Kathleen Willey that she had been sexually harassed by the President. The hypothesis to be evaluated is that Clinton was guilty of harassment, which would explain why she accused him. On the other hand, the contradictory hypothesis that he did not harass her would explain his protestations of innocence. I have not included in figure 5 possible alternative explanations, such as that Willey made the accusation for political reasons and that Clinton denied the accusation simply to protect his reputation. In figure 5, the evidence is exactly balanced, so that the

explanatory coherence program ECHO finds the competing hypotheses that Clinton harassed Willey and that he did not do so equally acceptable – they get the same low activation.

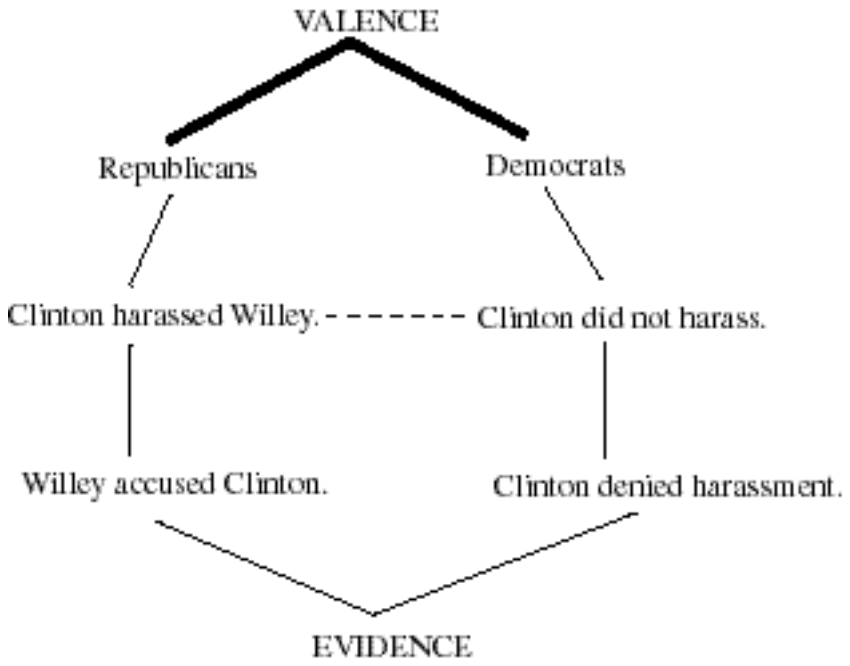

Figure 5. Emotional coherence in the assessment of whether President Clinton harassed Kathleen Willey. Thick lines are valence links, which may be positive or negative depending on attitudes toward Democrats and Republicans.

HOTCO 2, however, reaches very different conclusions depending on whether the Democrats or Republicans are favored by receiving a positive valence through a link with the VALENCE unit. If Democrats are favored by means of an excitatory valence link, the Democrat evaluation unit receives positive valence and activation, which suppresses the activation of the hypothesis that Clinton was guilty, so the program concludes that Clinton did not harass Willey. On the other hand, if Republicans are favored by means of an excitatory valence link, the Republican evaluation unit receives positive valence and activation, which then supports the activation of the hypothesis that Clinton was guilty.

Thus the behavior of the HOTCO 2 simulation is in accord with the findings of Westen and Feit that emotional attitudes predicted people's judgments of guilt and innocence. The subjects in the Westen and Feit studies obviously had many more values and beliefs than the bare-bones HOTCO 2 simulation, but it suffices to show how people's inferences about Clinton could arise from a combination of cognitive and emotional constraints. As in the Simpson simulation, HOTCO 2 is not simply engaging in wishful thinking, because if it is given a large amount of evidence against Clinton then it finds him guilty even if it has a pro-Democrat bias.

Further empirical support for emotional coherence is provided by studies of stereotype activation reported by Sinclair and Kunda (1999). They found that participants who were praised by a Black individual tended to inhibit the negative Black stereotype, while participants who were criticized by a Black individual tended to apply the negative stereotype to him and rate him as incompetent. According to Sinclair and Kunda, the participants motivation to protect their positive views of themselves caused them either to suppress or to activate the negative Black stereotype. Another study found similar reactions from students who received low grades from women professors: the students used the negative stereotype of women to judge female professors who had given them a poor grade as less competent than male professors who had given them an equally poor grade (Sinclair and Kunda, in press).

Figure 6 shows the structure of a simplified HOTCO 2 simulation of aspects of the experiment in which praise and criticism produced very different evaluations of the individual who provided them. Without any evidence input that the evaluation is good or bad, the program finds equally acceptable the claims that the evaluator is competent or incompetent. However, a positive evaluation combines with the motivation for self-enhancement to generate positive judgments of the evaluator and blacks, while a negative evaluation combines with self-enhancement to generate negative judgments of the evaluator and blacks. In the simulation shown in figure 6, the positive valence of the *I*

am good unit supports activation of the accurate-evaluation unit, which activates the competent-manager unit and suppresses the black stereotype. HOTCO 2 thus shows how thinking can be biased by emotional attachment to goals such as self-enhancement. Hence the mental mechanism of integrated cognitive and affective constraint satisfaction that is postulated by the theory of emotional coherence appears to be psychologically realistic.

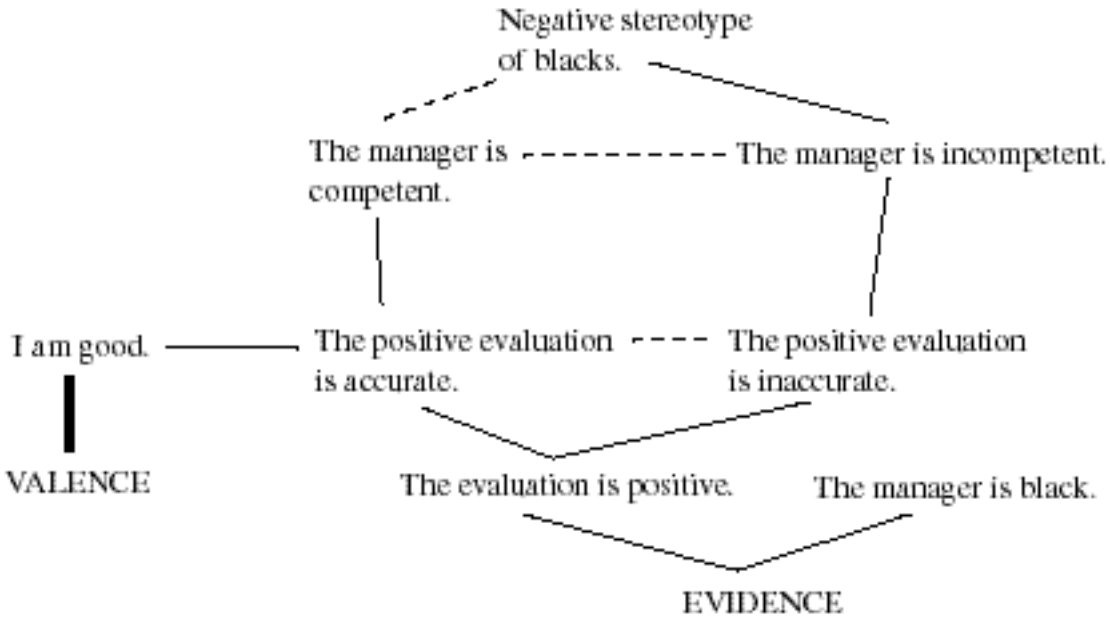

Figure 6. Evidential and valence associations leading to the motivated inhibition of the negative black stereotype. Solid lines are excitatory links and dashed lines are inhibitory.

Limitations of the HOTCO Model

Although the HOTCO model can simulate the results of psychological experiments as well as real-life legal decisions, it is obvious that it captures only a small part of the interactions of emotion and cognition. The theory of emotional coherence is nothing like a general theory of emotion. Both the theory and the computational model deal primarily with positive and negative valence, neglecting the many different kinds of positive and negative emotions. For example, different people's reactions to the

Simpson case included being happy that he was acquitted, or angry that Mark Fuhrman had lied, or sad that Nicole had been killed. HOTCO can model general reactions of happiness and sadness as overall coherence or incoherence, but it does not provide an account of how specific emotions directed toward particular objects can arise. In contrast, the ITERA model of Nerb and Spada (2001) can differentiate between sadness and anger as emotional reactions to events. Neither HOTCO nor ITERA can differentiate between myriad other emotions experienced by people, such as shame and pride. Emotional coherence is not intended as an alternative to appraisal theory, which provides a general account of how different emotions are elicited by different evaluations of events and situations (Scherer, Schorr, and Johnstone, 2001). Rather, it specifies some of the computational mechanisms by which appraisal might take place. Thagard and Nerb (forthcoming) discuss some of the relations between connectionist models, dynamic systems theory, and appraisal theory.

From a biological perspective, the HOTCO model is simplistic in many respects. It uses localist representations in which a whole concept or proposition is represented by a single unit, rather than distributed representations in which multiple neurons collectively represent multiple concepts or propositions. HOTCO units spread activation to each other symmetrically, without the unidirectional action and spiking behavior of real neurons. Morever, HOTCO makes no attempt to model neuroanatomical organization, such as the arrangement of neurons in particular parts of the brain such as the neocortex and the amygdala. Brandon Wagar and I are currently working on a much more neurologically accurate model of the interactions between emotion and cognition. Yet another limitation of the HOTCO simulations performed so far is that they model only the thought processes of a single juror, neglecting the social interactions that are part of group decision making. See Thagard (2000, ch. 7) for a computational model of how scientists reach consensus.

Conclusion

Despite these limitations, the theory of emotional coherence provides a psychologically plausible account of the decision made by the jurors to acquit O. J. Simpson. The two cold-cognitive explanations I considered, based on explanatory coherence and on probability theory, neglect the emotional considerations that appear to have been part of the psychological processes of the jurors. But the jurors did not engage in pure wishful thinking either: their emotional biases were integrated with considerations of explanatory coherence to produce a judgment that was in part emotion-based and in part evidence-based.

What *should* the jurors have been thinking? Members of a jury are supposed to be impartial, with no emotional biases for or against the accused. Hence it would seem illegitimate for the jurors to have biases that affect their interpretation of the evidence. If truth is one of the aims of legal deliberation, and if emotional bias helps to prevent the jury from arriving at true answers, then having emotions influence the assessment of evidence and explanatory hypotheses would seem to be normatively inappropriate. Moreover, if fairness is also an aim of legal deliberation, and emotional bias leads some involved parties to be treated unfairly, then the emotional part of emotional coherence seems to be doubly undesirable. Emotion only seems to be a normatively appropriate part of coherence judgments when emotional bias is inspired by fairness concerns, as in my account of reasonable doubt based on valuing acquitting the innocent over convicting the guilty.

I am not, however, trying to exclude emotion from legal thinking. Even scientific thinking is permeated by emotion (Thagard forthcoming), and it would be unreasonable to expect jurors to shut down the emotional reactions that are an ineliminable part of human thought (Damasio, 1994). All we can hope for is that the process of jury selection should tend to avoid the inclusion of jurors with strong

emotional biases, and that the conduct of trials by the prosecution, defense, and presiding judge should emphasize evidence and alternative explanations rather than emotional appeals. Juror decision making would then still be a matter of emotional coherence, but the emotional component would be minor compared to the rational assessment of the acceptability of competing hypotheses based on explanatory coherence. According to Posner (1999, p. 325): "It would be misleading to conclude that good judges are less 'emotional' than other people. It is just that they deploy a different suite of emotions in their work from what is appropriate both in personal life and in other vocational settings." Further work on the theory of emotional coherence should contribute to understanding of how emotions can enhance rather than undermine the quality of legal and other kinds of inference.

Acknowledgements: I am grateful to Ray Grondin, Josef Nerb, Cameron Shelley, Craig Smith, and anonymous referees for helpful comments, and to the Natural Sciences and Engineering Research Council of Canada for financial support.

References

Abelson, R. (1963). Computer simulation of "hot" cognition. In S. Tomkins (Ed.), *Computer simulation of personality* (pp. 277-298). New York: John Wiley and Sons.

Allen, R. J. (1991). The nature of juridical proof. *Cardozo Law Review, 373,* 373-422.

Allen, R. J. (1994). Factual ambiguity and a theory of evidence. *Northwestern University Law Review, 88,* 604-660.

Bugliosi, V. (1997). *Outrage: The five reasons why O. J. Simpson got away with murder*. New York: Island Books.

Byrne, M. D. (1995). The convergence of explanatory coherence and the story model: A case study in juror decision. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the seventeenth annual conference of the Cognitive Science Society* (pp. 539-543). Mahwah, NJ: Erlbaum.

Cochran, J. L., Jr. (1997). *Journey to justice.* New York: One World.

Cooley, A., Bess, C., & Rubin-Jackson, M. (1995). *Madam Foreman: A rush to judgment?* Beverly Hills, CA: Dove Books.

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

Damasio, A. R. (1994). *Descartes' error*. New York: G. P. Putnam's Sons.

Dershowitz, A. M. (1997). *Reasonable doubts: The criminal justice system and O. J. Simpson case*. New York: Touchstone.

Just, D. (1998). Excluding prejudicial evidence from criminal juries. *http://www.users.bigpond.com/justd/prejev.htm*.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kunda, Z. (1990). The case for motivated inference. *Psychological Bulletin, 108*, 480-498.

Kunda, Z. (1999). *Social cognition.* Cambridge, MA: MIT Press.

Lempert, R. (1986). The new evidence scholarship: analyzing the process of proof. *Boston University Law Review, 66,* 439-477.

Neapolitain, R. (1990). *Probabilistic reasoning in expert systems*. New York: John Wiley.

Nerb, J., & Spada, H. (2001). Evaluation of environmental problems: A coherence model of cognition and emotion. *Cognition and Emotion, 15,* 521-551.

Newman, L. S., Duff, K., Schnopp-Wyatt, N., & Brock, B. (19997). Reactions to the O. J. Simpson verdict: "Mindless tribalism" or motivated inference processes? *Journal of Social Issues, 53,* 547-562.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufman.

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

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition, 49,* 125-163.

Petrocelli, D. (1998). *Triumph of justice: The final judgment of the Simpson saga.* New York: Crown.

Posner, R. A. (1999). Emotion versus emotionalism in law. In S. A. Blandes (Ed.), *The passions of law* (pp. 309-329). New York: New York University Press.

Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology, 59,* 229-241.

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion*. New York: Oxford University Press.

Schiller, L., & Willwerth, J. (1997). *American tragedy: The uncensored story of the Simpson defense*. New York: Avon Books.

Shapiro, R. L. (1996). *The search for justice*. New York: Warner Books.

Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*, 885-904.

Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin, 26*, 1329-1342.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435-467.

Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.

Thagard, P. (1999). *How scientists explain disease.* Princeton: Princeton University Press.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Thagard, P. (forthcoming). The passionate scientist: Emotion in scientific cognition. In P. Carruthers & S. Stich & M. Siegal (Eds.), *The cognitive basis of science.* Cambridge: Cambridge University Press.

Thagard, P., & Nerb, J. (forthcoming). Emotional gestalts: Appraisal, change and emotional coherence.

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science, 22*, 1-24.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*, 547-567.

Westen, D., & Feit, A. (forthcoming). All the president's women: Affective constraint satisfaction in ambiguous social cognition. Unpublished manuscript, Boston University.

Appendix: Technical Details

The explanatory coherence program ECHO creates a network of units with explanatory and inhibitory links, then makes inferences by spreading activation through the network (Thagard, 1992). The activation of a unit *j*, a_j, is updated according to the following equation:

$a_j(t+1) = a_j(t)(1-d) + net_j(max - a_j(t))$ if $net_j > 0$, 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), *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:

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

In HOTCO, units have valences as well as activations. The valence of a unit u_j is the sum of the results of multiplying, for all units u_i to which it is linked, the activation of u_i times the valence of u_i, times the weight of the link between u_i and u_j. The actual equation used in HOTCO to update the valence v_j of unit *j* is similar to the equation for updating activations::

$v_j(t+1) = v_j(t)(1-d) + net_j(max - v_j(t))$ *if* $net_j > 0$, $net_j(v_j(t) - min)$ *otherwise.*

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

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

Updating valences is just like updating activations plus the inclusion of a multiplicative factor for valences.

HOTCO 2 allows units to have their activations influenced by both input activations and input valences. The basic equation for updating activations is the same as

the one given for ECHO above, but the net input is defined by a combination of activations and valences:

$$net_j = \sum_i w_{ij} a_i(t) + \sum_i w_{ij} v_i(t) a_i(t).$$

ECHO and HOTCO both proceed in two stages. First, input about explanatory and other relations generates a network of units and links. The LISP input for all simulations used in this paper is available on the Web at http://cogsci.uwaterloo.ca/coherencecode/cohere/hotco-input.lisp.html. Second, activations and (for HOTCO) valences are updated in parallel in accord with the above equations. Updating proceeds until the network has settled, i. e. when all activations have reached stable values. Running the network shown in figure 2 takes 112 cycles of updating, which requires less than 1 second on a Macintosh G4. Running the emotional coherence network shown in figure 4 takes 410 cycles of updating, which requires less than 2 seconds.