
PAUL THAGARD

TESTIMONY, CREDIBILITY, AND EXPLANATORY
COHERENCE

ABSTRACT. This paper develops a descriptive and normative account of how people respond to testimony. It postulates a *default pathway* in which people more or less automatically respond to a claim by accepting it, as long as the claim made is consistent with their beliefs and the source is credible. Otherwise, people enter a *reflective pathway* in which they evaluate the claim based on its explanatory coherence with everything else they believe. Computer simulations show how explanatory coherence can be maximized in real-life cases, taking into account all the relevant evidence including the credibility of whoever is making a claim. The explanatory-coherence account is more plausible both descriptively and normatively than a Bayesian account.

1. INTRODUCTION

When someone tells you something, should you believe it? This problem is ubiquitous in law, business, science, and everyday life. It arises in law when jurors and judges have to decide whether to believe the testimony of a witness. It arises in business when buyers, sellers, and others make claims that may be dubious. The problem arises in science when researchers report experimental results that may be true but alternatively might be the result of fraud or incompetence. Finally, in everyday life we usually accept the statements of friends, colleagues, and acquaintances, but we sometimes need to doubt their veracity. All of these are cases of testimony, where people usually accept but sometimes reject what is told to them.

Many philosophers have noted the heavy dependence of human knowledge on testimony (see, for example, Coady, 1992; Audi, 1997; Lipton, 1998). But there is currently no general theory of how people do and how people should respond to testimony. This paper develops a framework for both a descriptive, psychological theory of testimony response and a prescriptive, philosophical theory of how people can best respond to testimony. In the spirit of naturalistic epistemology, the prescriptive theory is intimately related to the

descriptive theory, in that it describes how people evaluate testimony when they are doing it best.

Section 2 provides a general account of how people respond to testimony. It postulates a *default pathway* in which people more or less automatically respond to a claim by accepting it, as long as the claim made is consistent with their beliefs and the source is credible. When these two conditions are violated, however, people enter a *reflective pathway* in which they evaluate the claim based on its explanatory coherence with everything else they believe. People usually do and always should accept a claim if and only if accepting the claim maximizes explanatory coherence. A person's making a claim can always be explained in different ways, for example by the hypotheses that the person really believes the claim and has evidence for it, or alternatively by the hypotheses that the person does not believe the claim and is trying to deceive the listeners. Listeners need to decide which of the competing explanations is most coherent with all the evidence, including what is known about the credibility of the person making the claim. You should believe a claim made by a person if its truth is part of the best explanation of all the evidence relevant to the claim (Lipton, 1998).

Sections 3–5 develop in detail the explanatory-coherence theory of reflective response to testimony. I show how it applies to real-life cases, including a business case involving an Internet bulletin board and a legal case involving a police officer's testimony. Computer simulations show how explanatory coherence can be maximized in these and other cases, taking into account all the relevant evidence including the credibility of whoever is making a claim. Section 6 provides an extended discussion of credibility, arguing that it is different from reliability construed as a conditional probability. Section 7 contrasts the explanatory-coherence theory of testimony with the Bayesian theory that says that testimony should be evaluated using the resources of probability theory. After describing a Bayesian simulation of the business case previously discussed, I argue that the explanatory-coherence account is more plausible both descriptively and normatively than the Bayesian account.

2. THE PROCESS OF RESPONDING TO TESTIMONY

In the course of a day, people hear or read hundreds of statements from other people and various media, including television, radio, newspapers and the Internet. I believe most of what I hear without

much reflection, for example the TV's weather forecast, the newspaper's report of the sports scores, and my sons' descriptions of their activities at school. Occasionally, however, I encounter a claim that is dubious, for example when my son says that he does not have much homework when I know he has a big assignment due. At that point I start to wonder about why he might be saying that he does not have much homework, for example because he wants to play video games instead. In addition, when I encounter a report from a source that I know to be unreliable such as a supermarket tabloid, I do not automatically accept the claim.

A general theory of testimony has to be able to explain both how testimony is usually accepted automatically but also how it sometimes provokes extensive reflection about the claim being made and the claimant who is making it. Figure 1 is a proposal about the psychological processes that arise in people responding to testimonial claims. It suggests that when people encounter a claim made by a person or other source there is an unconscious assessment of the claim and its source. Usually, the claim passes these two tests and is accepted without reflection. The top part of Figure 1 illustrates the *default pathway* in which claims are quickly accepted. However, if the claim is inconsistent with what the person believes or if the source is known to lack credibility then reflection is triggered. The result, shown in the bottom part of Figure 1, is a much more extended process of assessing the explanatory coherence of the claim, which is discussed in Section 4. I call this bottom part the *reflective pathway*. The hypothesis that people tend to accept testimony by default has previously been suggested by Burge (1993), Diller (2000), Fricker (1994), Lipton (1998), and Reid (1970). Figure 1 goes beyond these

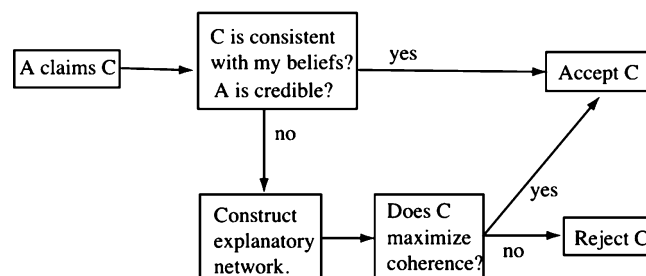


Figure 1. Dual-pathway model of response to testimony in which a person or other source A makes a statement C. The top part is the default pathway, and the bottom part is the reflective pathway.

suggestions in showing the relation between the default and reflective pathways.

From a descriptive, psychological perspective, Figure 1 is oversimplified in several respects. It omits a prior process by which an utterance is interpreted as a claim rather than a different kind of utterance such as a joke. If a person utters a sentence while winking or giggling, you will infer that the sentence is intended to make you laugh, not as a claim that you are supposed to believe. Similarly, the overall behavior and context of sources can itself raise questions about their credibility, for example if a person is making a claim while avoiding eye contact and sweating profusely. Hence the credibility check in the box in the top row of Figure 1 may include both general information about the source and particular, contextual information about it.

When you hear an utterance that is inconsistent with what you already believe, your response is usually emotional, ranging from mild surprise or puzzlement to stronger reactions such as annoyance or even anger. Prescriptively, it seems appropriate not to accept claims that are inconsistent with what you believe. But such claims should not be automatically rejected either, because there may be more evidence for the claim than for what you believe: the fact that a credible source states the claim is one such piece of evidence. So what you ought to do is to consider more thoroughly whether to accept or reject the claim, as shown in the reflective pathway of Figure 1.

Do people actually do this? I know of no experiments that show that people become reflective when a claim is made that is inconsistent with their beliefs, as opposed to simply rejecting the claim. There is, however, considerable evidence that people are often inclined not to accept claims that are incompatible with their personal goals (Kunda, 1990). For example, when coffee drinkers are given a mock newspaper article that links cancer with caffeine consumption, they are less inclined to believe the claim than people who do not drink coffee. This kind of motivated inference is not normatively appropriate, but the fact that it triggers a more reflective kind of inference makes it plausible that inconsistency would also have the same kind of effect. Descriptively, then we have four kinds of reflection triggers: lack of credibility of the source, non-credible behavior of the source, inconsistency of the claim with other beliefs, and incompatibility of the claim with the hearer's goals. Prescriptively, all of these seem like good grounds for additional, more reflective processing, except possibly for the last. I am not sure whether motivation should or should not be a reflective trigger. It might seem on the negative side that

using practical concerns to trigger reflection would create a bias toward believing what you want to believe, because if you like a conclusion you will not query it. On the other hand, if motivation is excluded from the reflective pathway, as it should be, then it will not bias the evaluation of the explanatory coherence of claims, so that triggering reflection on the basis of motivation will not contribute to the acceptance of defective claims. Moreover, from the perspective of practical rationality, which directs us toward accomplishing our goals, it may be useful to challenge information that is incompatible with those goals.

Descriptively, the overall process depicted in Figure 1 appears accurate, in that people do not challenge most of the claims that they encounter every day. We can ask, however, whether the default pathway is prescriptively appropriate: perhaps it would be epistemologically better always to use the reflective pathway, for the sake of obtaining fewer false beliefs. According to a recent study, lies are frequent in everyday conversation; 60% of participants in an experiment lied during a 10-min conversation, and did so an average of almost three times (Feldman et al., 2002). Our epistemic goals, however, go beyond just avoiding false beliefs; we want also to have a large number of important and useful beliefs. If we tried to reflect about every one of the statements that we encounter, we would be greatly restricted in the number of beliefs we could acquire. Moreover, everyday life would become impractical, in that we would spend so much time reflecting that we would never have time to do anything. We would be like the psychiatrist in the joke who, after a friend greeted him with “Hi, how are you?” muttered: “I wonder what she meant by that?” So I conclude that dividing response to testimony into the dual pathways displayed in Figure 1 is prescriptively as well as descriptively appropriate.

Much more contentious is the nature of the reflective pathway in that figure. The rest of the paper will argue for an explanatory-coherence interpretation of reflective response to testimony. I return to the general question of the justification of beliefs based on testimony in the concluding Section 8.

3. CASE STUDY 1: AN INTERNET BULLETIN BOARD

I begin with a real case that shows the need to evaluate claims made by other people. I regularly follow an Internet bulletin board that discusses a small Canadian company, Oncolytics Biotech. This

company is investigating the use of a naturally occurring virus, called the reovirus, to treat cancer. The reovirus has been amazingly successful in killing many kinds of cancer cells in test tubes and in mice. Moreover, a phase 1 clinical trial that was primarily aimed at testing for the safety of administering the reovirus to humans found evidence of viral activity affecting tumors in 12 of 18 patients treated. Additional trials are underway to determine the effectiveness of the reovirus treatment on patients with various forms brain cancer. For a description and analysis of the discovery that the reovirus is capable of killing cancer cells, see Thagard (2002).

The Oncolytics Biotech bulletin board can be found on the Web site, <http://www.stockhouse.ca/>. Almost every day there are dozens of messages posted anonymously by Stockhouse members with obscure names such as *Dave54*. Some regular participants are very enthusiastic about the prospects of the Oncolytics Biotech based on their expectations that the reovirus will prove to be a successful and lucrative therapy; the main enthusiasts include Bioeye, Blue“chip”, PullDaTrigger, and Matdul (I am not making these names up). Other participants have been much more skeptical about the company's prospects, for example Dandak and BeammeupToddy. Many of the postings have little content, with pumpers saying that the stock will go up and dumpers saying that it will go down. But there have also been messages that are highly informative about the science behind the reovirus treatment of cancer and the commercial activities of the company that is developing it.

Suppose you read the messages posed on the Oncolytics Biotech board, perhaps because you are thinking of investing in the company or perhaps because you have an interest in new cancer treatments. The problem of testimony arises whenever a member makes a substantive claim: should you believe it or not? For example, if someone posts that some positive experimental results will be announced the next day, you need to decide whether the poster is sufficiently credible that you can believe what is claimed. I have found Bioeye to be one of the most consistently well informed and reliable contributors to the bulletin board, so when she makes statements I tend to believe them (one of the other members reported without response that Bioeye is female; most of the other posters seem to be male). On the other hand, the main dumpers such as Dandak and BeammeupToddy never seem to have much information to contribute, so I have found no reason to believe their claims that the company is worthless. I should note, however, that despite some good scientific results the company's stock price is much lower in 2004 than when it peaked in 2000.

The descriptive question, then, is: Why do I usually believe Bioeye's claims and dismiss Dandak's claims? The normative question is: Am I justified in believing Bioeye's claims and dismissing Dandak's? Notice that the problem of testimony is particularly acute in Internet bulletin boards and chat rooms because there is little information about the identity, background, and motivations of the anonymous participants. In contrast to legal witnesses, there is no way of determining the identity of the people posting messages and whether they are in a position to know what they claim is true. Nevertheless, it is sometimes possible to assess the believability of different claims, and the next section analyses this case in terms of explanatory coherence.

The analysis to come concerns the reflective pathway, but much of what I read on the bulletin board can be accepted via the more direct default pathway. If there are postings by Bioeye, Matdul, or Blue"chip", all of whom have established credibility from a history of good messages, then I accept their claims without much reflection as long as the posting is consistent with what I already believe. See Section 7 for a discussion of how to evaluate the credibility of various people.

4. TESTIMONY AND EXPLANATORY COHERENCE

When Bioeye, Dandak or other member makes a claim that triggers reflection, I need to ask why they are making the claim. The positive explanation in Bioeye's case is that she made the claim because she believed it, and she believed it because she was in a position to know that it was true. For example, she was the first to report in September, 2001, that the announcement of the results of the phase I clinical trial would be delayed beyond the time period that the company had previously specified. It therefore seems to me that she has some good contacts in the company that put her in a position to know what it is doing. Of course, I would be rather gullible if I did not take into account possible alternative explanations of why Bioeye is making the claim that she did. She may simply be mistaken, having gotten her information from an unreliable source. More seriously, she may be trying to trick the members of the board into buying or selling stock so that she can increase her own trading profits. What I need to do is to try to decide whether the best explanation of Bioeye's making the claim is that she believes it because it is true, or alternatively that she believes it even though it is false or that she knows it is false but wants

to mislead people. Similarly, when I read Dandak's posts, which routinely predicted without much evidence the precipitous decline of the stock price for Oncolytics Biotech, I need to ask why he is making such claims. Because he does not seem to have any evidence for them, I tend to conclude that he is simply trying to depress the stock price, presumably because he has shorted the stock, i.e. borrowed shares to sell with the prospect of buying them back later at a much lower price.

The structure of the inferences to believe Bioeye and disbelieve Dandak can be analyzed using the theory of explanatory coherence and the computer model ECHO. They have already been applied to a great many examples inference in science, law, and everyday life (see for example Thagard, 1989, 1999, 2000). The theory of explanatory coherence consists of the following principles:

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.

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. I will not present the technical details here, since they are available elsewhere (Thagard, 1999, 2000).

Appendix A list the input given to ECHO to simulate the evaluation of a claim by Bioeye that a clinical trial has been completed. The explanatory structure of the case is best grasped through Figure 2, which shows the relevant propositions and the coherence relations among them. The only piece of evidence in the network is that Bioeye says that the trial is done. The crucial question is why she says that. Figure 2 displays two possible explanations, one that she says it because she believes it, and the alternative that she says it because she is lying. That she is lying would be explained by her desire to manipulate the stock price. On the other side, there are two available explanations of why she believes that the trial is done: that she is mistaken as the result of some erroneous information or inference, and alternatively that she has good evidence that the trial is done. The top-level hypothesis that the trial really is done would be part of the explanation of why she has good evidence that the trial is done. All of these explanatory relations are represented in Figure 2 by solid thin lines. The contradictions in the explanatory network are represented in Figure 2 by dotted thick lines: the hypothesis that the

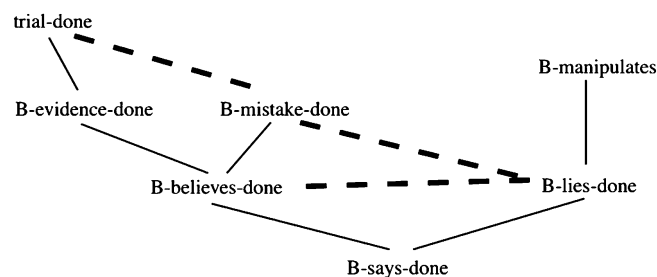


Figure 2. Explanatory coherence network of propositions determining whether to believe Bioeye. Solid lines indicate coherence relations, and dotted lines indicate incoherence relations. See Appendix A for the propositions.

trial is done contradicts both that Bioeye is mistaken and that the trial is done and that she is lying about whether it is done.

By itself, the network shown in Figure 2 is not sufficient to determine solidly which explanations are acceptable. In accord with explanatory coherence principle E4 (Data Priority), the unit representing *B-says-done* has an excitatory link with a special DATA unit that is always activated, so *B-says-done* is activated (in more complicated examples, however, data can be rejected, so explanatory coherence is not foundationalist). Running the network shown in Figure 2 leads to a weak conclusion that Bioeye is lying and that it is false that the trial is done. Activation spreads from *B-says-done* to *B-lies-done* and *B-believes-done*, which suppress each other's activation because of the inhibitory link between them, and then to all other hypotheses. Eventually, after dozens of iterations of spreading activation among units, the network settles (i.e. the activation of all units stabilizes) leaving *B-lies-done* with the slight advantage.

What is missing from the simulation shown in Figure 2 is a hypothesis about Bioeye's credibility. Anyone following the Oncolytics Biotech bulletin board regularly has experience of Bioeye making claims that are consistent with available scientific information and that usually turn out to be true. Thus a crucial part of the explanatory-coherence inference about whether to believe Bioeye requires some information about her credibility.

Figure 3 shows an expanded version of Figure 2 that shows how credibility fits into the explanatory-coherence account. If the unit representing the hypothesis that Bioeye is credible is supported by DATA, then it tends to become activated and thereby deactivates the hypotheses that Bioeye is mistaken or lying. The result is acceptance of the hypotheses that Bioeye has evidence that the trial is done and

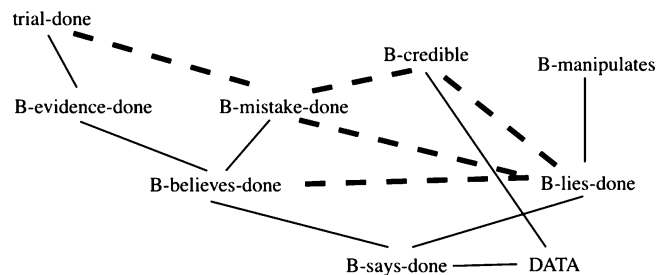


Figure 3. Explanatory coherence network of propositions determining whether to believe Bioeye, taking into account her credibility. Solid lines indicate coherence relations, and dotted lines indicate incoherence relations.

that the trial is in fact done. The hypothesis that Bioeye is credible is crucial to the overall assessment because it militates against the hypotheses that go against the hypothesis that she is telling the truth. Here the DATA unit does not represent any particular piece of information: it is merely a programming device for providing activation to units that represent propositions based on observational evidence, and the link with the unit representing B's credibility assumes that there is some such evidence that supports it.

Figure 3 is an advance on Figure 2 in that it takes credibility into account, but it does not say much about how credibility is determined, taking it simply as a kind of data. I will argue in Section 6, however, that inferences about credibility are more complex. First, however, I want to present another case study of inference in which testimony is rejected.

5. CASE STUDY 2: O. J. SIMPSON AND MARK FUHRMAN

To further illustrate how the evaluation of testimony can be based on explanatory coherence, consider the role of Detective Mark Fuhrman in the notorious trial of O. J. Simpson, who was accused of killing his wife. Fuhrman was one of the detectives who arrived at Simpson's house after the murder was reported, and he claimed to have found a bloody glove in Simpson's back yard. The glove belonged to Simpson and the blood matched that of the murdered wife, so this was a strong piece of evidence implicating Simpson. It seemed that the best explanation of the presence of the bloody glove in Simpson's yard was the hypothesis that he had murdered his wife.

At the trial, however, the defense succeeded in raising major questions about Fuhrman's credibility (Schiller and Willwerth, 1997; Thagard, 2003). He denied being a racist and using the word "nigger", but evidence was found that he frequently used the word and even bragged about harassing black men accompanied by white women. Some members of the jury said afterwards that after the revelations about Fuhrman they could not take seriously anything he said. The structure of the explanatory-coherence inference about Fuhrman is shown in Figure 4. The hypothesis that the bloody glove was found in Simpson's yard competes with the hypothesis that Fuhrman planted the glove, just as the hypothesis that Fuhrman is credible competes that he lied because he wanted to frame Simpson. The evidence that Fuhrman was a racist provides a convincing explanation of why he would do so. In case study 1, explanatory

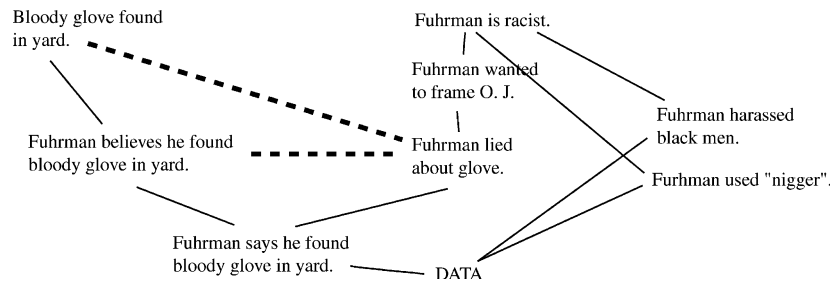


Figure 4. Explanatory-coherence analysis of why members of the jury in the trial of O. J. Simpson did not believe Fuhrman.

coherence explained why I believed Bioeye, and here it explains why members of the jury did not believe Fuhrman. Inferences about the credibility of Fuhrman are not shown in Figure 4, although they were surely made by the jury. Let us now investigate the nature of credibility in greater detail.

6. CREDIBILITY

As my discussion of Bioeye and Dandak indicated, the credibility of claimants is a crucial part of evaluating the claims that they make. But what is credibility and how is it assessed? This section argues first that the credibility of a speaker is not the same as reliability, which can be construed as a conditional probability. Rather, credibility should be understood as a dispositional psychological property of a person to be inferred by various means, including classification and analogy as well as enumerative induction.

The first crucial point to notice is that the credibility of a person is not a global property of the person, but is relative to particular topics (Fricker, 1995, p. 405). People may be credible on specific topics in which they have expertise, but not credible on other topics. For example, I have a friend who is a very reliable source of information on sports, but useless as a source on movies and politics. Bioeye was very credible on biological issues and questions involving developments at Oncolytics Biotech, but was caught out by other bulletin board members when she ventured opinions about legal matters involving patents. So whenever we speak of a person's credibility, we should consider it as relative to a particular topic. I shall use the formula $Cr(X, T)$ to represent the credibility of person X on topic T .

The simplest way of analyzing credibility would be as reliability construed as a conditional probability. Define $Rel(X, T)$ to be the

probability that a claim on topic T is true given that it was made by person X, in symbols $P(C/X \text{ claims } C)$. Reliability can be understood as a frequency: the ratio of claims by X on topic T that turned out to be true to all the claims made by X on topic T. For example, the reliability of Bioeye on biomedical information is the ratio of her true claims on this topic to all her claims. In principle, this kind of reliability could be calculated if one kept track of all her utterances on a topic and calculated what proportion of them are true. Presumably Bioeye would end up with a high reliability score while Dandak would have a low score. We might be tempted to posit the identity $Cr(X,T) = Rel(X,T)$.

There are several reasons why such an identification would be a mistake. First, we rarely have data that allow us to calculate reliability. An Internet bulletin board is unusual in that a reader could in principle look at all the claims made by a member over several years and then estimate what proportion are true. But even when such a record is available, no one has the time or interest to go back and make the calculation. In everyday life, there is no such permanent record of utterances so calculating a reliability ratio is impossible. Perhaps a person could do a rough estimate of a speaker's reliability based on a small sample of recent utterances, but people more commonly use such a sample to make explanatory inferences about a person's character, as described below.

The second reason for viewing credibility as different from reliability is that there are a number of different non-enumerative ways of inferring credibility. Consider, for example, Matdul, a relatively new member of the board. When he (or she) began to post to the board in October 2001, I quickly recognized him as a credible source because of the kinds of biological information he presented. As the result of background research I did for my paper on Patrick Lee (Thagard, 2002), I knew enough about the biology that I could tell from content and style that Matdul knew what he was talking about. His posts resembled those of Bioeye in biological expertise. In contrast, here is a typical post from Dandak (May 7, 2002): "It is going DOWN, DOWN DOWN!!! I told you so! But you did not listen!?!@#%\$#@". BeammeupToddy was even worse (Jan. 14, 2002, *sic*): "ROFLMAO HOW STUPID can one pumper be??? You think someone will pay \$8 a share for restricted stock in company with \$3 share price, just suffer huge setback lose partner, and half only 21 months of cash left? GIVE YER HEAD A SHAKE BOY!!!" Both these quotes are exactly as they appeared in the Oncolytics Biotech bulletin board, and it should be obvious why no one would take them seriously. What little content they have is not backed by evidence, and

the claims are poorly stated. In contrast, Bioeye and Matdul write well and convey information in defense of their claims. Hence, even though I have not calculated the reliability of Bioeye, Matdul, Dandak, and BeammeupToddy, I have no difficulty in judging the former two to be credible and the latter two not to be credible. Credibility is estimated on the basis of content and style as well as track record, not conditional probability. Presumably there is some correlation between credibility and reliability, but in evaluating testimony the former is prior to the latter. Content and style can speak directly to character and credibility, without intermediate probabilistic judgments of reliability. My judgments of reliability are not simply based on whether I like what the poster is saying, as there are people with positive posts that I ignore because they are poorly informed.

The third reason for not identifying credibility with reliability is that judgments concerning the two are affected very differently by the discovery that a person has uttered false statements. Suppose that a politician has made 100 statements about the economy, and it turns out that 2 of these are revealed to be false, where both of them concern an industry whose companies heavily contributed to the politician's campaign funds. The reliability of the politician on the economy would still be very high, 0.98, but we would become very reluctant to believe future utterances on this topic. It might even be said that the politician had, in the common phrase, "lost all credibility". Thus unlike reliability, credibility is not a simple linear function of the truth ratio of utterances. In the political case, we would infer that the politician has a set of beliefs and motivations that incline him or her to lie on topics relevant to the companies that provide financial support. Perhaps a mathematically sophisticated reasoner would think more extensively about whether the politician's 100 statements are statistically representative of all of his or her utterances, but the more natural inference is to judge the politician to be a liar with little credibility.

Thus credibility is not well construed as conditional probability, but rather requires a judgment about a person's disposition to tell the truth on a particular topic. Like all dispositions, we infer the existence of credibility by a theoretical inference to the best explanation. A person X is judged to be credible partly on the basis of X's track record on true statements, but also on the basis of the style and content of those utterances and our background knowledge about X. In general, X is judged to be credible if a disposition to tell the truth on a topic is inferred to be the best explanation of why X's utterances are usually true and have other

desirable properties such as informativeness and comprehensibility. A judgment of credibility is thus itself based on explanatory coherence. In Bioeye’s case, I have noticed a couple of instances where she reported misinformation, for example that the CEO of Oncolytics Biotech was still on the board of another company, Adherex. Rather than update Bioeye’s reliability number, however, I made little alteration in my estimate of her credibility, since in this case she had clearly made a minor mistake based on not being aware of a recent change in the board of Adherex.

Figure 5 shows the structure of the judgment that Bioeye is credible, understood as a hypothesis about a dispositional property that can be ascribed to her. That Bioeye has this property is justified as the best explanation of the facts that she writes lucidly and informatively and is almost always correct in what she says. Figure 5 extends Figure 3 by having the node for credibility supported by evidence, rather than being treated as directly related to evidence. As in the earlier simulation of belief in Bioeye, the program ECHO concludes that she is right that the trial is done. The major difference between the two simulations is that here Bioeye’s credibility is treated as an explanatory hypothesis rather than as given.

A similar account applies to the Mark Fuhrman case. Once the members of the Simpson jury learned that Furhman had lied about using the word “nigger” and had bragged about harassing black men, he lost all legal credibility. It was not just that his credibility received a numerical decrement as it would have if the juries had been making a judgment of reliability. Rather, the jury concluded that his character was fundamentally flawed so he was not to be believed in his testimony about Simpson.

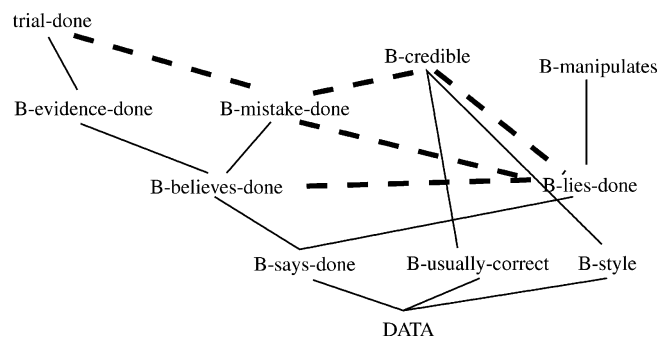


Figure 5. Full account of how a judgment about Bioeye’s credibility influences the judgment that what she says is true.

Thus the importance of credibility does not support a probabilistic interpretation of testimony. Let us now look in more detail at the main probabilistic alternative to my explanatory-coherence theory of testimony.

7. THE BAYESIAN ALTERNATIVE

The theory of explanatory coherence and the computational model ECHO provide a thorough and exact version of the doctrine that causal reasoning is inference to the best explanation. The main alternative theory of causal reasoning is provided by Bayesian networks, which use the resources of probability theory along with recently developed algorithms for computing probabilities to model causal reasoning (Pearl, 1988, 2000). The Bayesian approach is elegant and powerful, but there are reasons to think that it does not have much application to human thinking (Thagard, 2000, ch. 8; 2004). I will not attempt here a general comparison of Bayesian vs. explanatory-coherence approaches, but will describe a Bayesian simulation of the reasoning displayed in Figure 3. A Bayesian approach to evaluating testimony has been pursued by Goldman (1999, ch. 4).

Computationally, my Bayesian simulation produces the same basic result that ECHO does, accepting the hypothesis that Bioeye is right that the trial is done. Figure 6 shows the structure of the Bayesian network used in my simulations; it was produced using the program JavaBayes, an easy-to-use program due to Cozman (2001).

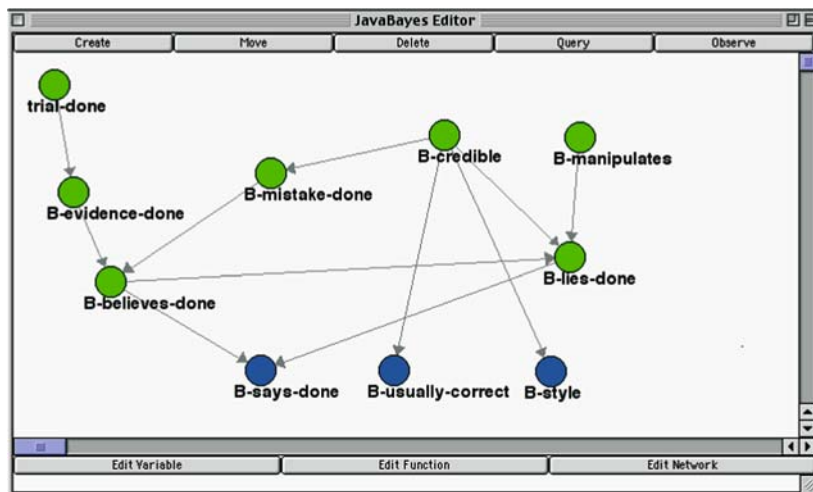


Figure 6. Bayesian simulation of the credibility of Bioeye using JavaBayes (Cozman, 2001). Arrows indicate causality.

To produce this simulation, I first had to identify the relevant nodes representing variables, which were adapted from the analysis in Appendix A. Next, I had to establish causal connections between nodes, which were also adapted from the explanation and contradiction relations in Appendix A. Finally, with much greater difficulty, I had to specify for each node with one or more arrows going into it the relevant conditional probabilities.

There are several reasons why the simulation in Figure 6 is implausible as an account of human thinking about the credibility of testimony. First, the simulation requires a host of conditional probabilities that people would be hard pressed to specify. Setting up the network requires specifying such conditional probabilities as $P(\text{B has evidence that the trial is done} / \text{trial is done})$, and also more complex ones. Because there are three arrows going into the node for *B-lies-done*, defining the function for calculating its probability requires specifying eight conditional probabilities, for example the probability that *B-lies-done* is false given that *B-manipulates* is true, *B-believes-done* is false, and *B-credible* is true. Despite having a reasonably good understanding of probabilities in domains where there are samples from populations of events, I have only the foggiest idea of what numbers to plug into the network. My rough guesses did, however, produce the desired result that the simulation concludes that the trial is done. Second, it is not clear what interpretation of probability to use. For epistemological purposes, it would be desirable to have objective probabilities, but these are only well defined for frequencies of occurrences in specified populations of events. Alternatively, a Bayesian might claim to be working with subjective probabilities understood as degrees of belief, but there is abundant psychological evidence that people's degrees of belief do not usually conform to the theory of probability (e.g. Kahneman et al., 1982; Gilovich et al., 2002). Third, it is not clear that the network in Figure 6 is a legitimate Bayesian network, because it is hard to say whether the links involve the direct causal influence that such networks require. Moreover, given the difficulty of assessing the relevant kinds of probabilities, it is impossible to determine whether the network satisfies the Markov condition that "a variable represented by a node in the Bayesian Network is independent of all variables represented by its non-descendant nodes in the Bayesian Network, conditional on all variables represented by its parent nodes" (Bovens and Hartman, 2003, p. 69). Hence it is not obvious that an explanatory coherence network can be translated into a Bayesian network so easily as Figure 6 suggests. For further discussion of the relevant applicability of

Bayesian and explanatory coherence approaches to legal reasoning, see Thagard (2004).

I see explanatory coherence as an alternative to Bayesian epistemology, in contrast to theorists who assess coherentist epistemology within the framework of Bayesian networks (Bovens and Olsson, 2000; Olsson, 2002a, b; Bovens and Hartman, 2003). Given the difficulty of interpreting and assigning probabilities to statements in law, science, and everyday life, there is no reason to take Bayesian theory as canonical. In particular, both the descriptive and normative theories of testimony can more plausibly be based on explanatory coherence rather than probabilistic ideas. In contrast, Lipton (2001) argues for the compatibility of probabilistic and explanatory analyses.

8. THE JUSTIFICATION OF TESTIMONY-BASED BELIEF

To conclude, I address the general question of how belief based on testimony is justified. Following a distinction made by Coady (1992), approaches to justification can be classified as *reductionist* or *non-reductionist*. Reductionist views hold that the justification of beliefs based on testimony is not special to testimony, but rather relies on the independent justifiability of beliefs based on perception, memory, or inductive inference. Reductionists in this sense include Hume (1972), Fricker (1994, 1995), Lipton (1998), and Lyons (1997). Non-reductionists maintain that acceptance of testimony is justified in itself, unless there is some reason to think that the testimony is false. Reasons for holding a non-reductionist view range from the contention by Reid (1970) that God created people with a propensity to speak the truth to a kind of transcendental argument that if testimony were not reliable there would be no such thing as testimony (Coady, 1992).

My dual-pathway model of response to testimony-based belief is reductionist, in that both pathways depend prescriptively on kinds of inference that are broadly used in domains other than testimony. Consider first the default pathway. It may well be that the tendency for people to believe what they are told is innate, not because of divine providence, but because of natural selection. We can at least tell a story of human evolution that involves selection for people who generally tell the truth as well as for people who generally believe what they are told. Groups of such individuals would likely have been more successful in surviving and reproducing than groups consisting primarily of the duplicitous and incredulous. Human society does

include psychopaths who are adept at and enthusiastic about lying, but such people are less than 1% of the population (Hare, 1993). If it seems plausible that the vast majority of statements made by people and media sources are true, then the default pathway is fairly reliable as well as efficient. Evolution, then, provided people with the default pathway that is useful both socially and epistemically. I do not know whether this story is true, but it does suggest how the default pathway to testimony-based inference might be both innate and reliable. The study by Feldman, Forrest, and Happ (2002) that found frequent lying may not be representative of most social interactions, because it involved participants who were strangers to each other and who were instructed to present themselves as likeable and competent.

What about the reflective pathway? If reflection about testimony depends on explanatory coherence, then its justification is the same as such inference in general. As I argued at the end of Section 7, I do not think that the justification of explanatory inferences depends on some Bayesian inference that they produce probable conclusions. Rather the justification of people's use of explanatory coherence is more indirect. Scientists' use of inference to the best explanation can be justified by noting that it is a standard practice in scientific thinking, and that there is much evidence from the technological applications of science that it often produce theories that succeed in representing how the world really is. A similar justification might be given for use in ordinary social contexts: we frequently use inference to the best explanation to infer the mental states of others, and the frequent successes of social interactions show that this strategy must be at least approximately veridical. Whatever justification applies to inference to the best explanation based on explanatory coherence applies by extension to testimony-based inference of the sort I described. Of course, such inferences are sometimes mistaken, like all inductive inferences, but the general practice of explanatory-coherence-based inference, in testimony and other contexts, is legitimated.

In sum, my dual-pathway account of testimony-based inference is consistent with a reductionist view of testimony that deems it to be justified on general inductive grounds. Neither Reid's divine intervention nor Coady's transcendental argument is needed to justify our very general practice of believing what we are told. My form of reductionist justification is complicated, however, by the fact that it gives different inductive justifications for each of the default and reflective pathways. When someone tells you something, whether you do and should believe it depends on what you know about the claim and the person making it.

ACKNOWLEDGMENTS

I am grateful to Luc Bovens, Axel Gelfert, Peter Lipton, Erik Olsson and anonymous referees for comments on earlier drafts. The research was supported by the Natural Sciences and Engineering Research Council of Canada.

APPENDIX A: INPUT TO ECHO SIMULATION OF INFERENCE
ABOUT BIOEYE

1. Input for simulation of network in Figure 2

(proposition 'trial-done "Clinical trial completed.")
 (proposition 'B-says-done "Bioeye says trial completed.")
 (proposition 'B-lies-done "Bioeye lies about trial.")
 (proposition 'B-manipulates "Bioeye wants stock up.")
 (proposition 'B-believes-done "Bioeye believes trial is done.")
 (proposition 'B-mistake-done "Bioeye is mistaken.")
 (proposition 'B-evidence-done "Bioeye has evidence trial is done.")
 (proposition 'B-cred-sci "Bioeye is credible about science.")
 ; alternative explanations
 (explain '(B-believes-done) 'B-says-done)
 (explain '(B-lies-done) 'B-says-done)
 (explain '(B-evidence-done) 'B-believes-done)
 (explain '(B-mistake-done) 'B-believes-done)
 (explain '(trial-done) 'B-evidence-done)
 (explain '(B-manipulates) 'B-lies-done)
 ; contradictions
 (contradict 'B-lies-done 'B-believes-done)
 (contradict 'B-lies-done 'trial-done)
 ; the evidence
 (data '(B-says-done))

2. Additional input for simulation in Figure 3

(contradict 'B-cred-sci 'B-lies-done)
 (contradict 'B-cred-sci 'B-mistake-done)
 (data '(B-says-done B-cred-sci))

3. Alternative additional input for simulation in Figure 5

(explain '(B-cred-sci) 'B-style)
 (explain '(B-cred-sci) 'B-usually-correct)
 (contradict 'B-cred-sci 'B-lies-done)
 (contradict 'B-cred-sci 'B-mistake-done)
 (data '(B-says-done B-usually-correct B-style))

REFERENCES

- Audi, R.: 1997, 'The Place of Testimony in the Fabric of Knowledge and Justification', *American Philosophical Quarterly* **34**, 405–422.
- Bovens, L. and S. Hartman: 2003, *Bayesian Epistemology*, Clarendon Press, Oxford.
- Bovens, L. and E. J. Olsson: 2000, 'Coherentism, Reliability, and Bayesian Networks', *Mind* **109**, 685–720.
- Burge, T.: 1993, 'Content Preservation', *Philosophical Review* **102**, 457–488.
- Coady, C. A. J.: 1992, *Testimony: A Philosophical Study*, Clarendon Press, Oxford.
- Cozman, F. J.: 2001, 'JavaBayes: Bayesian Networks in Java', <http://www-2.cs.cmu.edu/~javabayes/>.
- Diller, A.: 2000, 'Evaluating Information Found in Journal Articles,' in A. J. Nepomuceno, F. Quesada and F.J. Salguero (eds.), *Logic, Language, and Information: Proceedings of the First Workshop on Logic and Language*, Editorial Kronos, Sevilla, pp. 71–78.
- Feldman, R. S., J. A. Forrest and B. R. Happ: 2002, 'Self-Presentation and Verbal Deception: Do Self-presenters Lie More?', *Basic and Applied Social Psychology* **24**, 13–171.
- Fricker, E.: 1994, 'Against gullibility,' in B. K. Matilal and A. Chakrabarti (eds.), *Knowing from Words*, Kluwer, Dordrecht, pp. 125–161.
- Fricker, E.: 1995, 'Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony', *Mind* **104**, 393–411.
- Gilovich, T., D. Griffin and D. Kahneman (eds.): 2002, *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press, Cambridge.
- Goldman, A.: 1999, *Knowledge in a Social World*, Oxford University Press, Oxford.
- Hare, R. D.: 1993, *Without Conscience: The Disturbing World of the Psychopaths among Us*, Pocket Books, New York.
- Hume, D.: 1972, *Enquiries Concerning Human Understanding and Concerning the Principles of Morals* (2nd edn), Oxford University Press, Oxford.
- Kahneman, D., P. Slovic and A. Tversky: 1982, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York.
- Kunda, Z.: 1990, 'The Case for Motivated Inference', *Psychological Bulletin* **108**, 480–498.
- Lipton, P.: 1998, 'The Epistemology of Testimony', *Studies in History and Philosophy of Science* **29**, 1–31.
- Lipton, P.: 2001, 'Is Explanation a Guide to Inference? A Reply to Wesley C. Salmon,' in G. Hon and S. S. Rakover (eds.), *Explanation: Theoretical Approaches and Applications*, Kluwer, Dordrecht, pp. 93–120.
- Lyons, J.: 1997, 'Testimony, Induction, and Folk Psychology', *Australasian Journal of Philosophy* **75**, 163–178.
- Olsson, E. J.: 2002a, 'Corroborating Testimony, Probability, and Surprise', *British Journal for the Philosophy of Science* **53**, 273–288.
- Olsson, E. J.: 2002b, 'What Is the Problem of Coherence and Truth?', *Journal of Philosophy* **99**, 246–272.
- Pearl, J.: 1988, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufman, San Mateo.
- Pearl, J.: 2000, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge.

- Reid, T.: 1970, *An Inquiry Into the Human Mind*, University of Chicago Press, Chicago.
- Schiller, L. and J. Willwerth: 1997, *American Tragedy: The Uncensored Story of the Simpson Defense*, Avon Books, New York.
- Thagard, P.: 1989, 'Explanatory Coherence', *Behavioral and Brain Sciences* **12**, 435–467.
- Thagard, P.: 1999, *How Scientists Explain Disease*, Princeton University Press, Princeton.
- Thagard, P.: 2000, *Coherence in Thought and Action*, MIT Press, Cambridge, MA
- Thagard, P.: 2002, 'Curing Cancer? Patrick Lee's Path to the Reovirus Treatment', *International Studies in the Philosophy of Science* **16**, 179–193.
- Thagard, P.: 2003, 'Why wasn't O.J. Convicted? Emotional Coherence in Legal Inference', *Cognition and Emotion* **17**, 361–383.
- Thagard, P.: 2004, 'Causal Inference in Legal Decision Making: Explanatory Coherence vs. Bayesian networks', *Applied Artificial Intelligence* **18**, 231–249.

Department of Philosophy
University of Waterloo, Waterloo, Ont.,
Canada N2L 6G1
E-mail: pthagard@uwaterloo.ca