

# Bayesian Informal Logic and Fallacy

Kevin B. Korb  
School of Computer Science  
and Software Engineering  
Monash University  
Clayton, Victoria 3168 AUSTRALIA  
korb@csse.monash.edu.au

The theory of probability is at bottom nothing more than good sense reduced to a calculus.

— Pierre Laplace (1814)

**Abstract.** Bayesian reasoning has been applied formally to statistical inference, machine learning and analyzing scientific method. Here I apply it informally to more common forms of inference, namely natural language arguments. I analyze a variety of traditional fallacies, deductive, inductive and causal, and find more merit in them than is generally acknowledged. Bayesian principles provide a framework for understanding ordinary arguments which is well worth developing.

## 1 Introduction

Sophists, and the exercise of their profession — sophistry, have received very bad press, ever since Plato, who, together with his mentor Socrates, was an implacable foe of the sophists' ideas and methods. Indeed, the pair were so successful that labeling one a “sophist” has been a favored *ad hominem* attack on an antagonist ever since.<sup>1</sup> But the press, as often happens, have got hold of a half-truth and represented it as the whole. The sophists engaged in a kind of applied philosophy, for money — a practice which has only recently re-emerged in philosophical “therapy” and bioethical and environmental consulting. In particular, the sophists instructed people how to argue persuasively, a matter of considerable interest in the world's first democracies. Despite this focus on persuasion, the sophists were never so uninterested in the truth as Socrates and Plato would have us believe, nor were their sophistical arguments ever so unlikely to lead to the truth.

Here I shall not attempt to rehabilitate the reputations of sophists and their sophistries per se, as mine is not a historical exercise, but a logical one. Rather, I shall attempt to rehabilitate the reputations of certain *fallacies*, modes of reasoning which have commonly been taken to lead from truth to falsity, or, at any rate, to provide no guidance towards the truth.<sup>2</sup> Some of this work has been done previously, since the status of the various fallacies has always been a central and controversial issue for *informal logic*, the study of arguments in their natural setting (natural language) rather than in the formal setting of mathematical logic.<sup>3</sup>

---

<sup>1</sup>See, for example, his *Euthydemus*. Even before Plato, Aristophanes castigated the sophists (and Socrates) in his play *The Clouds*.

<sup>2</sup>Fallacies are often *defined* as forms of reasoning which are invalid — meaning, in the case of deductive argument, those that do not *necessarily* preserve the truth of the premises, and, in the case of inductive argument, those that do not lead to probable truth.

<sup>3</sup>For a brief history of informal logic see Johnson and Blair (1997).

I shall treat three varieties of reasoning: deductive, inductive or statistical, and causal. All three are host to a large variety of fallacies. And all three have given rise to substantial literatures within cognitive psychology, as psychologists have attempted to identify the fallacies to which people are prone and to understand the reasons why we are susceptible in those ways (e.g., for work on deduction, induction and causality respectively see Wason, 1960; Tversky, Slovic, and Kahneman, 1982; and Sperber, Premack and Premack, 1995). It is perhaps worth noting that my treatment of such a broad range of types of reasoning is intended to be indicative rather than comprehensive; the extension of Bayesian analysis beyond the select cases I examine should, however, frequently be straightforward.

Some efforts to explain the results of the cognitive psychology of human error have aimed at defending ordinary modes of reasoning using an evolutionary line of thought. Essentially, the thought is that human modes of reasoning must be efficacious in arriving at the truth, since arrival at the truth is a precondition of survival and reproduction in many actual scenarios. Therefore, despite the claims of the psychologists that many of these forms of reasoning fail to match our best normative standards, including Bayesian reasoning principles, and so are fallacious, they are to be endorsed as good, if imperfect, discoverers of the truth. Arguments along these lines are due to L.J. Cohen (1981) and Chase, Hertwig and Gigerenzer (1998).

My approach is quite different: rather than supposing that there is something *wrong* with the normative standards that the psychologists appeal to when describing some human forms of reasoning as illusory, I shall suppose that there is something *right* with those standards. In particular, I suppose that Bayesian principles are indeed *normative*, prescribing how we ought to reason. In the case of some “fallacies,” however, what is illusory is the Bayesian standard: the Bayesian principles have either not been applied or they have been misapplied. When applied, Bayesian reasoning fully or partially endorses the “fallacy.” This kind of point is perfectly compatible with the possibility that less than fully normative heuristics provide evolutionarily useful guidance in decision making. It is a continuing difficulty for Bayesian theory that its principles are very easy to misunderstand and misapply, as has been repeatedly demonstrated in the philosophical, statistical and psychological literatures. I hope that my treatment of the fallacies may help in this regard.

## 2 What is a Good Argument?

Let us first consider what makes a good argument. Chris Hamblin, in his *Fallacies* (1970), launched an influential attack on the standard alethic conception of a good argument as a sound argument — one with only true premises and valid inferences. The first objection, and a telling one, is that such a criterion may be accidentally fulfilled, our premises may be true even though we have no good reason to believe that they are. In such a case, we wouldn’t want to accept that the argument was good; the concept of an *accidentally* good argument appears to be empty. The goodness of our arguments is an achievement that we are responsible for, and so if our premises in fact need to be true, that truth must be something we can account for, rather than luck into.

Such considerations lead quite naturally to an epistemic criterion for good argument: an argument is good just in case its premises are *known* to be true and its inferential force apparent. Against both this epistemic conception of argument and its alethic predecessor, Hamblin argues that they ignore the argumentative context and, in particular, the relation between the argument and its intended audience. Instead, Hamblin encourages us to adopt a *dialectical* view: a good argument is one that *persuades* us of the truth of its conclusion — and so has *accepted* premises and compelling inferences (1970, p. 242). “One of the purposes of argument, whether we like it or not, is to convince” (p. 241). The classic view of argument, whether alethic or epistemic, aims exclusively at identifying the normative character of good

arguments.

Ralph Johnson (1996, chapter 9) expresses a number of reservations about involving dialectical criteria. Knowledge of the intended audience is required in order to determine what premises might be acceptable as a background to argumentation. Johnson wonders whether we are in fact stymied without having an audience in mind; in practice, people seem to produce good arguments without anyone in particular in mind. But I think it should not be too hard to accept that in such cases there is an implicit audience in mind; in any case, such arguments are clearly produced within a cultural context that settles for practical purposes the unargued acceptability or not of a large range of propositions. Nevertheless, Johnson is surely right that dialectical criteria *on their own* will not suffice (p. 178): “Suppose that I discover that my audience accepts a proposition which I know — or strongly believe — to be false, but which would, if accepted, provide strong support for the conclusion. According to dialectical criteria, it seems that I not only may but should use that proposition.” The relevance of dialectical standards for good arguments does not dispell the relevance also of normative standards.

I suggest that the best view of good argument is a fusion of Hamblin’s point that good arguments are persuasive and the classic view that they are normatively correct. An argument can fail to be a good one by employing identifiably false premises; it can fail to be good by employing weak inferences; and even when employing exclusively true, known and understood steps and premises, it can fail to be good by failing to persuade, by being directed at the wrong audience. As Ralph Johnson remarks, “[the] more fundamental purpose of argument is *rational persuasion*” (1996, p. 173). And indeed that purpose can only be served by arguments that are jointly rational (normatively grounded) and persuasive.<sup>4</sup>

### 3 Bayesian Reasoning

My focus here will be on the normative. More specifically it will be on inductive arguments, from a Bayesian perspective, rather than from the more common point of view, that the logic of arguments can best be understood as an application of the logic of mathematics, or formal logic. Although this is the less common perspective, it is not exactly new — dating from before the 18th century and Bishop Butler’s pronouncement that “probability is the very guide to life”. Regardless of its heritage, and despite considerable activity in developing a Bayesian scientific methodology, the Bayesian perspective on good argument has yet to be articulated.<sup>5</sup>

In order to develop such a perspective I first need to describe the normative standard that I intend to apply; so, I briefly describe Bayesian reasoning and its conditions for application.

Bayes’ theorem, a discovery of Thomas Bayes (1763/1958), reports the relation between the probability of a hypothesis (conclusion) given its evidence —  $P(h|e)$  — and its probability prior to any evidence  $P(h)$  and its likelihood  $P(e|h)$ , the probability that the hypothesis implies for the given evidence. In particular,

$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)} \tag{1}$$

The theorem is just that and is not controversial. Bayesian evaluation theory is controversial: it asserts that the proper way to evaluate hypotheses just is to adopt the probabilities

---

<sup>4</sup>In a series of papers on automated argumentation I, and my collaborators, have called such arguments *nice* rather than good — admittedly, just to give rise to the acronym *NAG* (Nice Argument Generator); e.g., Korb *et al.* (1997).

<sup>5</sup>Not that there has been no work in this area. The work of Polya (1968) in particular comes to mind. But although his ideas are suggestive and aimed in the same general direction as my own thoughts, they appear to have been incomplete thoughts.

conditional upon the available evidence — as supplied by Bayes’ theorem — as our new posterior probabilities; that is, we can abandon our prior distribution  $P(\cdot)$  in favor of our posterior distribution  $P'(\cdot) = P(\cdot|e)$ . This move to a posterior distribution is called *Bayesian conditionalization*.

Given this view, the simplest way of understanding the concept of the confirmation or support offered by some evidence is as the difference between the prior and posterior probabilities of a hypothesis; that is,  $e$  supports  $h$  just in case  $S(h|e) =_{df} P(h|e) - P(h) > 0$  (cf. Howson and Urbach, 1993, p. 117). A second measure of support, the ratio of likelihoods  $e$  given  $h$  over  $e$  given not- $h$ , is equally defensible:<sup>6</sup>

$$\lambda(e|h) =_{df} \frac{P(e|h)}{P(e|\neg h)}.$$

It is a simple theorem that the likelihood ratio is greater than one if and only if  $S(h|e)$  is greater than zero.  $\lambda(e|h)$  (or, simply,  $\lambda$ ) can be understood as a degree of support most directly by observing its role in the odds-likelihood version of Bayes’ theorem:

$$O(h|e) = \lambda O(h).$$

This simply asserts that the conditional odds on  $h$  given  $e$  should equal the prior odds adjusted by the likelihood ratio. Since odds and probabilities are interconvertible ( $O(h) =_{df} P(h)/P(\neg h)$ ), support defined in terms of changes in normative odds measures changes in normative probabilities quite as well as  $S(h|e)$ .  $\lambda$  has a significant advantage over  $S(h|e)$  however: it is simpler to calculate. Indeed, since a likelihood is just the probability of evidence given a hypothesis, and since hypotheses often describe how a causal system functions given some initial condition, finding the probability of the evidence assuming  $h$  is often a straightforward computation. What a likelihood ratio reports is the normative *impact* of the evidence on the posterior probability, rather than the posterior probability itself (i.e., the other necessary ingredient for finding the posterior is the *prior* probability of  $h$ ). However, confirmation theory is primarily concerned with accounting just for rational *changes* of belief, and so  $\lambda$  turns out to be a useful tool for understanding confirmation.

There are two other “tools” that a Bayesian analysis of argumentation delivers (or requires):

- *Priors*. Some critics of Bayesianism (e.g., Gigerenzer and Murray, 1987) take the manifest relevance in many cases of considerations about prior probabilities, together with the silence of pure Bayesian principles on the subject of how to choose priors, to imply the inadequacy of Bayesian theory. But this has matters exactly backwards. Unadulterated (unextended) Bayesian theory indeed does not specify how to find priors. But it is precisely Bayes’ Theorem which makes plain the relevance of the prior probability of a hypothesis for posterior belief, and so the inadequacy of methods ignoring priors. Considerations about rational priors do come into Bayesian analysis, as will be seen below, by way of possible extensions to unadulterated Bayesian theory.
- *Probabilistic Relevance (Conditional Dependence)*. One proposition is probabilistically relevant to another just in case conditioning on one changes the probability of the other — i.e.,  $a$  and  $b$  are probabilistically relevant if and only if  $P(a|b) \neq P(a)$  (or,

---

<sup>6</sup>Note that this is what Good called a Bayes factor (Good, 1983) and is not the same as what many statisticians call the *likelihood ratio*, which is instead a ratio of *maximum* likelihoods. That assumes that the hypothesis under consideration and its alternative are incomplete (indefinite) and can be parameterized so as to maximize the probability of the observed evidence. Here, however, I shall assume that we are dealing with definite hypotheses.

equivalently,  $P(b|a) \neq P(b)$ ).<sup>7</sup> Probabilistic relevance may vanish (or appear) between two propositions when the truth value of some further proposition(s) becomes known. For example, if John and Jane both ate the same soup at a restaurant, then one coming down with food poisoning will change the probability that the other will; however, if the common causal factor — that bacterial state of the soup — becomes known independently, then the health of John and Jane are rendered independent of each other. That is, their states of health are conditionally independent given the state of the soup. In the philosophical literature, this is known as “screening off,” following Hans Reichenbach (1956). Conversely, two independent propositions may become dependent given the value of a third proposition. For example, in judging whether to respond to a burglary alarm, a security company may discount the probability of a burglary upon learning that a small earthquake has struck, even though burglaries and earthquakes, without an alarm, are (approximately) independent. The importance of probabilistic relevance to understanding some fallacies will also be seen below.

## 4 The Justification and Limits of Bayesian Reasoning

### 4.1 Probabilistic Coherence

The key justification for Bayesian principles has been, and remains, the “Dutch book” arguments due to Frank Ramsey (1931). These arguments demonstrate an inability by a bettor violating the probability axioms to avoid a *guaranteed* loss from a sequence of fair bets, regardless of the outcomes of the individual bets. The assumptions required to obtain the Dutch book result are minimal: the bettor’s subjective degrees of belief must violate one of the probability axioms; the bets offered are individually fair (the odds are got by the ordinary ratio of probabilities given by the bettor); the bettor must be willing to take either side of a fair bet; the bettor must be willing to take any number of such fair bets.

Some have found the arguments insufficient reason to take the probability axioms as a normative guide for subjective degrees of belief. For example, various commentators have suggested that a willingness to take any number of bets, even when they are fair, is itself irrational, and have naively claimed that therefore the Dutch book is defective (e.g., Bacchus et al., 1990; Chihara and Kennedy, 1979). However, the Dutch book does not presuppose that making such a series of bets is always rational; the point is rather that the subsequent guarantee of losses appears to be attributable only to the initial incoherence and not the willingness to take fair bets. It matters not that incoherent bettors can manage to avoid guaranteed losses by refraining from betting; the point is that, should they bet, their subjective beliefs are advising them wrongly.

### 4.2 Conditionalization

Matters are less clear cut for the justification of conditionalization. David Lewis (reported in Teller, 1973) produced a diachronic Dutch book, intended to show that violating conditionalization will lead to guaranteed losses on combinations of fair conditional bets. The argument turns out to fail in a variety of contexts, in particular when the new evidence reveals more about the weakness of one’s conditional probability structure than about the empirical world.

As Colin Howson (2000) has pointed out, conditionalization is the invalid inference rule:

$$\frac{P(B|A) = r, P'(A) = 1}{P'(B) = r} \quad (2)$$

---

<sup>7</sup>Also equivalently,  $a$  and  $b$  are probabilistically related just in case their mutual information measure is non-zero (see Cover and Thomas, 1991).

This inference will be valid when and only when a third premise is adopted:

$$P'(B|A) = P(B|A) \tag{3}$$

What distinguishes varieties of Bayesians from each other, as well as Bayesianism from its antagonists, is just what view is adopted concerning the range of applicability of equation (3). Extreme anti-Bayesians adopt the simple stance that (2) is wrong and (3) is therefore a fallacy, and so suppose themselves justified in ignoring Bayesian conditionalization universally. Extremist Bayesians, in some cases in reference to the diachronic Dutch book, take (2) to be universally valid and, hence, (3) to be universally true. But it is clear, as Ramsey (1931) himself pointed out, that there are many circumstances where the acquisition of new evidence will rationally upset our conditional probability structure. The simplest examples are those of Howson (1997), where adherence to (3) implies a *logical* inconsistency. But even before Howson produced such examples, it should have been obvious that the extremist Bayesian stance is hugely implausible, for it implies that a rational agent must in its very beginning adopt a complete conditional probability structure which can in the future never be rationally revised and only slavishly followed.

The normative standing of conditionalization arises from the fact that the retention of conditional probabilities, although not justifiable universally, is clearly defensible across a wide range of cases. This is most prominent in understanding scientific method: it would be an unusual case for the measurement  $e$  of an experimental outcome to lead us to revise its probability of discovery conditional upon any hypothesis  $h$ , that is, the likelihood  $P(e|h)$ . This is in part because likelihoods are just the kinds of things that our hypotheses generally supply to us. For example, if we assume a coin-tossing exercise to be a binomial process  $h$ , then the probability of any reported binary sequence  $e$  under that process is readily calculable, and is  $P(e|h)$ . There is ordinarily no room, simply upon discovering the particular outcome  $e$ , to argue for a revision of  $P(e|h)$ .<sup>8</sup> In many non-scientific cases, although the probabilities involved may be more obscure or qualitatively understood, the same principle applies: if the conditional probability structure is well thought-out in advance — possibly being given by a clear causal theory — then observing an experimental outcome will not disturb it, and (3) applies, with conditionalization being justified.

It is plain that such reasoning is not itself Bayesian. Bayesian principles obligate conditionalization given (3), but are silent on the whys and wherefores of adopting a prior probability distribution, and therefore of adopting a conditional probability structure (which is a function of the joint prior). The normative principles of Bayesian theory, in order to gain any useful leverage, must be supplemented by a theory of rational change of conditional belief structure — when, and why and how, (3) is violated.

I shall not here attempt any such supplementary theory. By developing and applying Bayesian theory to a wide range of cases in informal logic, however, it will become clear that such a theory must allow for the correctness of (3) across scientific and non-scientific domains of understanding, providing a large, if circumscribed, range for Bayesian reasoning.

### 4.3 Priors

The single most influential criticism of Bayesianism has been and remains: where do prior probabilities come from? There is a broad range of answers supplied by Bayesians; indeed, the different flavors of Bayesianism are largely determined by how one answers this question. Extreme objectivists (such as Carnap, 1950, and maximum entropists like Jaynes, 1989) find prior probabilities in the structure of language; extreme subjectivists (such as de Finetti, 1937) find priors in the wind. Moderates, such as Lewis (1980) and Good (1983), find them where they can get them.

---

<sup>8</sup>There are other cases, however, where the likelihoods are not so simply understood; cf. Korb (1994).

In the face of such variety, and controversy, over legitimate sources of prior probabilities, many prefer to opt out of the game entirely, and seek some method which doesn't require a subjective choice of one probability over another, nor obscure reasons for one variety of prior over another. Thus, for example, orthodox statistical inferences, such as significance tests and confidence interval estimations, do not require prior probability distributions. A confusing factor is that any specific such inference can be recast as a Bayesian inference, and, working Bayes' theorem in reverse, the corresponding prior probability can be inferred.<sup>9</sup> So, a sequence of such non-Bayesian inferences will determine a prior probability distribution — or else be demonstrably incoherent. Although many inferential methods do not require an explicit prior commitment to probabilities, they nevertheless, in application, require an implicit commitment to either probabilities or incoherence. This makes for an amusing game for Bayesian critics of those alternatives: finding cases where the prior probabilities implied by an inferential performance are manifestly absurd. The other side of the game is to define cases where the prior probabilities are irresistible. The operation of simple gambling apparatus provides indefinitely many examples. And in such cases the Bayesian analysis is more or less universally accepted.

#### 4.4 Formalism

It seems plausible that one of the reasons that Bayesian analysis outside the simple gambling cases has been widely resisted is that Bayesian analysis is founded on the application of a formal calculus: the simple cases are clearly and non-controversially formalizable, whereas the difficult cases all have non-formal aspects.

Consider the analogy of formal deductive logic (FDL). FDL, also, for a long time was considered a major candidate for formalizing human inferences, including even scientific inferences in the hypothetico-deductive methods of Nagel (1961) and Hempel (1965) and in the falsificationist method of Popper (1934). It has been typical of logic textbooks to claim, at least implicitly in the examples if not also explicitly, that FDL provides the normative standard for analyzing natural language arguments. That is, an argument is to be considered good (sound) if and only if both its premises are true and its translation into FDL produces a formally valid inference. However, this model for understanding natural language arguments has largely come unglued since 1970. Instruction in logic was widely assumed to provide benefits to students' ability to reason and argue. However, the empirical evidence suggests otherwise (van Gelder, 2000), and since around 1970 the related movements of informal logic and critical thinking have made substantial inroads in that supposed use of formal logic (cf. such texts as Kahane, 1971, and Scriven, 1976).

One substantial difficulty for FDL is that there is no univocal translation of typical natural language arguments into the language of formal logic: dealing with the ambiguities, complexities and hidden assumptions of ordinary arguments turns out to be more complex than anything to be found in deductive logic per se. But a more telling difficulty is that natural language arguments clearly and non-controversially come in degrees of strength, whereas the tools of FDL yield in all cases a simple two-valued verdict: valid or invalid. Precisely that same difficulty arises in all deductivist attempts to understand scientific inferences: in the end, for example, all Hempel's theory could say of any theory was that it was confirmed

---

<sup>9</sup>That is, such orthodox inferences issue in claims, implicit or explicit, about posterior probability, such as that  $P(h|e)$  is almost certainly false; since likelihoods are generally not controversial, we can apply Bayes' theorem to obtain the denied prior probability  $P(h)$  (cf. with equation 1):

$$P(h) = \frac{P(h|e)P(e)}{P(e|h)}$$

or not.<sup>10</sup>

So, if the formal methods of deductive logic are relatively useless for understanding natural language arguments or scientific reasoning, why should we believe that the formal methods of the probability calculus can do any better? Well, of course, regarding the last difficulty, the probability calculus was specifically designed to cope with *degrees of strength* for propositions, so at least Bayesian analysis is not so obviously the wrong tool. Nevertheless, it is certainly correct that Bayesian analysis, at least in the first instance, is formal and human reasoning is not, so it may seem that the tool and the problem remain quite unmatched. What I aim to show is that the formal, quantitative tools of Bayesian analysis support counterparts that are qualitative and semi-formal; these can shed light on what is normative and productive in analyzing arguments, without providing a complete solution to the problems raised by human argumentation. Bayesian philosophers have already done much to make such a case in the study of scientific method specifically.

#### 4.5 Successes in the Bayesian Analysis of Scientific Method

The concept of the likelihood ratio provides a simple but effective tool for analyzing the impact of evidence on conclusions (hypotheses). For example, it makes clear why Karl Popper's (1959) insistence that scientific hypotheses be subjected to severe tests makes sense. Intuitively, a severe test is one in which the hypothesis, if false, is unlikely to survive; that is, whereas the hypothesis predicts some outcome  $e$ , its competitors do not. Since the hypothesis predicts  $e$ ,  $P(e|h)$  must be high; since its competitors do not,  $P(e|\neg h)$  must be low. These jointly imply that the likelihood ratio is very high. Therefore, a severe test will be highly confirmatory if passed and highly disconfirmatory if failed — and so provides the most efficient approach to testing a hypothesis.

A second example is the preference which experimental scientists exhibit *ceteris paribus*, when confronted by two possible tests of a theory, for that test which is most different from one that the theory has previously passed.<sup>11</sup> For example, Eddington was faced with two approaches to testing Einstein's general theory of relativity (GTR) in 1919, either repeating the analysis of Einstein himself of the precession of Mercury's perihelion or checking the predictions which GTR made of a "bending" of starlight by the mass of the sun, observable during a total eclipse. Despite the fact that astronomical observations of the motion of Mercury are cheaper and simpler, Eddington famously chose to observe the starlight during the eclipse over the Atlantic. Why?

Intuitively, we can say that this is because testing a new experimental consequence, as opposed to a repeated experiment, offers a more severe test of the theory: the predicted new outcome is less likely to come about than the repeated outcome, and so when it does come about it offers greater confirmation of the theory under evaluation. A more formal Bayesian analysis directly supports this intuition. Where  $d$  is the prior experimental result,  $e$  its repetition and  $f$  some new variety of test result, the *ceteris paribus* clause above implies  $P(d|h) = P(e|h) = P(f|h)$ . Since  $e$  is closer to  $d$  than is  $f$ ,  $P(e|d) > P(f|d)$ . It follows directly from the Bayesian understanding of confirmation that  $f$  then has greater confirmatory power than does  $e$ , after observing  $d$ :<sup>12</sup>

$$P_d(h|f) = \frac{P_d(f|h)P_d(h)}{P_d(f)} > \frac{P_d(e|h)P_d(h)}{P_d(e)} = P_d(h|e)$$

---

<sup>10</sup>Similarly, all Popper could say of any theory was that it was falsified or not falsified, despite some extraordinary contortions to do with the "corroboration" of theories — which never quite managed to mean as much as confirmation.

<sup>11</sup>Franklin (1986) gave a prior, more complex, Bayesian analysis of this case.

<sup>12</sup>I employ  $P_d(\cdot)$  for simplicity rather than  $P(\cdot|d)$ , which is warranted by the fact that probabilities conditional upon some event satisfy the probability axioms.



Since

$$P_d(f) < P_d(e)$$

by assumption above and

$$P_d(f|h) = P_d(e|h)$$

because (what is typical, at any rate) we can take the probability of the evidence  $f$ , given that we know whether or not  $h$  is true, to be independent of the probability of the evidence  $e$  ( $h$  screens off  $f$  from  $e$ ). More applications of such Bayesian analysis to scientific method can be found in Howson and Urbach (1993) and Korb (1992).

## 5 Logical Fallacies

So, Bayesian analysis has a clear and unique justification in dealing with situations of uncertainty, in Ramsey's Dutch book treatment. Furthermore, it has had notable success in aiding our understanding of scientific inference. Clearly, Bayesian reasoning is at least a plausible candidate for assisting with our understanding of human inferences as manifested in ordinary argument.

Many of the forms of reasoning classically identified as fallacious have long been held *not* to be fallacious under various circumstances by various commentators. What has been missing, however, is a unifying theory supporting them which applies to more than a handful of cases. I believe Bayesian analysis, and, in particular, the examination of plausible prior probabilities and the relevant likelihood ratios, offers the skeleton of such a unifying theory. I will present Bayesian analyses of a few commonly discussed fallacies, illustrating in turn the application of likelihood ratios, probabilistic relevance and considerations about prior probabilities. The general applicability of these *types* of analysis elsewhere should be clear.

### 5.1 Affirming the Consequent: Reasoning with the Likelihood Ratio

Affirming the consequent is perhaps the most blatant of fallacies. It takes an assumed conditional and, instead of applying to it the asserted truth of its antecedent as in *modus ponens*, it attempts to reverse the process, as in:

All humans are mortal.  
Socrates is dead.

---

Therefore, Socrates is human.

Despite being the most blatant of fallacies, and disparaged in popular texts on logic, this form of argument is *pervasive* in the sciences:

If evolutionary theory is true, we would expect to find antibiotic-resistant strains of bacterial diseases appearing.  
We do find that!

---

So, indeed, evolutionary theory is supported.

The only reason we do not see this particular argument with any frequency is that evolutionary theory by now has been so well confirmed that the argument is utterly unnecessary.<sup>13</sup>

---

<sup>13</sup>Never mind that there are those with minds closed against evolutionary theory; empirically based arguments will never have a significant impact if you start out with either sufficiently biased priors or sufficiently warped conditional probabilities. (And, if to the anti-Bayesian this appears to be a fatal concession — because it shows that it is possible to be an irrational Bayesian — that is a common but mistaken conclusion. It is the demand for a Bayesian theory that answers all problems about rationality in all circumstances that is most irrational.)

This kind of reasoning is so wide-spread that, in addition to being a named fallacy, it has a name as an accepted form of reasoning, namely, “inference to the best explanation” (IBE). Probabilistically, what is going on is an inference of the following form:

$P(e|h)$  is high.  
 $e$  is observed.

---

Therefore,  $P'(h)$  is high.

Comparing this with the Bayesian requirement for confirmation, that the likelihood ratio be high, we see that there is a suppressed premise, that the alternative hypothesis should fail to support the evidence, i.e., that  $P(e|\neg h)$  is low.<sup>14</sup> In the case of Socrates, there are a great many explanations of Socrates’ death alternative to Socrates being a human, as many as there are species, and so IBE fails.<sup>15</sup> In the case of evolution, there is no serious alternative explanation to the development of antibiotic resistance — the alternatives to some form of evolutionary theory, plausibly construed, make such a development a miracle. Hence, the Bayesian criterion for high confirmation of evolution theory is satisfied.

This analysis not only endorses the value of “affirming the consequent” in some circumstances, as others have done before, but also clarifies the demarcation between such circumstances and those others which led to its identification as a fallacy in the first place. Affirming the consequent is emphatically *not* a fallacy. It is a central feature of scientific inference, in fact, quite properly deserving its special title as “inference to the best explanation.” But neither is it an unexceptionable form of inference. It is only in Bayesian considerations of likelihood and prior that a principled distinction between good and bad applications of this form of reasoning has ever been made.

## 5.2 The Appeal to Popularity: Probabilistic Relevance

In some circles (or circumstances) it is popularly believed that there are witches; in others, it is believed that hairless aliens walk the planet. If a bald appeal to the popularity of a belief were enough to establish its acceptability, then reasonable beliefs and the arguments for them would ebb and flow with the tides of fashion.

Nevertheless, there seems to be *some* merit to the appeal to popular belief. Johnson (1996) points out a *direct* relevance between popular belief and (propositions concerning) the outcome of democratic elections! But even when there is no direct (or indirect) causal chain leading from popular belief to the truth of a proposition, there may well be a common cause that relates the two, making the popularity of a belief a (possibly minor) reason to shift one’s belief in a conclusion. Presumably, a world in which no one believes in witches would support a moderately smaller rational degree of belief in them than one in which many do, at least prior to the development of science. In general, open mindedness is at least partially characterized by a willingness to be persuaded that one’s beliefs may be wrong, and perhaps in the first instance simply by the sheer number of people of different mindedness. The Bayesian point of view accommodates all of the above observations, and again supports the demarcation of telling and pointless appeals to popularity via the Bayesian standard of probabilistic relevance.

As  $\lambda$  (and the related measures of Bayesian confirmation) comes in degrees, another immediate consequence of the Bayesian analysis is that relevance also comes in degrees. In our world the popular belief in witches, for example, can be understood to be relevant, but only in a minor way that can be swamped by (or screened off by) other, more strongly relevant,

---

<sup>14</sup>To be sure, it is also required that the prior  $P(h)$  not be too low prior to asserting  $P'(h)$  is high.

<sup>15</sup>Interestingly, the proponents of IBE, such as Gilbert Harman (1965) and Peter Lipton (1991), seem not to have appreciated these additional requirements, which in fact imply that IBE taken as a general rule is defective.

information. That is, our situation may well be describable via Figure 1. Popularity of a belief may well in general be associated with the truth of what is believed; so, lacking any clear scientific judgment (say, during the Dark Ages), common belief in the efficacy of witchcraft may well rationally lift our own belief, if only slightly. Nevertheless, given an improved understanding of the science of belief and a science of natural phenomena (say, now), the popular belief is no longer relevant for deciding whether witches exist or not: the former account for both the belief and the reality.

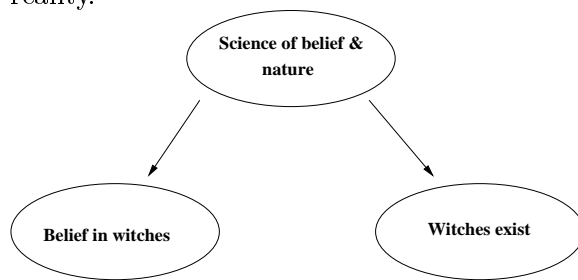


Figure 1: Belief in witches screened off

### 5.3 Ad Hominem: Probabilistic Relevance and Priors

Argument *ad hominem* is directing criticism at the presenter of some original argument rather than at the original argument itself. Its characterization as a fallacy implies that there is an attempt to deflect attention from the original argument — that is, that the *ad hominem* is a form of red herring, irrelevant to the original question. Walton (1989, p. 151) gives a nice example:

[An] instance of [an] *ad hominem* imputation of bias occurred during a debate on abortion in the Canadian House of Commons: “It is really impossible for the man, for whom it is impossible to be in this situation, to really see it from the woman’s point of view.”

As Walton notes, it is of course correct that a man cannot be in the woman’s situation, but the suggestion that this makes it impossible for the man to “see” the situation from the woman’s point of view fails to address anything substantive the man may have said.

Despite its being a “classic” fallacy, many have expressed strong doubts that such argument should generally be considered fallacious. Joseph (1906) pointed out that it is standard practice in court to consider the reliability of witnesses, which would not be defensible were argument *ad hominem* fallacious without reservation. The real question is whether or not, in a particular case at hand, the question raised about the human arguer is relevant to the original issue. If  $a$  is the *ad hominem* attack and  $h$  the original hypothesis under contention, what we would like to know is whether or not the truth of  $a$  would rationally influence our belief in  $h$ ; i.e., whether  $P(h|a) \neq P(h)$ .

If a liar is exposed in court, then surely that is relevant. If an anti-abortionist is exposed as male, the relevance is, at best, obscure. In these cases, the relevance (or irrelevance) is based upon the origin of the testimony. Although reference to the origin of an argument has commonly been denounced as the “genetic fallacy,” so long as the plausibility or acceptability of any of the premises in an argument relies to any degree upon the believability of the one putting them forward, argument *ad hominem* and the “genetic fallacy” will be pointed and relevant. The impulse to discount these forms of argument as fallacy appears to stem from an over-idealized view of argumentation: one where the merits of every argument can be assessed in abstraction from all context. In the real world, considerations of time and circumstance are inescapable, and ignoring the reliability of witnesses is simply irresponsible.

## 5.4 Epistemic Direct Inference: Probabilistic Relevance and Priors

This leads us straight to the issue of *direct inference* (or, the “statistical syllogism”) since it is a primary way by which we assess probabilities relative to sources, and so a plausible source of some of the prior probabilities needed to operate Bayes’ theorem. Formally stated, direct inference is the inference rule:

*Direct Inference (DI)*: If the frequency of objects of type  $A$  within the reference class of objects of type  $B$  is  $r$ , that is,  $F(A|B) = r$ , then if object  $x$  is selected from  $B$  with a physically uniform randomizing device (and the agent knows that), then  $P(x \in A) = r$ .

This is what endorses, for example, the assertion that the probability of drawing a red ball from an urn containing equal number of red and blue balls (and the result being from a known physically uniform selection device) is  $1/2$ . This is not a controversial rule. However, since it requires a *physically* random selection device, its preconditions are rarely satisfied.

A more widely applicable rule generalizes direct inference to cases where the properties of the selection device are unknown:

*Epistemic Direct Inference (EDI)*: Suppose  $F(A|B) = r$  and the agent has no reason to believe that the selection process for  $x \in B$  is biased with respect to membership in  $A$  (this is a kind of total evidence condition). In particular, the agent does not know of any proper subset  $C \subset B$ , with a known frequency  $F(A|C)$ , both that  $F(A|C) \neq F(A|B)$  and that  $x \in C$ . Then it can defeasibly infer that  $P(x \in A) = r$ .

EDI is of course a fallible, inductive rule. As such, it has been subjected to attack, for example by the notable Bayesian Isaac Levi (1980). It is easy to develop cases where it will lead astray. And it is quite common for the preconditions of EDI to be satisfied by multiple, competing reference classes. Thus, for example, we may know that 10% of the English are Catholics and that 5% of academics are the same, without knowing anything useful about the intersection set, English academics. This is the “problem of the reference class,” which has yet to find a generally compelling solution. Regardless, EDI is a rule which humans and other animals use widely and successfully, at any rate to all appearances. And compelling or not, there are proposals which are probably already a satisfactory first step towards modestly successful autonomous agents, such as that of Reichenbach (1949), to employ the smallest reference class for which we have reliable statistics. It is in fact hard to imagine how a species in a complex environmental niche could survive for very long if it did not use EDI or some heuristic rule that mimicked it. The burden is surely on the critics to supply an alternative rule that is as useful as EDI.

Given EDI, we can make immediate sense of arguments *ad hominem*, as either irrelevant or relevant. A successful *ad hominem* shows that the presenter of the argument comes from a reference class other than the one presumed, and indeed a reference class that biases matters, whether favorably or unfavorably. Thus, an appeal to expert authority identifies a favorable bias for a prior probability to be attached to an assertion attributed to the authority. The appeal is reasonable and relevant if the authority’s expertise is pertinent to the assertion at issue. Similarly, an unfavorable biasing factor is introduced when a witness is discovered to be a liar. Even then, the argument *ad hominem* may fail to be relevant if the original argument in fact does not actually rely upon the credibility of its presenter — if, for example, all of its premises are common knowledge or previously accepted. In such a case an *ad hominem* argument cannot but be irrelevant.

**The Value of Saying** An interesting related phenomenon is the “value of saying.” When a proposition is stated, as opposed to being left implicit in an argumentative exchange, there

are at least two distinct effects on the hearer:

- An attentional effect — the hearer’s cognitive apparatus will be differently deployed, in order to interpret the statement. Semantic and probabilistic connections with the statement will be followed, or followed further, resulting in some shift in the degrees of belief in associated propositions.
- A probability shift — the hearer’s degree of belief in what is said may shift. In particular, the fact that the interlocutor has made the utterance will tend to shift the hearer’s belief in the statement up or down, depending upon the credibility of the speaker, in an application of EDI. That is, if the speaker has a credibility lower than the hearer’s initial belief in the statement, the belief in the statement will diminish, and vice versa. The degree of shift will depend both upon the discrepancy between the speaker’s credibility and that initial belief and upon the conviction with which the hearer holds that initial belief, as reflected in the hearer’s conditional probabilities.

## 5.5 Logical Fallacies in Sum

It should be fairly clear already that the Bayesian tools for argument analysis — prior probabilities, likelihood ratios and relevance — cannot reasonably be expected to resolve all issues arising in informal logic. For example, any semantic analysis of arguments is simply outside the scope of this kind of Bayesian analysis. So, the classic fallacy *A gray elephant is a gray animal, so a small elephant is a small animal* depends upon an equivocation in the predicate “is small.” As such, no Bayesian analysis per se will spot the problem.

The prerequisites for Bayesian conditionalization to support a hypothesis — stable conditional probabilities, some significant prior probability, and a likelihood ratio greater than one — provide tools for analyzing the merits of various kinds of argument, as I have shown. The analysis can clearly be extended to additional varieties of argument, providing a common and principled theme to this work of informal logic. Semantics, the origin and revision of conditional probability structures, and the origin of prior probabilities are matters that lie beyond any ordinary Bayesian analysis. So too does the story of how normative arguments may or may not persuade the target audience, to which rhetoric and psychology contribute. Demanding, or expecting, Bayesian principles either to deliver the *complete* story of argumentation or else to stand exposed as inadequate sets an impossible standard for any formal system. Bayesian analysis clearly can deliver much of what is important. But the point of this Bayesian analysis completed will be simply to provide a useful *framework* for identifying and dealing with inferential relations between statements, one which supplements and must be supplemented by considerable additional analytic resources.

## 6 Statistical Fallacies

The psychological literature on statistical fallacies has been developed largely by Bayesian psychologists, that is, by psychologists who accept Bayesian principles as their normative standard. Hence, we should expect to find that the fallacies which they identify are indeed fallacies from the Bayesian perspective. However, it ain’t necessarily so. Consider Kahneman and Tversky’s best known heuristic: the Representativeness Heuristic. This was introduced in order to explain why many people reason in a way obviously violating probabilistic rules. According to Tversky and Kahneman (1973), with this heuristic “an event is judged probable to the extent that it represents the essential features of its parent population or generating process.” Thus, in one notorious experiment, a hypothetical Linda, described as an active leftist when a student, is judged more likely to have subsequently become a feminist bank

teller than... a bank teller! — a conclusion violating the rules of probability. Tversky and Kahneman’s extremely plausible suggestion is that people find the concept of Linda more stereotypical of the subset of feminist bank tellers than of bank tellers generally, and then they substitute this statement of stereotypicality for the requested response of which vocational outcome is more likely.

The moral that Tversky and Kahneman would like their readers to draw, of course, is that this latter substitution of stereotypicality for probability is a mistake, that we humans have got it fundamentally wrong. For example, in summarizing their work on representativeness, they write “In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all” (Kahneman and Tversky, 1972). I would like to suggest that this is quite a wrong interpretation (and not just because of the sexist implication that women are superior to men).

It has long been known that strict Bayesian inference is computationally complex; that is the main reason why computerized expert systems, until quite recently, never used proper probability computations (cf. Neapolitan, 1990, chap. 4). Indeed, Bayesian inference has been proven to be NP-hard (Cooper, 1990), which means in practice that complete algorithms to perform such inference require exponential increases in time as the problem to be dealt with increases (e.g., as the number of variables increases). In short, it is too much to expect computers to perform strict Bayesian inference; it would then be absurd to expect humans to do so, when relative to computers we are so much worse at arithmetical computations. If computers require heuristic short-cuts to approximate full Bayesian inference, then so much more do humans require heuristic short cuts.

So what might such heuristic short cuts look like — rules that approximately follow the normative standard, at least in common cases, even if they break down under less usual circumstances? Given the performance pressures on humans and their ancestors in evolution, in addition to yielding true or “good enough” answers in run-of-the-mill circumstances, it would undoubtedly also be advantageous if such heuristics were quick and easy to compute. We can call such rules *fast and frugal* heuristics. So, what are they? Well, that is a moot point, but one well worth investigating in the cognitive psychology of human inference. A very promising candidate for one though is the representativeness heuristic. I suggest that it has all of these characteristics: (1) It is relatively fast. For obvious reasons humans have evolved substantial capacities to recognize and categorize objects and events. The use of stereotypes is clearly a part of this set of abilities. And there is compelling evidence, for example, that humans are better and faster at categorizing stereotypical members of classes than non-stereotypical members (Rosch, 1976; Rips, 1975). (2) Given all of the mental capacity already dedicated to recognition and categorization, and its relation to stereotypicality, frugality follows. (3) It is also clear that in a very broad range of cases, the stereotypical answer is also the right one. This is clear on evolutionary considerations, since we (and other animals) demonstrably employ such reasoning tendencies and so were they commonly seriously suboptimal we (and such other animals) would no longer exist. But also on statistical grounds, use of the stereotype (mode) of a class to infer properties of an unknown member of the class must in a very large range of cases yield a practically useful answer (see, for example, Holte, 1993, and Korb and Thompson, 1994, for statistical studies showing the effectiveness of classifiers drawing upon such extremely simple classification models). You can and should judge a book by its cover, if you haven’t the time to examine its contents.<sup>16</sup>

Many other heuristics investigated by Bayesian psychologists, such as availability and anchoring (see Kahneman, Slovic and Tversky, 1982), have the same properties. Indeed, the

---

<sup>16</sup>I am indebted to Gigerenzer and Todd (1999) for the phrase “fast and frugal heuristic,” and indeed they discuss some such heuristics that have the same sort of characteristics as representativeness, availability and other heuristics examined by Bayesian psychologists — i.e., accuracy in common cases, ease and speed of computation — although curiously they do not investigate those heuristics themselves.

Bayesian analysis of the logical fallacies can be viewed in the same way: many of the fallacies are fast and frugal heuristic techniques for inference, which nevertheless can lead us astray. Considering how we might improve our reasoning abilities, a question at least underlying much of the work of cognitive psychology, it would be more useful to examine the circumstances in which our heuristics fail us, and what might flag such circumstances, rather than simply decry our inability to reasoning cogently. Labelling these heuristics as fallacies tends to act as a substitute for critical reasoning about them, rather than an encouragement to reason critically.

## 7 Causal Fallacies

### 7.1 Post Hoc

The classic “causal fallacy”, dreaded by many generations of logicians, starting with the first one, Aristotle, is *post hoc ergo propter hoc* — that is, the argument that “because B happens after A, it happens because of A” (Aristotle, *Rhetoric*). As Pinto (1995) points out, this characterization is ambiguous. Many recent writers have interpreted it as an inference concerning particular token events: because A in fact occurred before B, it caused B (e.g., Rottenberg, 1991; Copi and Cohen, 1990). No doubt a more useful interpretation is available in terms of event types: that because events of type A are followed by events of type B, As cause Bs. This is a more useful interpretation because in token form the fallacy is strictly counterfactual — it is one which people do not in fact commit; from the *mere* fact that a particular event preceded another, practically *no one* concludes a causal relationship. How often is it said that John Wilkes Booth caused the death of Kennedy, or the invention of the piano initiated World War II? The denunciation of *post hoc* in these terms is simply fatuous.<sup>17</sup> What *might* ground a singular causal claim is the acceptance of the corresponding claim relating event types. If we may be warranted in claiming that consuming anti-oxidents prevents cancer in general, then we may be warranted in asserting the same in particular cases as well.

### 7.2 Correlation and Causation

The denunciation of *post hoc*, in type language, is closely allied to the dictum of many statisticians: “correlation does not imply causation.” Indeed, Walton (1989) simply identifies the two: *post hoc* occurs “when it is concluded that A causes B simply because one or more occurrences of A are correlated with one or more occurrences of B.” This identification of Walton’s is mistaken because a defining characteristic of *post hoc* is the known temporal precedence of cause to effect, whereas correlation is a symmetrical relation. Of course, those who oppose the statisticians’ dictum, including me, do not propose that we infer causal relations which do not respect temporal ordering; rather, we propose that causal relations may be inferred even lacking explicit temporal knowledge. In artificial intelligence methods to do this are known as causal discovery algorithms. Now, if these methods can be demonstrated to be successful, and so the statisticians’ principle demonstrated to be false or misleading, then it would be a simple corollary that the denial of *post hoc* also must collapse.

The denial of the basis for causal discovery — learning causal structure from correlational information — continues in the statistical literature (see the recent debate in McKim and Turner, 1997); however, since the mid-1980s, the case against it has thinned considerably. A

---

<sup>17</sup>Nor does imposing a condition of temporal proximity save the charge of fatuousness. Although people commonly infer that hitting a light switch turns on a light, it is hardly just a temporal relation they are relying upon, since analogous inferences are withheld regarding the overwhelming majority of simultaneous cases of hitting a switch around the world.

clear and compelling philosophical argument for causal discovery was already put by Glymour *et al.* (1987, part I). But more compelling for some will be the many and varied successes of causal discovery algorithms in practice, which have been reported over the last decade in the artificial intelligence literature. Every volume of the annual conference *Uncertainty in AI* in the 1990s contains multiple reports of new and successful applications of causal discovery algorithms. The technique has been remarkably successful for something founded on a mere fallacy! Despite this, the nay-sayers continue to be unimpressed; for example, Humphreys and Freeman (1996) suggest that it is all a Baconian (Cartesian?) dream of mechanizing thought, which will all come crashing down in a heap. Of course, a Humean concern might give any inductive program pause, but no reason has yet been produced for thinking causal discovery is in that respect any more vulnerable than any other scientific activity.<sup>18</sup>

What causal discovery does is precisely to distinguish those cases of correlation between A and B which are best explained by a causal structure between them from those which are best explained otherwise. How can that be done? The basic insight was codified already by Hans Reichenbach (1956). In his “Principle of the Common Cause” he asserted (1956, p. 157): “If an improbable coincidence has occurred, there must exist a common cause.” By this formulation Reichenbach did not mean to be ruling out the possibility that, given coincident phenomena, one might cause the other, directly or indirectly. Rather, this formulation becomes active just in case we can rule out such a connection — so, the Common Cause Principle in effect is: reliable, reproducible *coincidences* do not occur — magic does not exist. The statisticians’ denial of the rationality of inferring causation from correlation simply leaves the reliable, persistent correlation inexplicable, which is tantamount to endorsing magic.

Following Reichenbach, the causal inference is based upon probabilistic independencies that are revealed in observed correlations. Thus, considering a system with three variables, there are only the following possible causal structures (assuming the pairs  $\langle A, B \rangle$  and  $\langle B, C \rangle$  are directly related):

- a.  $A \rightarrow B \rightarrow C$
- b.  $A \leftarrow B \leftarrow C$
- c.  $A \leftarrow B \rightarrow C$
- d.  $A \rightarrow B \leftarrow C$

Reichenbach called (c) a “fork open toward the future” and (d) a “fork open toward the past” and pointed out that they support distinct conditional independence structures: (c) — as well as (a) and (b) — have  $A$  and  $C$  marginally dependent, but independent conditioned upon  $B$  (i.e.,  $P(A|C) \neq P(A)$  but  $P(A|C, B) = P(A|B)$ ); (d), exactly to the contrary, has  $A$  and  $C$  marginally independent, but conditionally dependent (i.e.,  $P(A|C) = P(A)$  and  $P(A|C, B) \neq P(A|B)$ ).

All causal discovery algorithms are ultimately based upon this simple distinction. And although this basis may seem quite small, making only a binary distinction between two sets of models, the recursive application of the principle over models with many variables turns out to be very powerful (see Verma and Pearl, 1990). Since different causal models give distinct probabilities to data reflecting a given conditional independence structure, they have different likelihoods, and so Bayes’ theorem yields different posterior probabilities. For example, if some observational evidence  $e$  were to strongly support the conditional dependence of (d) (that  $P(A|C, B) \neq P(A|B)$ ), then the likelihood of (d) on that evidence would be much greater than the likelihood of any of the three other models, and on Bayesian grounds (d) would be strongly confirmed. Such Bayesian inference reasoning can be automated, and quite

---

<sup>18</sup>My rebuttal to Humphreys and Freeman (1996) is in Korb and Wallace (1997).



effectively for discovering the true model underlying the data (see, e.g., Wallace and Korb, 1999). And, so, Bayesian analysis once again exposes the denunciation of a mode of reasoning as fallacious as itself fallacious.

## 8 On Improving Our Probabilistic Reasoning

So, we have seen that Bayesian reasoning can deliver more informed and insightful verdicts on the merits and demerits of arguments in informal logic, the correctness of forms of statistical reasoning, and even whether a causal inference is justified or not. If Bayesian reasoning is so valuable, it would be good to learn how to learn it, or do it, because the evidence on the whole is that people find Bayesian reasoning very difficult to do (e.g., Nisbett, et al., 1987).

There are at least four different approaches that can be taken to using Bayesian calculation in practice. To illustrate, I present them in the context of an example of breast cancer diagnosis adapted from Hoffrage and Gigerenzer (1998; itself adapted from Eddy, 1982).<sup>19</sup>

### 8.1 Bayes' Theorem

The most direct way to do Bayesian reasoning is simply to employ Bayes' theorem. Suppose that we are presented with a woman appearing at a clinic whose initial test for breast cancer has proved positive. We are asked to estimate the chance that the woman indeed has breast cancer. As background, we are told that one in one hundred women appearing at the clinic have breast cancer and that the tests are positive given cancer 80% of the time and they also are positive 10% of the time in women without cancer.

The most common response to this kind of scenario is to assert that the woman has an 80% chance of having breast cancer. It is not clear *why* most people respond this way. One possibility is just that people tend to find conditionals confusing and, in particular, confuse conditionals and their converses. In this case, asserting an 80% chance of cancer confuses  $P(\text{cancer}|\text{positive test})$ , which is the posterior probability that we were asked for, with  $P(\text{positive test}|\text{cancer})$ , which is the likelihood given in the scenario as 0.8. Some Bayesians would like the confusion between posterior and likelihood to explain the entire mistaken tradition of maximum likelihood-oriented orthodox statistics, which is probably putting too much of a burden on a simple confusion. Tversky and Kahneman (1982) dubbed this mistake “base-rate neglect”, since it can arise by suppressing the term for the prior in Bayes' formula.

In any case, the correct computation using Bayes' formula is:

$$\begin{aligned} P(\text{cancer}|\text{pos}) &= \frac{P(\text{pos}|\text{cancer})P(\text{cancer})}{P(\text{pos}|\text{cancer})P(\text{cancer}) + P(\text{pos}|\text{no cancer})P(\text{no cancer})} \\ &= \frac{.8 \times .01}{.8 \times .01 + .1 \times .99} \\ &= \frac{.008}{.008 + .099} \end{aligned}$$

---

<sup>19</sup>Although Gerd Gigerenzer is the author of one of these methods, curiously he has spent much energy and time questioning Bayesian reasoning as a normative standard (e.g., Gigerenzer and Murray, 1987; Gigerenzer, 1991; Gigerenzer and Todd, 1999). Most recently, he has objected to Bayesian reasoning on the ground that there is no plausible evolutionary story to tell about how such a reasoning ability *could* evolve, given its computational complexity, citing the work of Tooby and Cosmides in evolutionary psychology (e.g., in Chase *et al.*, 1998). Presumably, the idea is that, as a general principle, what cannot be done also cannot be held up as a normative standard of behavior, so Bayesian reasoning, being unavailable to us, offers us no standard. Perhaps the oddest twist in this story is that Gigerenzer's own technique for Bayesian reasoning promises to make Bayesian calculation far more accessible, using his “frequency formats” (Gigerenzer and Hoffrage, 1998; Gigerenzer, 1996). Gigerenzer is certainly right that for Bayesian principles to provide an effective standard they must be made possible to use.

$$= \frac{.008}{.107} \approx .075$$

It is easy to see why few people will get this computation right without resorting to a calculator or at least paper and pencil! Assuming that the difference between estimating the probability of cancer as 80% and as 7.5% is one that means something, what needs to be noticed is just the *need* to employ a computational device (such as paper and pencil) to get the answer right, rather than relying upon an intuitive judgment that gets the answer wrong. Another option is to learn one of the following three alternative methods of computation, each of which is notably psychologically simpler than the one above.

## 8.2 Frequency Formats

Hoffrage and Gigerenzer (1998) advocate the use of “frequency formats” to make Bayesian reasoning more intuitive. Basically, this involves multiplying the probabilities in any given case by a sufficiently large number so that the entire problem can be worked in whole numbers rather than fractions. Thus, we can take the breast cancer numbers and multiply them by 1000. We also lay out the problem in a classification tree (cf. Breiman et al., 1984), as in Figure 2.

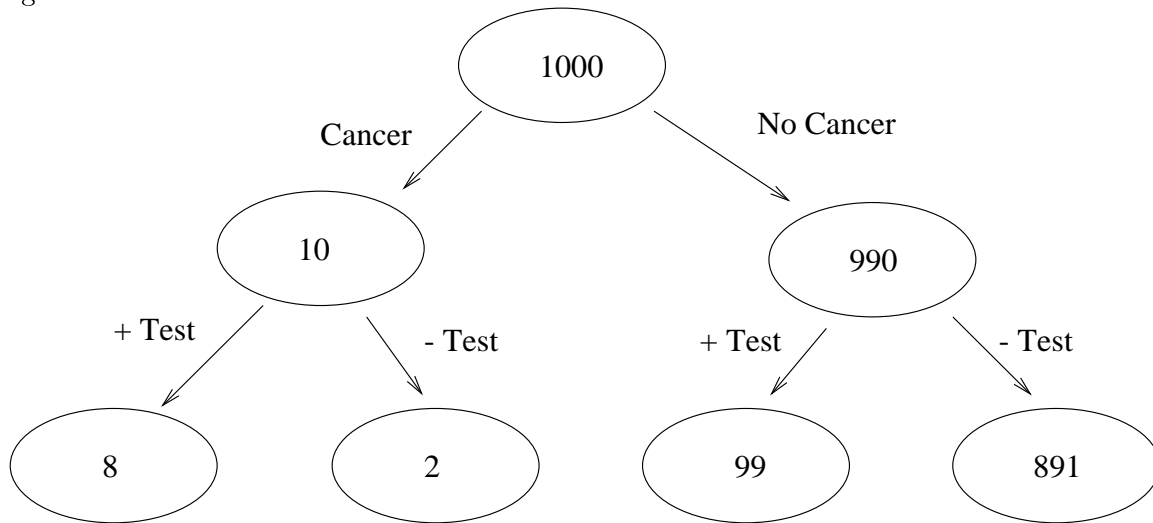


Figure 2: Classification tree for breast cancer

To construct the tree, we start with 1000 women. One percent are presumed to have cancer (the prior probability); that means the left branch yields 10 women with cancer and the right branch 990 without. Of the 10 women with cancer, 8 (80%) will test positive. Of the 990 without, 99 (10%) will test positive. Thus, confronted with a positive test and nothing else, we compute the probability of cancer as  $8/(8+99)$ . This is clearly easier to handle without computational devices than Bayes’ formula directly. Hoffrage and Gigerenzer (1998) report greater success in getting frequency formats into effective use than Bayes’ theorem, which is hardly surprising.

## 8.3 Odds-likelihood Bayes

Another way of simplifying Bayesian reasoning is to encourage people to think in terms of betting odds, rather than probabilities. In that case, the odds-likelihood form of Bayes’ theorem, which is far simpler to handle than the original form, can be employed exclusively. The breast cancer problem is then solved as follows.

The prior odds of cancer are:

$$O(\text{cancer}) = \frac{P(\text{cancer})}{P(\text{no cancer})} = \frac{1}{99}$$

The likelihood is:

$$\lambda = \frac{P(\text{pos}|\text{cancer})}{P(\text{pos}|\text{no cancer})} = \frac{.8}{.1} = 8$$

We apply odds-likelihood Bayes:

$$O(\text{cancer}|\text{pos}) = \lambda O(\text{cancer}) = 8 \times 1/99 = 8/99$$

In other words, we simply take the confirmatory power of the evidence, the likelihood ratio (Bayes factor) of  $\lambda = 8$ , and multiply that into the prior odds, yielding a posterior odds for cancer of 8:99. This is at least as simple as the use of frequency formats.

Furthermore, the odds-likelihood approach focuses attention on the two major factors in Bayesian reasoning better than any of the other approaches, since it simply involves multiplying them together — that is, multiplying the prior odds (equivalent to the prior probability) and the likelihood ratio (i.e., the confirmatory power of the evidence). Given these strengths, it would make good sense to try building a tutorial program for Bayesian reasoning around odds-likelihood reasoning. Nevertheless, I must report that in teaching first-year university students I have had more success with frequency formats than odds-likelihood Bayes; presumably, the latter should be reserved for more advanced students.

If in the end we are asked to produce a posterior probability, we will need to move from odds to probabilities through an additional conversion:

$$\begin{aligned} P(\text{cancer}|\text{pos}) &= \frac{O(\text{cancer}|\text{pos})}{1 + O(\text{cancer}|\text{pos})} \\ &= \frac{8/99}{107/99} \\ &= \frac{8}{107} \approx .075 \end{aligned}$$

## 8.4 Bayesian Networks

From the human point of view, far and away the simplest method of solving such probability problems is just to let a computer do them for us. Bayesian network technology has been developed by AI researchers and statisticians since the 1980s (e.g., Pearl, 1988; Neapolitan, 1990) which allows both exact and approximate Bayesian inference to be performed with essentially no burden on the human user. In recent years, this technology has been incorporated into PC programs with typical windows interfaces (e.g., Netica, at [WWW.NORSYS.COM](http://WWW.NORSYS.COM)). The breast cancer problem then is handled by: building a two-variable Bayesian net as in Figure 3; clicking on the test node and setting it positive; reading off the display a posterior probability of cancer at 0.075. This is as simple as it gets. The use of such networks in decision analysis and decision support is almost certainly going to become widespread.

One might wonder how such a tool is possible. After all, above I pointed out that Bayesian computations are NP-hard, that is to say, they grow exponentially complex and long as the problem size increases. There are two parts to the answer to this. First, a two-variable problem is computationally trivial (at least for the better and more capable non-humans); the computational complexity kicks in somewhere above ten variables (depending on the network structure and number of states per node). Second, methods of approximating exactly correct inference, including stochastic simulation over Bayesian networks, are available, and their improvement is an active area of research. These techniques make possible inference in complex problems that would otherwise be hopeless.

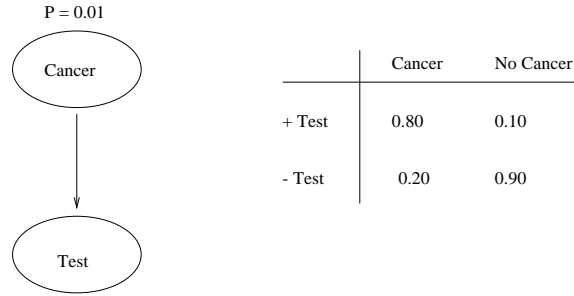


Figure 3: Bayesian network for breast cancer

## 9 Conclusion

Despite the fact that probability theory applies a formal calculus and, as such, comes no more naturally to people than do mathematical logic or differential and integral calculus, it should be clear that probability theory offers an unequaled guide to correct human inference. It can go a long way toward making sense not just of scientific inference and evaluation, but also of ordinary language arguments of a wide variety. Further, there is reason to believe that real progress can be made in developing pedagogical techniques for teaching probabilistic reasoning and its application to argumentation. Here I have only made a rudimentary start at applying Bayesian principles and methods to that task, but I believe it makes clear the opportunity for a more systematic application of Bayesian technique for informal reasoning. Perhaps, indeed, someone will even rehabilitate the standing of the sophists, removing the *ad hominem* invented by Plato so long ago.

## Acknowledgement

I would like to thank the London School of Economics and its Centre for the Philosophy of the Natural Sciences for the provision of facilities during early work on this paper. I thank Charles Twardy and Felicity Lawson for helpful comments.

## References

- Aristotle *On sophistical refutations*.
- Aristotle *Rhetoric*.
- Bacchus, F., H. Kyburg, and M. Thalos (1990). Against conditionalization, *Synthese*, 8,: 475-506.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman and Hall.
- Carnap, R. (1950). *Logical foundations of probability*. University of Chicago.
- Chase, V. M., Hertwig, R. and Gigerenzer, G. (1998) Visions of rationality. *Trends in Cognitive Sciences*, 2, 206-214.
- Cohen, L. J. (1981). Can human rationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 6, 248-249.
- Copi, I. and Cohen, C. (1990) *Introduction to logic*, tenth edition. Prentice Hall.
- Cover, and Thomas, (1991). *Elements of information theory*. New York: Wiley.
- de Finetti, B. (1937/80). Foresight: Its logical laws, its subjective sources. Translated by H. Kyburg in H. Kyburg and H. Smokler (Eds.) *Studies in subjective probability*, second edition (pp. 55-118). Huntington, NY: Krieger.
- Eddy, D.M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In Kahneman, Slovic and Tversky (1982), pp. 249-267.

- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases.” In W. Stroebe and M. Hewstone (Eds.) *European Review of Social Psychology*, vol 2 (pp. 83-115). New York: Wiley.
- Gigerenzer, G. (1996). The psychology of good judgment: Frequency formats and simple algorithms. *Journal of Medical Decision Making*, 16, 273-280.
- Gigerenzer, G. and Murray, D. J. (1987). *Cognition as intuitive statistics*. Lawrence Erlbaum.
- Gigerenzer, G. and Todd, P.M. (1999). *Simple heuristics that make us smart*. Oxford University.
- Glymour, C., Scheines, R., Spirtes, P. and K. Kelly (1987). *Discovering causal structure: Artificial intelligence, philosophy of science and statistical modeling*. Academic Press.
- Good, I.J. (1983). *Good thinking: The foundations of probability and its applications*. University of Minnesota.
- Hamblin, C. L. (1970). *Fallacies*. Newport News, VA: Vale Press.
- Harman, G. (1965). Inference to the best explanation. *Philosophical Review*, 74, 88-95.
- Hempel, C. (1965) *Aspects of scientific explanation*. New York: Free Press.
- Hoffrage, U., and Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, 73, 538-540.
- Holte, R. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11, 63-91.
- Howson, C. (2000). The logical basis of uncertainty. Presented at *Bayesianism 2000, King's College London, 11-12 May, 2000*.
- Jaynes, E. (1989). *Papers on probability, statistics and physics*, second edition, edited by R.D. Rosenkrantz. Kluwer.
- Joseph, H.W.B. (1906). *An introduction to logic*. Oxford: Clarendon.
- Kahneman, D., Slovic, P. and Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University.
- Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454.
- Korb, K.B. and Thompson, C. (1994). Primitive concept formation. *Second Australian and New Zealand Conference on Intelligent Information Systems* (pp. 362-366).
- Korb, K.B. and Wallace, C.S. (1997). In search of the philosopher’s stone: Remarks on Humphreys and Freeman’s critique of causal discovery. *British Journal for the Philosophy of Science*, 48, 543-553.
- Laplace, P. S. (1814/1951) *A philosophical essay on probabilities*, trans. F.W. Truscott and F.L. Emory. Dover.
- Lewis, D. (1980). A subjectivist’s guide to objective chance. In R. Jeffrey (Ed.) *Studies in inductive logic and probability* (Vol II, pp. 263-293). University of California.
- Lipton, P. (1991). *Inference to the best explanation*. Routledge.
- McKim, V. and Turner, S.P. (Eds.) (1997). *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences*. Notre Dame.
- Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace and Ward.
- Nisbett, R.E., Fong, G.F., Lehman, D.R. and Cheng, P.W. (1987). Teaching reasoning. *Science*, 238, 625-631.
- Polya, G. (1968) *Mathematics and plausible reasoning, volume 2*, second edition. Princeton University.
- Popper, K. (1934/59). *Logik der Forschung*, translated as *Logic of scientific discovery*. London: Hutchison.
- Ramsey, F. P. (1931). *The foundations of mathematics and other logical essays*, edited by R. B. Braithwaite. New York: Humanities Press.
- Reichenbach, H. (1949). *The theory of probability*, second edition. University of California.

- Reichenbach, H. (1956). *The direction of time*. University of California.
- Rips, L.J. (1975). Inductive judgments about natural categories. *Journal of Verbal Language and Verbal Behavior*, 14, 665-681.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M. and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Rottenberg, A.T. (1991). *Elements of argument: A text and reader*. Boston: St. Martin's Press.
- Teller, P. (1973). Conditionalization and Observation. *Synthese*, 26, 218-258.
- Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 4, 207-232.
- Tversky, A. and Kahneman, D. (1982). Evidential impact of base rates. In Kahneman, Slovic and Tversky (1982), pp. 153-160.
- van Gelder, T. (2000). The efficacy of undergraduate critical thinking courses: A survey in progress. At [HTTP://WWW.PHILOSOPHY.UNIMELB.EDU.AU/REASON/CRITICAL/](http://www.philosophy.unimelb.edu.au/reason/critical/)
- Verma, T., and Pearl, J. (1990). Equivalence and synthesis of causal models. *Proceedings of the Sixth Conference on Uncertainty in AI* (pp. 220-227). Morgan Kaufmann.
- Wallace, C.S. and Korb, K.B. (1999). Learning linear causal models by MML sampling. In A. Gammerman (Ed.) *Causal models and intelligent data management*. Springer Verlag.
- Walton, D. N. (1989). *Informal logic: A handbook for critical argumenation*. Cambridge University Press.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task, *Quarterly Journal of Experimental Psychology*, 12, 129-140.
- Wason, P. C. (1966). Reasoning. In B. M. Ross (Ed.), *New horizons in psychology*, I. Hammondsorth: Penguin.