The Combination of Evidence

Glenn Shafer
School of Business, University of Kansas, Lawrence, KS 66045

This article provides a historical and conceptual perspective on the contrast between the Bayesian and belief function approaches to the probabilistic combination of evidence. It emphasizes the simplest example of non-Bayesian belief-function combination of evidence, which was developed by Hooper in the 1680s.

I. INTRODUCTION

The theory of belief functions, also called the Dempster–Shafer theory, has attracted a great deal of attention in artificial intelligence during the past several years (see Gordon and Shortliffe[1] and Yager[2]). This theory is based on mathematical probability, but it uses mathematical probability in a more general way than the better known Bayesian theory does.

Unfortunately, many proponents of the Bayesian theory see their approach as the only legitimate way to use mathematical probability in the assessment and combination of evidence. They are therefore highly critical of the theory of belief functions, even though this theory is a generalization of the Bayesian theory. Some (e.g., Cheeseman[3]) go so far as to characterize the theory of belief functions as *ad hoc* and essentially nonprobabilistic.

This article attempts to give a broader historical and conceptual perspective on the relation between the belief-function and the Bayesian theories. On the historical side, we will see that the belief-function approach to the probabilistic combination of evidence is actually much older than the Bayesian approach; the basic non-Bayesian belief-function idea appeared in the 1680s in the work of George Hooper, whereas the Bayesian approach first appeared with Bayes's original paper in the 1760s and was well understood only after Laplace rediscovered it in the 1780s. On the conceptual side, we will see that non-Bayesian belief-function combination of evidence cannot be properly understood if it is seen in Bayesian terms.

We begin by studying the simple rules for the combination of evidence developed by Hooper and contrasting them with corresponding Bayesian rules later developed by Laplace. We then turn to the generalization of Hooper's rules developed by Lambert in the 1760s. After that we study and criticize recent Swedish attempts (Gärdenfors, Hansson, and Sahlin[4]) to interpret Hooper's and Lambert's rules in Bayesian terms. Finally, we explain how Dempster further

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. I, 155-179 (1986)
 CCC 0884-8173/86/030155-25$04.00

generalized Lambert's rule to a general rule for combining belief functions, and we extend our criticism of Bayesian interpretations of Hooper's and Lambert's rules to recent attempts to give Bayesian interpretations of Dempster's rule.

II. HOOPER

In 1699, George Hooper, who later became Bishop of Bath and Wells, published an article in the *Philosophical Transaction of the Royal Society*[5] entitled "A Calculation of the Credibility of Human Testimony."* In this article, Hooper formulated two rules relating the credibility of reports to the credibility of the reporters who make them.

These two rules are quite simple. The *rule for successive testimony* says that if a report has been relayed to us through a chain of n reporters, each having a degree of credibility p, then the credibility of the report is p^n. The rule for *concurrent testimony* says that if a report is concurrently attested to by n reporters, each with credibility p, then the credibility of the report is $1-(1-p)^n$. (Here $0 \leq p \leq 1$.) Thus the credibility of a report is weakened by transmission through a chain of reporters but strengthened by the concurrence of reporters.

Hooper's explanations of these rules were reminiscent of Pascal's arguments about equity in games of chance. But it is easy to motivate the rules in modern terms. We simply think of the performance of the reporters as random. There is a chance p that a given reporter will report faithfully and accurately. The performance of one reporter is independent of the performance of another. In the case of a chain of reporters, we can be confident of the report finally transmitted to us *only if* all the reporters in the chain have reported faithfully and accurately, and the chance of this is p^n. In the case of concurrent reporters, we can be confident of the report they concur in if at least one has reported faithfully and accurately, and the chance of this is $1-(1-p)^n$.

The applicability of Hooper's rules to actual testimony is, of course, very limited. In order to use the rule for successive testimony, we must know the length and nature of the chain of testimony, something we are unlikely to be certain about if we cannot fully credit our sources. And both rules depend on an unlikely independence in the motives and abilities of different reporters. The rules illustrate, nonetheless, important general aspects of the evaluation of evidence. The rule for successive testimony illustrates how the weight of an argument is diminished when it involves many uncertainties. And the rule for concurrent testimony illustrates how high probability results from the concurrence of independent arguments.

*Hooper had already published the main ideas of this article in 1689,[6] in a passage in his tract refuting the infallibility of the pope. The 1699 article was anonymous, and until recently its authorship was unknown to students of the history or probability, even though it was published in his collected works in 1757[7] and again in 1855. (In the 1855 edition,[8] the 1689 tract is on pp. 33–156 of Vol. 1, and the 1699 article follows on pp. 157–162. The relevant passages in the tract are on pp. 40–49.) Brown Grier[9] is responsible for calling Hooper's authorship to the attention of historians of probability.

Hooper's rules were repeated in many 18th century books and articles on probability.* We may assume that the authors of these works repeated the rules because of the general insights they offered, not because of any illusions about their literal applicability to problems of testimony.

A. Interpreting the Rule for Successive Testimony

Suppose a report comes to us through a chain of 12 witnesses, each with credibility 100/106. Then the credibility of the report will be $(100/106)^{12} \approx .5$. Hooper expressed this by saying that it will be "an equal Lay whether the report be true or no."

This simple betting interpretation is open to two related objections. First, there might be other evidence for or against the report that might affect our willingness to bet on it. Secondly, the report might be true in spite of a failure of integrity on the part of one or more of the reporters, so that our willingness to bet at even odds that at least one reporter failed in his duty should not imply an equal willingness to bet that the report is false.

These points are brought out more strongly if we lengthen the chain so as to make the credibility smaller. With a chain of 100 reporters, for example, we have a credibility of the final report of $(100/106)^{100} \approx .003$. This calculation should convince us that such a long chain of testimony is useless as evidence, but it may fail to persuade us to bet 997 to 3 against the truth of the report.

Contrary to Hooper's own assertion, then, the credibility p^n mentioned in Hooper's rule should not be thought of as fair odds for betting for or against the report. Instead it should be thought of as the probability given to the report by the testimony alone.

As I have already suggested, Hooper's rule for successive testimony can be thought of as a general rule for combining uncertainties in an argument. If an argument has n steps, and each has a chance p of being right, then the whole argument has a chance p^n of being right. (More generally, if the ith step has chance p_i of being right, then the whole argument has a chance $p_1 \ldots p_n$ of being right.) Here, perhaps, we should not call p^n the probability of the argument's conclusion. Instead, we should call it the probability provided the conclusion by the argument alone, or more simply, the probability of the argument.

B. Interpreting the Rule for Concurrent Testimony

Suppose two witnesses act independently and each witness has a 3/4 chance of testifying faithfully and accurately. Then there is a chance of

$$1 - \left(1 - \frac{3}{4}\right)\left(1 - \frac{3}{4}\right) = \frac{15}{16} = 0.9375$$

*For details concerning the influence of Hooper's rules, see Grier[9] and Shafer.[10] It should be noted that similar rules were independently discovered by James Bernoulli; see again Shafer.[10]

that at least one of the two will be faithful and accurate. If the two witnesses agree, then the proposition that at least one is faithful and accurate implies that their common report is true. Hooper's rule for concurrent testimony says that we may therefore interpret 15/16 as the probability given to this report.

Does this rule take full account of the agreement between the witnesses? Let us imagine that the experiment of having the two witnesses (or two similar witnesses) testify is repeated many times. We can say that about 15/16 of these times at least one of the two will be faithful and accurate. But perhaps we should consider only those times when the two witnesses agree. Will at least one of the two witnesses be faithful and accurate 15/16 of these times? Not necessarily.

Suppose, for example, that the experiments are arranged so that the two witnesses must always speak to the same question, each witness always knows the answer to the question, and he either tells the truth or lies. In this case, both witnesses will tell the truth $3/4 \cdot 3/4 = 9/16$ of the time, and both will lie $1/4 \cdot 1/4 = 1/16$ of the time; otherwise they will disagree. So they will be telling the truth only

$$\frac{\frac{1}{16}}{\frac{9}{16}+\frac{1}{16}} = \frac{9}{10}$$

of the times that they agree. And 9/10 is less than 15/16.

It seems clear that if the experiment really is arranged as described in the preceding paragraph, and if we know this, then we should not use the probability of 15/16 given by Hooper's rule. We should use the probability 9/10 instead, because it is the result of the analysis that takes the details of the situation more fully into account.

When, then, is Hooper's rule appropriate? Suppose our two witnesses agree but we do not know what, if any, rules they may be following with regard to what questions they will speak to or what they will say if they are not being faithful and accurate. Is Hooper's rule appropriate in this case? Unbeknown to us, the witnesses may be following the rules described above, so that their report is true only 9/10 of the time when they agree. Is it legitimate for us, in this situation, to say that the witnesses give the report a probability of 15/16? It might not be wise, certainly, for us to offer to bet on the report at odds of 15 to 1. Someone who knew the rules the witnesses were following could make money from us in the long run by accepting such offers.

Of course, it is always unwise to bet with someone who knows more than we do.

Moreover, there may be no clear rules that the witnesses are following. Often any rules about what the witnesses would do if they were not to speak faithfully and accurately to the question at hand will be more hypothetical than factual. Talk about such rules is talk about might-have-beens, or "counterfactual" talk. As van Fraassen[11] and other philosophers have stressed, such talk about might-have-beens is always controvertible, for it always involves the choice of one out of many possible but nonactual worlds.

Our idea of numerical probability derives from the picture of games of chance. Games of chance are played out step-by-step, and they always follow a "protocol," a set of rules specifying what may happen at each step (see Shafer[12]). The players know the protocol and keep abreast of what is happening in the game, and so at each step they have new probabilities for what will happen next. The existence of the protocol means that these new probabilities take the details of the situation fully into account, and hence these probabilities can be given a strong betting interpretation: a player who offers to bet at the odds set by these probabilities can be certain of breaking even in the long run no matter what strategy another player uses in deciding what bets to accept.

In practical problems talk about probabilities usually involves an imperfect comparison to the picture of games of chance. We construct "probability arguments" by showing in detail how our situation is comparable to a situation in a certain game of chance (see Shafer and Tversky[13]). Usually one reason for the imperfection of the comparison is the lack of a fully persuasive protocol for the real-world problem. We cannot take a particular aspect of the situation fully into account because we cannot give rules for what might have happened differently. In the case of the two witnesses, we may not be able to take their agreement fully into account because we cannot fully specify the conditions under which they might accidentally or mendaciously agree.

The imperfection of the comparison of real-world problems to the picture of games of chance means that there are often several probability arguments that compete for our attention. Often, as in the case of our example of the two witnesses, one analysis may seem to take the details of the situation more fully into account and yet be less convincing as an argument than a simpler analysis that makes fewer assumptions.

Hooper's rule for concurrent testimony is based on a very simple analysis and often provides a convincing probability argument. Like the rule for successive testimony, it can be applied not just to testimony but to arguments in general. If there are n independent arguments for a conclusion, and each has a chance p of being right, then all the arguments together constitute an argument that has a chance $1 - (1-p)^n$ of being right. So we call $1 - (1-p)^n$ the probability provided the conclusion by these arguments alone.

III. LAMBERT

Hooper's rule for concurring testimony is concerned with the case where two witnesses agree. Suppose they disagree. How is the credibility of one witness affected by the contrary testimony of another?

The Alsatian polymath Johann Heinrich Lambert addressed this question in his philosophical treatise *Neues Organon*,[14] published in 1764 (see Shafer[10]). Lambert gave a general rule for combining the credibilities of two witnesses, applicable whenever the testimony of both is directed to exactly the same question.

Lambert's rule is based on a fuller probabilistic model for testimony than Hooper's rule is. As we have seen, Hooper's rule can be based on a picture where a witness has a certain chance of being reliable. There is a chance p, say, that the witness will be faithful and accurate, and we may be confident of what he says; otherwise, his testimony is of no import. In Lambert's fuller picture, there is a chance p that the witness will be faithful and accurate, a chance q that he will be mendacious, and a chance $1-p-q$ that he will simply be careless. In the first case we may be confident of what he says; in the second case we may be confident of the denial of what he says; in the third case his testimony is of no import.

Suppose two witnesses act independently, the first with chances p_1 and q_1, the second with chances p_2 and q_2. Lambert's rule says that when both witnesses agree their testimony lends credibility

$$\frac{1 - (1 - p_1)(1 - p_2) - k}{1 - k} \tag{1}$$

to the report they are testifying to, and credibility

$$\frac{1 - (1 - q_1)(1 - q_2) - k}{1 - k} \tag{2}$$

to its denial, where

$$k = p_1q_2 + q_1p_2.$$

In the case where the second witness flatly contradicts the first, the resulting credibility of the report of the first is

$$\frac{1 - (1 - p_1)(1 - q_2) - k}{1 - k}, \tag{3}$$

while the credibility of the report of the second is

$$\frac{1 - (1 - q_1)(1 - p_2) - k}{1 - k}, \tag{4}$$

where

$$k = p_1p_2 + q_2q_2.$$

Expressions (3) and (4) are obtained from (1) and (2) by interchanging p_2 and q_2.

Expression (1) can be thought of as a generalization of Hooper's rule for concurrent testimony. Indeed, if we set $q_1 = q_2 = 0$, then Expression (1) reduces to $1 - (1-p_1)(1-p_2)$. It is also of interest to note the special case where $q_1 = q_2 = 0$ and the witnesses contradict each other. In this case Expression (3) reduces to

$$\frac{p_1 - p_1p_2}{1 - p_1p_2} \tag{5}$$

and Expression (4) reduces to

$$\frac{p_2 - p_1p_2}{1 - p_1p_2}.$$

Expressions (1)−(4) are derived as follows. Since the two witnesses act independently, there is a chance p_1p_2 that both testify faithfully and accurately, a chance p_1q_2 that the first testifies faithfully and the second mendaciously, etc. Altogether there are nine possibilities, with probabilities adding to one, as shown in Table I. The nine probabilities are shown schematically in Figure 1, where they represent the areas of nine rectangles that together form a unit square. If we find that the two witnesses agree, then we know that it is impossible for one to have been truthful and the other mendacious. This means we must eliminate possibilities 2 and 4 in Table II. Possibilities 1, 3, and 7 imply that the witness's testimony is true, and the total conditional probability of these possibilities,given the elimination of possibilities 2 and 4, is

$$\frac{p_1p_2 + p_1(1 - p_2 - q_2) + (1 - p_1 - q_1)\, p_2}{1 - p_1q_2 - q_1p_2},$$

which is equal to Expression (1). Similarly, possibilities 5, 6, and 8 imply that the testimony is false, and the total conditional probability of these possibilities is

$$\frac{q_1p_2 + q_1(1 - p_2 - q_2) + (1 - p_1 - q_1)\, q_2}{1 - p_1q_2 - q_1p_2},$$

which is equal to Expression (2). Expressions (3) and (4) follow similarly in the case where the witnesses disagree; in the case possibilities 1 and 5 are eliminated.

Table I.

Probability	Possibility
1. p_1q_2	Both truthful.
2. p_1q_2	First truthful, second mendacious.
3. $p_1(1-p_2-q_2)$	First truthful, second careless.
4. q_1p_2	First mendacious, second truthful.
5. q_1q_2	Both mendacious.
6. $q_1(1-p_2-q_2)$	First mendacious, second careless.
7. $(1-p_1-q_1)p_2$	First careless, second truthful.
8. $(1-p_1-q_1)q_2$	First careless, second mendacious.
9. $(1-p_1-q_1)(1-p_2-q_2)$	Both careless.

$(1-p_2-q_2)$	$p_1(1-p_2-q_2)$	$q_1(1-p_2-q_2)$	$(1-p_1-q_1)(1-p_2-q_2)$
q_2	p_1q_2	q_1q_2	$(1-p_1-q_2)q_2$
p_2	p_1p_2	q_1p_2	$(1-p_1-q_1)p_2$
	p_1	q_1	$(1-p_1-q_1)$

Figure 1.

The probabilistic model for testimony on which Lambert's rule is based can be more complete than the model for Hooper's rule, but it is still, in general, incomplete. In particular, the witness's behavior is incompletely modeled as long as $p + q < 1$. In such a case, if the witness is really following a fuller model, then someone who knows that model will be able to make money by betting with us.

An Example. My friend Ilse mentions to me that she will be busy late Sunday afternoon; she is going to church at 5:00 p.m. I wonder whether Ilse is Catholic. It is my impression that midwestern Protestant churches seldom have services in the afternoon, whereas Catholic churches schedule masses throughout the day. Now that I think of it, Ilse is from Austria, and Austria is 90% Catholic and only 10% Protestant. Her husband is from Los Angeles, but that does not suggest any particular religious affiliation.

Here is a simple analysis by Lambert's rule. Ilse's afternoon church services and her Austrian origin are two independent arguments for her being Catholic. I feel 80% certain that no Protestant churches in Lawrence have an afternoon service. So the afternoon service is like testimony that has an 80% chance of meaning Ilse is Catholic and a 20% chance of meaning nothing: $p_1 = .8$, $q_1 = 0$. Her Austrian origin is like testimony that has a 90% chance of meaning she is Catholic and a 10% chance of meaning she is Protestant: $p_2 = .9$, $q_1 = .1$. Substituting these values in Expression (1), we find that our evidence provides a probability of 97.8% that Ilse is Catholic.

IV. LAPLACE

Though Hooper's rules were popular in the eighteenth century, they did not survive into the nineteenth century. Instead, they were replaced by rules more in accord with the understanding of probability developed in the late eighteenth century by Bayes, Condorcet, and especially Laplace.

This "Bayesian" approach, as we now call it, seems to have first been applied to problems of testimony by Condorcet in the 1780s (Todhunter,[15] p. 400). The best early account is the one given by Laplace in Chapter XI of the second edition

Table II.

A. Chance .8 of meaning "1 or 2"; chance .2 of meaning nothing.
B. Chance .9 of meaning "1 or 4"; chance .1 of meaning "2 or 3."
C. Chance .98 of meaning "1 or 3"; chance .02 of meaning nothing.

of his *Théorie analytique des probabilitiés*,[16] published in 1814 (*Oeuvres de Laplace*, Vol. VII, pp. 455–470[17]).

The Bayesian approach to testimony involves consideration of prior probabilities. In addition to considering the credibility of the witness, we must also consider the prior or intrinsic probability of what the witness says. The Bayesian approach also requires a fuller probabilistic model of the witness's behavior than Hooper's or Lambert's approach does. In explaining Hooper's rule, we said that a witness's credibility p was the chance that the witness would report faithfully and accurately. We left open the possibility that the witness might also say something true by accident or as a result of a careless guess, but we said nothing about the chance of this. The Bayesian approach, in contrast, requires consideration of the total chance that the witness would make the report he did make, both under the hypothesis that it is true and under the hypothesis that it is false.

In order to see what is involved when we think in terms of the total chance that a witness would make the report he did make, let us consider the case of a single witness. Let us make the simplifying assumption that the witness will be accurate if and only if he does not deliberately lie, and let p denote his "veracity" —the probability that he will not lie. Let A denote the fact he reports, and let B denote the event that he reports it. Shifting to twentieth century nomenclature and notation, we now use "Bayes's theorem" to calculate "the conditional probability $P[A|B]$":

$$P[A|B] = \frac{P[A]P[B|A]}{P[A]P[B|A] + P[\bar{A}]P[B|\bar{A}]} \tag{6}$$

Here $P[A]$ is the prior or intrinsic probability of A—its probability based on evidence other than the witness's report. (The symbol $\bar{A}$ means "not A"; thus $P[\bar{A}] = 1-P[A]$.) How do we assign values to the conditional probabilities $P[B|A]$ and $P[B|\bar{A}]$?

The conditional probability $P[B|A]$ is the probability the witness will report A if it is true. In order to assess this, it is not enough to know how likely the witness is to be honest. We must also know how likely he is to say anything about A at all. Similarly, in order to assess $P[B|\bar{A}]$, we must know not only how likely he is to lie but also how likely he is to choose this particular lie.

The simplest way of handling this is to suppose that it is arranged that the witness will find out about A and will report either that A is true or that it is not—nothing less and nothing more. In this case the witness must either report A or lie if A is true, and hence $P[B|A]$ must equal his veracity p. Similarly, he must either report A or tell the truth if A is not true, and hence $P[B|\bar{A}]$ must equal his probability of lying, $1-p$. Thus

$$P[A|B] = \frac{P[A]p}{P[A]p + P[\bar{A}](1-p)} \,. \tag{7}$$

If we further assume that $P[A] = 1/2$, then this reduces to $P[A|B] = p$; the final probability of the report is simply the witness's veracity.

Laplace considered the following generalization of the preceding model. One begins with n mutually exclusive possibilities—say n different balls that might be drawn from an urn. The witness is supposed to find out which of these is true (he draws a ball) and then to report one of them as true. It is assumed that if he decides to lie then he is equally likely to choose any of the $(n-1)$ false possibilities to report. Thus $P[B|A]$, the probability he reports A if it is the true possibility, is again p. But $P[B|\bar{A}]$, the probability he chooses A to report if it is not the true possibility, is only $(1-p)/(n-1)$. If we take the intrinsic probability $P[A]$ to be $1/n$, then Expression (6) becomes

$$P[A|B] = \frac{\frac{1}{n}\cdot p}{\frac{1}{n}\cdot p + \frac{n-1}{n}\cdot\frac{1-p}{n-1}} = p.$$

Again the final probability is the witness's veracity (*Oeuvres de Laplace*, Vol. VII, pp. 457–458[17]).

Yet other models are possible, and perhaps more realistic. The important point is that the Bayesian analysis requires we have some model. It is not enough to assess the witness's honesty. We must also assess how likely he is to choose this particular lie if he is lying.

Now let us review Laplace's treatment of chains of testimony. Consider first the case of two witnesses. We again assume that each witness has veracity p and is accurate unless he deliberately lies. And we again assume that there is a protocol that specifies an event A and requires each witness to say simply whether or not A is true—nothing less and nothing more. The first witness reports to the second; the second reports to us. Let B denote the event that the second witness tells us A. If A is true, then B will happen if both witnesses tell the truth and also if both lie—for the first lies by saying "not A" and the second then lies by saying "A": Thus $P[B|A] = p^2 + (1 - p)^2$. Similarly, $P[B|\bar{A}] = 2p(1 - p)$. If we again take the prior probability $P[A]$ to be 1/2, then Expression (6) yields

$$P[A|B] = \frac{\frac{1}{2}(p^2 + (1 - p)^2)}{\frac{1}{2}(p^2 + (1 - p)^2) + \frac{1}{2}2p(1 - p)} = p^2 + (1 - p)^2.$$

In the case of a chain of three witnesses, similar assumptions yield the probability $p^3 + 3p(1-p)^2$ for the final report; in the case of n witnesses, a probability of

$$\sum_{0 \leq k < n/2} \binom{n}{2k} p^{n-2k}(1-p)^{2k} = \frac{1}{2} + \frac{1}{2}(2p-1)^n. \tag{8}$$

This is Laplace's rule for successive testimony.*

Finally, let us review Laplace's rule for concurrent testimony. We now suppose witnesses follow a protocol which requires that each find out whether A is true and then report to us either that it is or is not—nothing more and nothing less. Let B be the event that all n witnesses report A is true. The witnesses act independently, and each has veracity p. If A is true, then B happens only when all the witnesses tell the truth; if A is false, B happens only when they all lie. Thus Expression (6) yields

$$P[A|B] = \frac{P[A]p^n}{P[A]p^n + P[\bar{A}](1 - p)^n} . \tag{9}$$

If we again set $P[A] = 1/2$, then this reduces to $p^n/(p^n+(1-p)^n)$ (*Oeuvres de Laplace*, Vol. VII, p. 464[17]).

A. Bayesian and Non-Bayesian Arguments

The name "Bayesian" has become current only during the last few decades, and there is room for debate about its meaning. But everyone would agree that Laplace's arguments are Bayesian, while Hooper's and Lambert's are not. What distinguishes the two?

We cannot answer that the Bayesian arguments are distinguished by their subjectivity, for the "credibilities" used by Hooper's rules are as subjective as the numbers we put into the Bayesian formulas. And it would be superficial to answer that the Bayesian arguments are distinguished by their use of Bayes's theorem or by their use of prior probabilities. Bayes's theorem is an inessential tool for exposition, and the use of prior probabilities is only one aspect of the Bayesian demand for a full probability model.

This demand for a full probability model is the fundamental distinguishing feature of the Bayesian approach. The fundamental advantage of the Bayesian approach is that by using a full probability model we can take all the evidence into account. The fundamental disadvantage is that we sometimes lack evidence for some of the judgments that are called for. In the case of testimony, for example, the Bayesian analysis demands a judgment of the likelihood that the witness will choose to lie in a particular way if he lies. If there is something to be said about this likelihood, then it is an advantage that the Bayesian analysis asks about it. But if there is nothing to be said about it—if our evidence does not tell us anything about the value or even the meaningfulness of such a likelihood, then the demand that we supply it is a disadvantage.

Proponents of the Bayesian approach sometimes suggest that it is illegitimate to ignore any likelihoods or prior probabilities. We cannot, for example, ignore the likelihood that the witness will choose the particular lie, for this likelihood is

*The identity (8) may be proven by mathematical induction. Laplace did not, however, derive the expression $1/2 = 1/2\,(2p - 1)^n$ from a binomial sum. Instead, he solved a difference equation. See pp. 466–467 of Vol. VII of the *Oeuvres de Laplace*[17].

part of what is going on. But this is fundamentally wrong. Its plausibility depends on the pretense that the facts in the problem we are studying really are governed by probabilities as in a game of chance, so that the Bayesian analysis merely mirrors an objective reality. Actually, the picture of the witness choosing a particular lie according to given probabilities may be very far from what is going on.

The same pretense that every probability that can be named has an objective reality also tempts us to give Bayesian explanations of non-Bayesian arguments. Consider, for example, the special case of Lambert's rule that is obtained when we set $p_1 = p_2 = p$ and $q_1 = q_2 = 1-p$ in Expression (1). The substitution yields $p^2/(p^2 + (1-p)^2)$, which agrees with Laplace's rule (9) with $P[A] = 1/2$. We are tempted to say that Lambert's argument implicitly takes the prior probability of A to be 1/2. But this formulation overlooks the fundamental difference in logic between Lambert's and Laplace's arguments. Only Laplace's argument uses a model in which the truth or falsehood of A is determined by chance.

In several recent papers, I have suggested that probability arguments, whether Bayesian or non-Bayesian, should be thought of as thought experiments, in which actual situations are compared to pictures of chance (see Shafer,[18] and Shafer and Tversky[13]). Such a thought experiment is convincing as an argument to the extent that the details of the picture capture the significant and assessable items of evidence in the actual situation. In this context, Bayesian arguments are characterized by the fact that they attempt to fit the whole reality being investigated to the picture of chance. Non-Bayesian probability arguments attempt more modest comparisons.

B. Probability after Laplace

Laplace's contribution to probability was monumental, but it was not monolithic. It had both Bayesian and non-Bayesian elements. On the philosophical side, Laplace developed what we now call a subjective, Bayesian interpretation of probability; he stressed that probability is a measure of our knowledge, and he introduced what we now call "Bayes's theorem" as a fundamental, axiomatic rule for calculating the probabilities of causes. But on the practical side he developed both Bayesian and non-Bayesian methods in the theory of errors. Apparently he did not distinguish the two as we do now. He did not see the inconsistency between his fundamental Bayesianism and statistical methods that seem non-Bayesian to us (see Stigler[19]).

Laplace's nineteenth century successors did come to distinguish his Bayesian method (the "inverse method," as they called it) from the other methods in the theory of errors, and to recognize that the philosophical justification of these other methods required a different foundational approach to probability. They found this different approach in the "frequency" interpretation of probability. Thus was born the great dichotomy of philosophical viewpoint that has dominated probability for the past century and a half: on the one hand, Laplace's subjective Bayesianism; on the other, the "objective" frequentist philosophy that underlies the modern vocabulary of statistical tests and confidence methods.

The important point is that the nineteenth century re-emergence of non-Bayesian probability did not challenge Laplace's incorporation of Bayesian ideas into the subjective conception of probability. There was no revival of Hooper's rules and related eighteenth century ideas, which were non-Bayesian but clearly could have only subjective significance. Laplace made subjectivism Bayesian, and so it has remained.

The frequentist philosophy proved well-suited to support the development, in the early and mid-twentieth century, of non-Bayesian ideas in mathematical statistics. But it did not provide a basis for the extension of these ideas to those disciplines, such as philosophy and jurisprudence, that are concerned with probability judgment rather than with frequencies in repeatable experiments. Twentieth century scholars in these disciplines have, for the most part, taken the Bayesian picture for granted, even at times when this picture was an object of scorn among most statisticians.

Today, as the still young and growing discipline of statistics seeks to expand its breadth of application, statisticians have widely recognized the inadequacy of the frequentist philosophy and the pervasiveness of the practical need for subjective probability judgment. But this has not broken down the objective frequentist vs. subjective Bayesian dichotomy. It has simply resulted in the conversion of many statisticians from frequentism to Bayesian subjectivism.

I believe we need to escape from the frequentist vs. Bayesian dichotomy. Both sides of this dichotomy combine important truths with unreasonable dogmas. The Bayesian side is right to insist that probability judgment is subjective, but wrong to insist that subjective probability judgments are meaningful only within full probability models. The frequentist side is right to stress that probability judgments vary in the quality of their supporting evidence and that probability arguments depend on the strength of such evidence, but wrong to pretend to limit probability judgment to cases where this evidence takes the strong form of observed frequencies.

In order to break out of this sterile dichotomy, we need to revive the idea of non-Bayesian subjective probability judgment.

V. EKELÖF

In the early 1960s, Per Olof Ekelöf, a Swedish professor of law, began to study how mathematical probability could be used in jurisprudence. Dissatisfied with the relevance of what he could learn about probability from philosophers and statisticians, he undertook to formulate his own ideas, and soon he had invented three rules for the combination of probabilities—a rule for chains of evidence, a rule for concurring evidence, and a rule for conflicting evidence.* The first two of these rules are, as it turns out, formally identical to Hooper's rules for successive

*Ekelöf first presented his rules in 1963, in the first edition of Volume IV of *Rättegang*,[20] his Swedish legal textbook. Ekelöf (1964)[21] is an early presentation in English, and Ekelöf (1981)[22] is a more mature presentation in German. Ekelöf (1983)[23] includes an interesting account of the evolution of Ekelöf's ideas.

and concurrent testimony. The third rule, the rule for conflicting evidence, is related to a special case of Lambert's rule*

In addition to reviving Hooper's and Lambert's rules, Ekelöf also studied practical aspects of the legal application of these rules. In particular, he studied the role of auxiliary facts and circumstances that affect the weight of evidence. Auxiliary facts are facts that do not bear directly on what one is trying to prove but strengthen or weaken arguments that do bear on it. Evidence about a witness's eyesight is an example. Ekelöf stressed that auxiliary facts are often necessary to support a probability argument and that they cannot themselves be subjected to formulas but must instead be weighed intuitively.

Here is an example, from Ekelöf,[21] that illustrates Ekelöf's rule for chains of evidence. Suppose there are two uncertainties involved in a witness's testimony. It is uncertain whether he is accurately reporting what he thought he saw. And it is uncertain whether he really did see what he thought he saw. Suppose there is a chance 3/4 that he reports his observation accurately and a chance 3/4 that he observes accurately. Ekelöf suggests that the argument from his statement to his observation and hence to the truth of what he says then has overall credibility $(3/4)^2 = 9/16$.

Here is an example, again from Ekelöf,[21] that illustrates Ekelöf's rule for concurring evidence. Consider two items of evidence that each give credibility 3/4 to the conclusion that a car was traveling more than 60 miles per hour before an accident: the length of its skid marks and the observation of a witness. Ekelöf suggests that both together give the conclusion credibility $1-(1-3/4)^2 = 15/16$.

As these examples show, Ekelöf's rules, while formally identical to Hooper's rules for successive and concurring testimony, are meant to apply to more general kinds of evidence. Ekelöf's Swedish colleagues have suggested that these rules are concerned in general with "evidentiary mechanisms," and with the event that such a mechanism is working correctly. A witness is an evidentiary mechanism that is working correctly if he is testifying faithfully and accurately. Skid marks are an evidentiary mechanism that is working correctly if the causal factors that affect the relation between speed and the length of the marks are within normal limits.

A. Halldén and Edman's Bayesian Interpretation

In the early 1970s, the Swedish philosophers Sören Halldén and Martin Edman formulated assumptions under which Bayesian probabilities can satisfy relations suggested by Ekelöf's (or Hooper's) rules for concurring and conflicting evidence. In this section I shall review and criticize Halldén and Edman's Bayesian formulation and, in particular, their approach to the rule for concurring evidence.†

*Ekelöf has given several versions of his rule for conflicting evidence. One of these versions is the special case of Lambert's rule given by Expression (5). Stenlund has given a Bayesian treatment of this version in the same spirit as Halldén and Edman's treatment of the rule for concurring evidence. See Stening (p. 112)[24] and Goldsmith (p. 216).[25]

†Nils-Eric Sahlin, in private correspondence, has pointed out to me that Halldén and Edman do not consider themselves Bayesians. Yet they seem to take for granted the full probability models that characterize Bayesian arguments.

Let A denote the event that an evidentiary mechanism works correctly, let H denote a hypothesis, and let E denote the event that the evidentiary mechanism indicates H to be true. From a Bayesian point of view, we might say that Ekelöf's (or Hooper's) rules concentrate on $P[A]$, the probability that the mechanism works correctly. They do not consider $P[H|E]$, the probability of what the mechanism indicates, or even $P[A|E]$, the probability the mechanism worked correctly taking into account what it indicated and how likely what it indicated is on other grounds.

Halldén and Edman were intrigued by the idea that we shall consider the event A rather than the event H. Perhaps, they suggested, the probability of A is of greater relevance to the law than the probability of H. In a sense, A is the event that the evidence has proven H. Since courts are allowed to draw conclusions only if they prove them, it is the probability of A that indicates whether the court is entitled to conclude H, not the probability of H. Halldén and Edman took for granted, however, the Bayesian assumption of a full probability model, and so they considered it appropriate and possible to evaluate the probability of A on all the evidence; they were interested in $P[A|E]$ rather than $P[A]$.

Consider how these ideas apply to concurring evidence. If A_1 is the event that one witness reports faithfully and accurately, A_2 is the event that a second reports faithfully and accurately, and the two are independent, then Hooper's rule uses the formula

$$P[A_1 \vee A_2] = 1 - (1-P[A_1])(1-P[A_2]), \tag{10}$$

where "$A_1 \vee A_2$" means "A_1 or A_2." How close can we come to Expression (10) if the probabilities for the A_i are based on all the evidence? Halldén and Edman suggested we consider the inequality

$$P[A_1 \vee A_2|E_1E_2] \geqslant 1 - (1-P[A_1|E_1])(1-P[A_2|E_2]), \tag{11}$$

where E_i is the event that the ith witness reports H, and "E_1E_2" means "E_1 and E_2." When will Expression (11) hold?

Halldén,[26] who first posed this question, proved the following theorem.

Theorem 1. If $P[E_1E_2\bar{A}_1]$ is nonzero, so that probabilities conditional on $E_1E_2\bar{A}_1$ E_1E_2, E_1, and E_2 exist, and if the relations

$$P[A_2|E_2E_1\bar{A}_1] = P[A_2|E_2] \tag{12}$$

and

$$P[A_1|E_1E_2] \geqslant P[A_1|E_1] \tag{13}$$

hold, then Expression (11) holds. (*Proof:* We can write

$$\begin{aligned} P[A_1 \vee A_2|E_1E_2] &= P[A_1|E_1E_2] + P[\bar{A}_1A_2|E_1E_2] \\ &= P[A_1|E_1E_2] + P[\bar{A}_1|E_1E_2]\, P[A_2|E_1E_2\bar{A}_1] \\ &= 1 - (1-P[A_1|E_1E_2])\,(1-P[A_2|E_1E_2\bar{A}_1]). \end{aligned}$$

If we apply Expressions (12) and (13) to the last expression, we obtain (11). Q.E.D.)

Halldén suggested that conditions (12) and (13) are a reasonable formulation of the idea that the evidentiary mechanisms represented by A_1 and A_2 are independent.

It is evident from the proof of Halldén's theorem that the theorem will still hold if Expression (12) is weakened to the inequality

$$P[A_2|E_1E_2\bar{A}_1] \geqslant P[A_2|E_2]. \tag{14}$$

It is not clear, however, just when we can expect Expressions (12), (13), or (14) to hold.

Recognizing the need for further analysis of Halldén's conditions (12) and (13), Edman[27] formulated and proved the following theorem.

Theorem 2. Suppose again that $P[E_2E_2\bar{A}_1] > 0$, so that the conditional probabilities in Expressions (12) and (13) exist. Then Expressions (12) and (13) are implied by the following set of conditions.

$\mathscr{E}1$) If $X_1 \varepsilon \{A_1, \bar{A}_1E_1\}$ and $X_2 \varepsilon \{A_2, \bar{A}_2E_2\}$, then H, X_1, and X_2 are mutually independent.

$\mathscr{E}2$) $HA_1 = A_1E_1$ and $HA_2 = A_2E_2$.

If $P[H] > 0$, $P[A_1] > 0$, and $P[A_2] > 0$, then the inequality in Expression (13) is strict. (*Proof:* To derive Expression (12), we write

$$\begin{aligned}
P[A_2|E_2E_1\bar{A}_1] &= \frac{P[A_2E_2E_1\bar{A}_1]}{P[E_2E_1\bar{A}_1]} = \frac{P[A_2E_2E_1\bar{A}_1]}{P[A_2E_2E_1\bar{A}_1] + P[\bar{A}_2E_2E_1\bar{A}_1]} \\
&= \frac{P[HA_2E_1\bar{A}_1]}{P[HA_2E_1\bar{A}_1]+P[\bar{A}_2E_2E_1\bar{A}_1]} = \frac{P[HA_2]P[E_1\bar{A}_1]}{P[HA_2]P[E_1\bar{A}_1]+P[\bar{A}_2E_2]P[E_1\bar{A}_1]} \\
&= \frac{P[HA_2]}{P[HA_2] + P[\bar{A}_2E_2]} = \frac{P[A_2E_2]}{P[A_2E_2] + P[\bar{A}_2E_2]} = P[A_2|E_2].
\end{aligned}$$

To derive Expression (13), we write

$$\begin{aligned}
P[A_1|E_1E_2] &= \frac{P[A_1E_1E_2]}{P[E_1E_2]} = \frac{P[A_1E_1E_2]}{P[HE_1E_2] + P[\bar{H}E_1E_2]} \\
&= \frac{P[HA_1(A_2\text{v}\bar{A}_2E_2)]}{P[H(A_1\text{v}\bar{A}_1E_1)(A_2\text{v}\bar{A}_2E_2)] + P[\bar{H}\bar{A}_1E_1\bar{A}_2E_2]} \\
&= \frac{P[HA_1]P[A_2\text{v}\bar{A}_2E_2]}{P[H(A_1\text{v}\bar{A}_1E_1)]P[A_2\text{v}\bar{A}_2E_2] + P[\bar{H}\bar{A}_1E_1[P[\bar{A}_2E_2]} \qquad (15) \\
&\geqslant \frac{P[HA_1]}{P[H(A_1\text{v}\bar{A}_1E_1)] + P[\bar{H}\bar{A}_1E_1]} = \frac{P[HA_1]}{P[HA_1] + [\bar{A}_1E_2]}
\end{aligned}$$

$$= \frac{P[A_1E_1]}{P[A_1E_1] + P[\bar{A}_1E_1]} = P[A_1|E_1].$$

Since $P[A_2 v \bar{A}_2 E_2] = P[A_2] + P[\bar{A}_2 E_2]$, the inequality is strict unless $P[A_2] = 0$ or the numerator of Expression (15) is zero, and this is possible only if $P[H] = 0$ or $P[A_1] = 0$. Q.E.D.)*

Condition $\mathscr{C}1$ is Edman's way of making precise the idea of "independent evidentiary mechanisms." Condition $\mathscr{C}2$ says that if the mechanism is working correctly, then it will report H if and only if H is true.

But is condition $\mathscr{C}1$ plausible? Ekelöf's rule for concurring evidence seems intuitively reasonable whenever the event that the first evidentiary mechanism works correctly is independent of the event that the second works correctly. Condition $\mathscr{C}1$ goes far beyond this. It requires that when the mechanisms are not working they behave independently of each other and of the truth or falsehood of H. Indeed, in the presence of $\mathscr{C}2$, $\mathscr{C}1$ is equivalent to the four following conditions:

- $\mathscr{F}1$) H, A_1, and A_2 are mutually independent.
- $\mathscr{F}2$) The joint distribution for H, A_1, and E_1 is independent of A_2, and the joint distribution for H, A_2, and E_2 is independent of A_1.
- $\mathscr{F}3$) E_1 and E_2 are independent given either $H\bar{A}_1\bar{A}_2$ or $\bar{H}\bar{A}_1\bar{A}_2$.
- $\mathscr{F}4$) E_1 and H are independent given $\bar{A}_1$, and E_2 and H are independent given $\bar{A}_2$.

(The statement "A and B are independent given C" should be understood to mean that $P[ABC]P[C] = P[AC]P[BC]$; this is equivalent to $P[AB|C] = P[A|C]P[B|C]$ if $P[C] > 0$. Similarly, the statement "the joint distribution of A, B, and C is independent of D" should be understood to mean that $P(ED) = P(E)P(D)$ for any event E that is determined by the determination of A, B, and C.) I contend that it would be unusual, in a situation where we did have a full probability model for H, A_1, A_2, E_1, and E_2 for conditions $\mathscr{F}3$ and $\mathscr{F}4$ to hold.

Consider condition $\mathscr{F}4$ in particular. It says that $P[E_1|H\bar{A}_1] = P[E_1|\bar{H}\bar{A}_1]$ and $P[E_2|H\bar{A}_2] = P[E_2|\bar{H}\bar{A}_2]$. These equalities can be weakened to inequalities: if $\mathscr{C}2$, $\mathscr{F}1$, $\mathscr{F}2$, and $\mathscr{F}3$ holds then Expressions (12) and (13) and hence (11) will hold provided that

$$P[E_1|H\bar{A}_1] \geq P[E_1|\bar{H}\bar{A}_1] \text{ and } P[E_2|H\bar{A}_2] \geq P[E_2|\bar{H}\bar{A}_2]. \qquad (16)$$

But should we expect Expression (16) to hold? It would not hold, for example, if there were any tendency for the evidentiary mechanisms to provide incorrect information when they are not working correctly. (Suppose, to take an extreme example, that $\mathscr{C}1$, $\mathscr{F}1$, $\mathscr{F}2$, and $\mathscr{F}3$ hold, that $P[H] = P[A_1] = P[A_2] = 0.5$, but

*Notice that since A and E imply H, $P[H|E] \geq P[A|E]$. Thus, $P[A|E]$, the "probability that H is proven," is a lower bound for $P[H|E]$, the "probability that H is true." Notice, however, that by attending to $P[A|E]$ instead of $P[H|E]$ one does not avoid dependence on the prior probability $P[H]$.

that $P[E_i|H\bar{A}_i] = 0$ and $P[E_i|\bar{H}\bar{A}_i] = 1$ for $i=1,2$. In this case $P[A_1 v A_2|E_1E_2] = P[A_1|E_1] = P[A_2|E_2] = 0.5$, and so Expression (11) does not hold.)

When we use Hooper's or Ekelöf's rules are we assuming Expression (16)? Does the implausibility of Expression (16) invalidate arguments based on these rules? No. Here, as in our discussion of Laplace above, we must insist that the non-Bayesian arguments have their own logic. They do not use a model that gives reality to conditional probabilities such as those in Expression (16), and hence they do not depend on relations between such conditional probabilities.

B. Equality or Inequality?

Hooper's rule, Expression (10), is an equality, but Halldén and Edman's argument produces Expressions (11), an inequality. The inequality can be explained by saying that the concurrence of the second evidentiary mechanism increases the credibility of the first (see, for example, Freeling and Sahlin,[28] p. 66). Should we rely on this intuitive idea in general and therefore always construe Hooper's rule as an inequality?

Again, we must answer no. Making the idea that the concurrence mutually increases the credibility of the evidentiary mechanisms into a genuine argument requires a full probability model. Indeed, as we have seen, it requires questionable assumptions about the behavior of the mechanisms when they are not working correctly. If we are not willing to make these assumptions, we have only a hint of an argument.

We should, in general, resist the temptation to think of the numbers given by non-Bayesian probability arguments as bounds on more precise but unknown probabilities. In the case of the chain of 100 reporters each with credibility 100/106, for example, we are tempted to say that $(100/106)^{100} \approx .003$ is a lower bound for the probability of the report, which may be higher because of other favorable arguments. But there may just as well be other negative arguments. The meager support lent to the report by the chain of testimony does not imply a strong probability against the report; we cannot say that a denial of the report has probability .997. But it is otiose to explain this by saying that there is a true probability for the report known to lie between .003 and 1.

VI. DEMPSTER

Both Hooper's and Lambert's rules apply only in problems where the arguments being combined bear directly on a single proposition, so that only two possibilities need be considered. In the story about my friend Ilse, for example, both arguments bore directly on whether Ilse is Catholic, and so we considered only the possibility that she is and the possibility that she is not. The idea of combining independent probability arguments can, however, be generalized to the case where the different arguments bear on different propositions and so more than two possibilities must be considered. This generalization was first formulated by A. P. Dempster.[29]

The nature of the generalization can be indicated by example. Returning to the story of Ilse, let us now drop the assumption, implicit on our earlier analysis, that she is Catholic if she is attending Catholic mass. If we retain the assumption that she is either Catholic or Protestant, then we may consider four possibilities:

1) Ilse is Catholic and plans to attend Catholic mass Sunday afternoon.
2) Ilse is Protestant but plans to attend Catholic mass Sunday afternoon.
3) Ilse is Protestant but plans to attend a Protestant service Sunday afternoon.
4) Ilse is Catholic but plans to attend a Protestant service Sunday afternoon.

The argument that Protestant services are not usually held on Sunday afternoons is now an argument for "1 or 2," whereas Ilse's Austrian origin is an argument for "1 or 4." We must also make explicit a third argument: the argument that Ilse is probably attending a church of her own faith. This argument is based on the fact that people usually attend a church of their faith and on the fact that Ilse mentioned the church service in a way that suggested that it was not a special event for her. It is an argument for "1 or 3." I would give this argument, standing on its own, 98% credibility.

Altogether, then, we have three independent arguments, each of which may be right or wrong. The probabilities for what each of these arguments mean are given in Table II. Since the arguments are independent, we may multiply these probabilities to find probabilities for what they mean together. The eight resulting probabilities are listed in Table III. The first line of this table indicates that there is a probability of .7056 that all three arguments are correct, in which case "1" is true, etc.

The third line of Table III requires special notice. As this line indicates, the probabilities for the arguments and their independence imply that there is a probability .0784 that arguments *A* and *C* are correct but argument *B* is wrong, but the conclusions to which the arguments point show that this is impossible. (It is impossible for "1 or 2," "2 or 3," and "1 or 3" to all be true; we represent this impossibility by the symbol "$\emptyset$.") So we must condition on this impossibility by eliminating the probability .0784 and multiplying the other probabilities by $1/(1-.0784) = 1.085$.

Table III.

Chance	Meaning
$.8 \times .9 \times .98 = .7056$	(1 or 2) and (1 or 4) and (1 or 3) = 1
$.8 \times .9 \times .02 = .0144$	(1 or 2) and (1 or 4) and (nothing) = 1
$.8 \times .1 \times .98 = .0784$	(1 or 2) and (2 or 3) and (1 or 3) = $\emptyset$
$.8 \times .1 \times .02 = .0016$	(1 or 2) and (2 or 3) and (nothing) = 2
$.2 \times .9 \times .98 = .1764$	(nothing) and (1 or 4) and (1 or 3) = 1
$.2 \times .9 \times .02 = .0036$	(nothing) and (1 or 4) and (nothing) = 1 or 4
$.2 \times .1 \times .98 = .0196$	(nothing) and (2 or 3) and (1 or 3) = 3
$.2 \times .1 \times .02 = .0004$	(nothing) and (2 or 3) and (nothing) = 2 or 3

We find, therefore, a total probability of

$$(1.085)(.7056 + .0144 + .1764) = .9727$$

for possibility "1," the possibility that Ilse is Catholic and the Sunday afternoon service is a Catholic mass. And we find a total probability

$$(1.085)(.7056 + .0144 + .1764 + .0036) = .9766$$

for the proposition that Ilse is Catholic. This differs only slightly from the probability .978 that we found in Section III, where we neglected the possibility that Ilse might be attending a service of another faith.

Of course, the story about Ilse contains many other assumptions that can be examined. Why do I assume that she is either Catholic or Protestant? Was she telling me the truth when she mentioned the church service? Is her husband's growing up in Los Angeles really irrelevant?, etc. The evidence for each assumption can be brought into the analysis, but always at the cost of complicating the analysis and multiplying the number of possibilities considered.

The fact that the evidence for assumptions can be brought into the analysis in this way partially refutes Ekelöf's insistence that the assessment of auxiliary facts cannot be formalized. Ekelöf is correct, however, to insist that the formalization can never be complete. There will always be presumptions, circumstances, and auxiliary facts that remain outside the formal framework and provide the basis for the judgments made within that framework.

A. DEMPSTER'S GENERAL RULE OF COMBINATION

In order to use Dempster's rule, we must always list the possible conclusions of the independent arguments we are combining and their probabilities. Suppose, for simplicity, that there are only two arguments. Denote the possible conclusions of the first argument by $A_1, \ldots, A_m$, and their probabilities by $r_1, \ldots, r_m$. (Here $\Sigma_{i=1}^{m} r_i = 1$.) Similarly denote the possible conclusions of the second argument by $B_1, \ldots, B_n$, and their probabilities by $s_1, \ldots, s_n$. The basic idea of Dempster's rule is that the two arguments together yield the conclusion "A_i and B_j" with probability $r_i s_j$. The combination "A_i and B_j" may be impossible, however, and we must condition on such impossibilities—i.e., we must eliminate their probabilities and multiply the other probabilities by

$$K = \frac{1}{1 - \Sigma\, r_i s_j}, \tag{17}$$

where the sum is over all pairs (i,j) such that "A_i and B_j" is impossible. Dempster's rule says, then, that the probability given to a proposition C by the two arguments together is

$$K\, \Sigma\, r_i s_j, \tag{18}$$

where the sum is over all pairs (i,j) and such that "A_i and B_j" is possible and implies C.

If we denote by Bel(C) the probability or "degree of belief" given C by the two arguments together, then Expressions (17) and (18) may be combined in the formula

$$\text{Bel}(C) = \frac{\Sigma \{r_i s_j | \text{"}A_i \text{ and } B_j\text{" is possible and implies } C\}}{1 - \Sigma \{r_i s_j | \text{"}A_i \text{ and } B_j\text{" is impossible}\}} .$$

This is equivalent to

$$\text{Bel}(C) = \frac{\Sigma \{r_i s_j | \text{"}A_i \text{ and } B_j\text{" is possible and implies } C\}}{\Sigma \{r_i s_j | \text{"}A_i \text{ and } B_j\text{" is possible}\}} .$$

B. RANDOMLY CODED MESSAGES AND BELIEF FUNCTIONS

Underlying Dempster's rule is the idea of an argument that has different possible meanings, each assigned a probability. We have illustrated this idea with examples of arguments about Ilse's religion. In Shafer[18] and Shafer and Tversky,[13] the idea is illustrated with the example of a "randomly coded message."

Someone chooses a code at random from a list of codes, uses the chosen code to encode a message, and then sends us the result. We consider the original or "plaintext" message to be completely reliable. We know the list of codes and the chance of each code being chosen—say the list is $c_1, \ldots, c_n$, and the chance of code c_i being chosen is r_i. Let Z denote the coded message we receive. We decode Z using each of the codes and find, in each case, a message that makes sense. Let A_i denote the message we get when we decode using c_i. We may say that the evidence provided by the received message "means" A_i with probability r_i. For a given proposition C, we may say that

$$\text{Bel}_1(C) = \Sigma \{r_i | A_i \text{ implies } C\} \tag{20}$$

is the probability or "degree of belief" given to C by this evidence alone. In Shafer,[30] functions that assign numbers to propositions in the manner of Expression (20) are called "belief functions."

Now suppose we receive a second message, again encoded by a randomly chosen code. This time the codes are $d_1, \ldots, d_n$, say, with probabilities $s_1, \ldots, s_n$. The corresponding possible plaintext messages are $B_1, \ldots, B_n$. This evidence, by itself, would give probability

$$\text{Bel}_2(C) = \Sigma \{s_i | B_i \text{ implies } C\}$$

to a proposition C.

If the codes for the two messages are chosen independently, then their possible meanings are the combinations "A_i and B_j," with probabilities $r_i s_j$. Eliminating the "A_i and B_j" that contradict each other, we arrive at Expression

(19) as the probability the two messages together give to C. The belief function Bel given by Expression (19) is the result of combining the belief functions Bel_1 and Bel_2 by Dempster's rule.

C. Williams and Good's Bayesian Interpretation

P. M. Williams[31] and I. J. Good[32] have independently discussed the story of the randomly coded message from a Bayesian point of view.

Let us review the story. A single plaintext message is encoded. The possible codes are $c_1, \ldots, c_m$, and the probability of c_i being used is r_i. We receive an encoded message Z, and A_i is the message obtained when Z is decoded using c_i. The argument of the preceding section boils down to saying that we continue to think of r_i as the probability for c_i after we have seen Z. This means that r_i becomes a probability for A_i being the correct plaintext message. Given any proposition B,

$$\Sigma \{r_i | A_i = B\} \tag{21}$$

is the total probability that the correct plaintext message is B, and given any proposition C, Expression (20) is the total probability that the correct plaintext message implies C.

Williams and Good, taking the same tack as Halldén and Edman in their Bayesian analysis of Ekelöf's rule for concurring evidence, have asked under what circumstances Expression (21) will be a Bayesian's probability for B being the correct plaintext message—i.e., under what circumstances the equation

$$P[B \text{ is the plaintext message } | Z \text{ is the received message}] = \Sigma\, r_i | A_i = B\} \tag{22}$$

will hold.

In order to carry out a Bayesian analysis, we must assign probabilities to possibilities for the correct plaintext message. Let us also assume that the choice of this plaintext message is independent of the choice of a code. Then

$$P[B \text{ is the plaintext message } | Z \text{ is the received message}]$$

$$= \frac{P[B \text{ is the plaintext message and } Z \text{ is the received message}]}{P[Z \text{ is the received message}]}$$

$$= \frac{P[B \text{ is the plaintext message}]\, P[Z \text{ is the received message } | B \text{ is the plaintext message}]}{P[Z \text{ is the received message}]}$$

$$= \frac{P[B \text{ is the plaintext message}]\, \Sigma\, r_i | A_i = B\}}{P[Z \text{ is the received message}]}.$$

This makes it clear that unless both sides of Expression (22) are zero, (22) can hold only if

$$P[B \text{ is the plaintext message}] = P[Z \text{ is the received message}]. \quad (23)$$

As Good (p. 343)[32] points out, Expression (23) does not seem to be a very natural assumption. Williams (p. 342,[31] and private correspondence, 1980) points out, however, that Expression (23) will hold under the following assumptions: (1) there are a finite number, say k, of messages that we initially consider to be possibilities for the plaintext message; (2) we assign each of these possible plaintext messages equal prior probability; and (3) each of the possible encoded messages can be decoded using any of the m codes and this always yields one of the possible plaintext messages. From assumptions (1) and (3) we see that

$$P[B \text{ is the plaintext message}] = \frac{1}{k}$$

for every possible plaintext message B. And it then follows by (3) that

$$P[Z \text{ is the received message}]$$
$$= \sum_{B} P[B \text{ is the plaintext message}]\, P[Z \text{ is the received message} \mid B \text{ is the plaintext message}]$$
$$= \sum_{i=1}^{m} \frac{1}{k} r_i = \frac{1}{k}$$

for every possible received message Z. (The summation over B is over all possible plaintext messages.)

We see, then, that the "belief-function" understanding of the story of the randomly coded message can be given a Bayesian justification only under very restrictive assumptions. This is not an unfamiliar situation. As we saw in Section IV above, Laplace gave a Bayesian analysis with equal prior probabilities that agreed with Lambert's rule for concurring testimony (with $p+q=1$). As we saw in Section V, Halldén and Edman gave a Bayesian justification for Hooper's rule for concurrent testimony that involved unreasonably restrictive assumptions. And similar situations are common in parametric statistical inference: there non-Bayesian methods frequently give results that accord with Bayesian analyses with equal prior probabilities for different parameter values (see, e.g., Cox and Hinkley, pp. 378–379[33]).

In all these cases it is tempting to say that the Bayesian analyses show what the non-Bayesian arguments really mean, and that the unreasonableness of the Bayesian assumptions show the non-Bayesian arguments to be wrong. This temptation should be avoided. The non-Bayesian arguments, though they do not pretend to be as definite as the Bayesian arguments pretend to be, have their own logic and can be convincing in their own right.

The author has profited from correspondence with Per Olof Ekelöf, Robert W. Goldsmith, Nils-Eric Sahlin, and Peter Williams. Research for this paper was partly supported by ONR Grant No. N00014-85-K-0490 and by a Guggenheim fellowship.

References

1. J. Gordon and E.H. Shortliffe, "A method for managing evidential reasoning in hierarchical hypothesis spaces," *Artificial Intelligence*, **26**, 323–358, 1985.
2. R.R. Yager, "Toward a general theory of reasoning with uncertainty. 1: Nonspecificity and fuzziness," *International Journal of Intelligent Systems*, **1**, 45–67, 1986.
3. P. Cheeseman, "In defense of probability," In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 1985, pp. 1002–1009,
4. P. Gärdenfors, B. Hansson, and N.-E. Sahlin, *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*, Lund, 1983.
5. G. Hooper, "A calculation of the credibility of human testimony," *Philosophical Transactions of the Royal Society*, **21**, 359–365, 1699.
6. G. Hooper, *A Fair and Methodical Discussion of the First and Great Controversy between the Church of England and Church of Rome concerning the Infallible Guide*, London, 1689.
7. G. Hooper, *The Works of the Right Reverend Father in God, George Hooper, D. D., Late Bishop of Bath and Wells*, Oxford, 1757.
8. G. Hooper, *Works*, 2 Vols., Oxford, 1855.
9. B. Grier, "George Hooper and the early theory of testimony," Department of Psychology, Northern Illinois University, Dekalb, IL, 1981.
10. G. Shafer, "Non-additive probabilities in the work of Bernoulli and Lambert," *Archive for History of Exact Sciences*, **19**, 309–370, 1978.
11. B.C. van Fraassen, *The Scientific Image*, Clarendon Press, 1980.
12. G. Shafer, "Conditional probability, *International Statistical Review*, **53**, 261–277, 1985.
13. G. Shafer and A. Tversky, "Languages and designs for probability judgment," *Cognitive Science*, 309–339, 1985.
14. J.H. Lambert, *Neues Organon*, Leipzig, 1764. Reprinted in 1985 by Olms as the first two volumes of Lambert's *Philosophische Schriften*.
15. I. Todhunter, *A History of the Mathematical Theory of Probability for the Time of Pascal to That of Laplace*, Cambridge, 1865. Reprinted in 1949 by Chelsea.
16. P.S. Laplace, *Theorie Analytique des Probabilities*, 2nd ed., Paris, 1814.
17. P.S. Laplace, *Oeuvres de Laplace*, Vol. VII, Paris, 1886.
18. G. Shafer, "Constructive probability," *Synthese*, **48**, 1–60, 1981.
19. S.M. Stigler, "Napoleonic statistics: The work of Laplace," *Biometrika*, **62**, 503–517, 1975.
20. P.O. Ekelöf, *Rättegang*, Vol. IV, Stockholm, 1963, 1969, 1973, 1977, 1982.
21. P.O. Ekelöf, "Free evaluation of evidence," *Scandinavian Studies in Law*, **8**, 45–66, 1964.
22. P.O. Ekelöf, "Beweiswert," In *Festschrift für Fritz Bauer*, W. Grunsky, R. Strürner, G. Walter, and M. Wolf, Eds., 1981, pp. 343–363.
23. P.O. Ekelöf "My thoughts on evidentiary value," In *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*, Lund, 1983, pp. 9–26.
24. A. Stening, *Bevisvärde*, Uppsala, 1975.
25. R.W. Goldsmith, "Studies of a model for evaluating judicial evidence," *Acta Psychological*, **45**, 211–221, 1980.
26. S. Halldén, "Indiciemenkanismer," *Tidsskrift for Rettsvitenskap*, **86**, 55–64, 1973.
27. M. Edman, "Adding independent pieces of evidence," In *Modality, Morality and*

Other Problems of Sense and Nonsense. Essays dedicated to Sören Halldén, Lund, 1973, pp. 180–188.

28. A.N.S. Freeling, and N.-E. Sahlin, "Combining evidence," In *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*, Lund, 1983, pp. 58–74.
29. A.P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," Annals of Mathematical Statistics, **38**, 325–339 (1967).
30. G. Shafer, *A Mathematical Theory of Evidence*, Princeton, NJ, 1976.
31. P.M. Williams, "Discussion of Belief Functions and Parmetric Models, by Glenn Shafer," *Journal of Royal Statistical Society, Series B*, **44**, 341–343, 1982.
32. I.J. Good, "Comment on Lindley's Paradox, by Glenn Shafer," *Journal of the American Statistical Association*, **77**, 342–344, 1982.
33. D.R. Cox. and D.V. Hinkley, *Theoretical Statistics*, London, 1974.