Part XI

Statistical Evidence

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

Two Theories of Probability

Glenn Shafer[1]

University of Kansas

In a recent monograph [8], I advocated a new theory--the theory of belief functions--as an alternative to the Bayesian theory of epistemic probability. In this paper I compare the two theories in the context of a simple but authentic example of assessing evidence.

The Bayesian theory is ostensibly the theory that assessment of evidence should proceed by conditioning additive probability distributions; this theory dates from the work of Bayes and Laplace in the second half of the eighteenth century. It is indisputably the dominant theory of epistemic probability today.

The theory of belief functions differs from the Bayesian theory in that it uses certain non-additive set functions in the place of additive probability distributions and in that it generalizes the rule of conditioning to a general rule for combining evidence. As a mathematical theory its apparent origin is rather recent and abrupt; it first appears in work of A. P. Dempster, published in the 1960's. (See, for example, [3].) But in both its mathematical form and its philosophical content it echoes pre-Bayesian eighteenth-century ideas propounded by James Bernoulli and J. H. Lambert. (See [9].)

The idea that evidence should be assessed by conditioning additive probability distributions can be interpreted in several ways. It might be interpreted to mean that we should anticipate our evidence--that before obtaining evidence we should have a probability distribution which includes a probability for that evidence and which therefore can be conditioned on that evidence. In the theoretical arguments for the Bayesian theory (the "coherence arguments"), the idea does seem to be interpreted in this way. But as I argue in Section 5 below, this interpretation produces an untenable theory; most often we simply cannot anticipate our evidence. A more plausible interpretation is that after obtaining evidence we should construct an additive probability distribution that includes a probability for that evidence. But as I

PSA 1978, Volume 2, pp. 441-465

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

argue in Section 7 below, even this idea is usually impossible to put into practice.

I conclude that the Bayesian theory of probability does not, in its practical form, really assess evidence by conditioning; instead it simply uses various rules of the additive probability calculus--principally Bayes' theorem--to organize probability judgments once evidence has been obtained. And since, contrary to what the formal explanation of these rules suggests, there is no probability distribution really being conditioned, this practical Bayesian theory can be defended only in terms of its success in organizing probability judgments. It is on a par, in this respect, with the theory of belief functions.

The paper is organized as follows. First I present the example that serves as a backdrop for the whole paper. Next I review the theory of belief functions and apply it to the example. Then I turn to the Bayesian theory, first discussing its theoretical problems and then applying it to the example. I conclude with a summary of the paper's main theses.

1. An Example

My house has been enlarged several times. There are several rooms that have been added within the past ten years, and the east and west sections of the older part of the house were apparently also built at different times. The east section has an oak floor, the west section a concrete floor. Which was built first?

A few months ago I did not know. But now I have muddled about and obtained some evidence: (1) The partition between the two sections, which was probably the outside wall when only one was there, appears to rest on the concrete floor. It definitely does rest on concrete, and though it might rest on a separate concrete foundation, it does not seem to do so; so far as I can tell there is no crack between the concrete on which it rests and the concrete floor. (2) The footing under the oak floor along the partition does not seem adequate to support an outside wall. (3) In the attic above the partition, I find stubs of what appear to have been rafters for a roof covering only the section with the concrete floor. (4) I gather from conversations with a neighbor that the original building was the office for a small gravel quarry. An oak floor seems out of place in such a building.

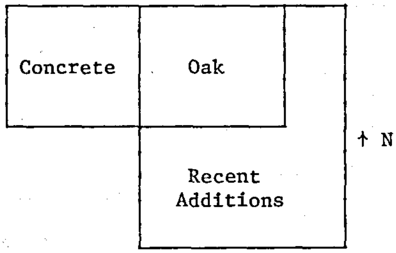

Figure 1. The House

I am now morally certain that the section with the concrete floor was built first.

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

I have just described my evidence in words, by listing four arguments for the section with the concrete floor having been built first. But, of course, this is only a description. It would be impossible to put into words all the details that determine the force of conviction conveyed by these arguments. Much of the evidence, especially in arguments (2) and (3), is essentially visual; if the reader wants to judge for himself the force of this evidence, he must come and look. And arguments (1) and (4) also leave much unsaid: my reasons for thinking the partition had been the outside wall, my impression of the reliability of my elderly neighbor's hesitant memory, my impression of the character of the small quarry, etc.

This example is similar in spirit to examples that James Bernoulli presented in Ars Conjectandi. There are several different items of evidence, corresponding to several different small worlds of experience and understanding that can be brought to bear on the matters in question. And our task is to assess the strength of each of these items of evidence and to combine these assessments into overall probability judgments. We shall see how our two theories of probability help us in this task.

2. The Theory of Belief Functions

The theory of belief functions provides two basic tools to help us make probability judgments: a metaphor that can help us organize probability judgments based on a single item of evidence, and a formal rule for combining probability judgments based on distinct and independent items of evidence.

Here is the metaphor for organizing our probability judgments. We think of our belief (or our "probability," if you will) as a whole and imagine committing parts of that whole ("probability masses") to various propositions. The amount we commit represents a judgment as to the strength of the evidence that specifically favors that proposition. It is not required that belief not committed to a given proposition should be committed to its negation, nor that belief committed to a given proposition should be committed more specifically.

Consider, for example, my elderly neighbor's hesitant report that my house was once the office for a small quarry. I have no evidence against what he says--indeed, other remnants of the quarry are nearby. Thus, I do not want to commit any of my belief to the denial of his report. On the other hand, his own uncertainty and my doubts about the reliability of his memory mean that I am willing to commit only part of my belief--a relatively small part, say .2--on the basis of his report. So I have a degree of belief of .2 that my house was once the office for the quarry and a degree of belief of zero that it was not.

(The Bayesian theory is often explained in terms of a similar metaphor of probability masses--cf., [2], p. 104. But in the Bayesian metaphor all one's probability must be committed to one side or the other of every question.)

Probability judgments are, of course, always difficult to make. They are, in the last analysis, a product of intuition. But the job is often easier if we can focus on a well-defined and relatively simple item of evidence, such as my neighbor's testimony, and make our judgments on the basis of that evidence alone. The theory of belief functions suggests that we try to break our evidence down into such relatively simple components, that we make probability judgments separately on the basis of each of these components, and that we then think about how to combine these judgments. This makes sense intuitively if the different components of evidence seem to involve (depend for their assessment on) intuitively independent small worlds of experience. And in this case the theory offers a formal rule for combining the probability judgments--Dempster's rule of combination.

We can fully develop these ideas, of course, only if we have a mathematical structure to represent the propositions about which we are making probability judgments. Following the usual practice of statisticians, we may obtain such a structure by setting out an exhaustive list Θ of mutually exclusive possibilities and interpreting each subset A of Θ as the "proposition" that the truth is in A. I call such a list Θ a "frame of discernment" in order to emphasize that we may always split each possibility in Θ into more specific possibilities and thus increase the number of propositions that Θ "discerns".

Having set out a frame of discernment Θ, we can then think about a given item of evidence by asking whether it points specifically to certain subsets, how much of our belief we should commit specifically to each subset, how much should be committed more specifically, etc. *Setting our total degree of belief for each subset equal to the measure* of all the belief committed either exactly to it or else to something more specific, we obtain a "belief function" representing that evidence. Dempster's rule is a rule for combining such belief functions. (See Chapters 2 and 3 of [8] for details.)

3. Constructing the Frame of Discernment

The choice of the frame of discernment Θ is of great intuitive as well as mathematical importance. For one thing, it is tied up with the attempt to analyze the evidence into intuitively independent components or "small worlds". Intuitively, Θ should be the meeting ground of these small worlds--it should be where they do come together. This is what we mean when we ask for the different items of evidence to be "independent" or "non-interacting". It is not that they should not interact at all; it is just that all their interaction should be in terms of the issues discerned by Θ. That is to say, Θ should "discern the interaction of the evidence". The practical problem in applying the theory of belief functions is to choose a frame Θ and a decomposition of the evidence so that this is so.

What issues are involved in the interaction of the evidence in the example concerning my house? There seem to be three such issues: which section was built first, whether the partition rests on the concrete floor, and whether the partition is the original outside wall. So I need a frame which will discern whether

A: the concrete section was built first or
$\bar{A}$: the oak section was built first,

B: the partition rests on the concrete floor or
$\bar{B}$: the partition rests on a separate concrete foundation,

C: the partition was formerly the outside wall or
$\bar{C}$: the former outside wall has been removed.

If these three dichotomies were regarded as three logically independent dichotomies in a formal language, then they would generate $2 \times 2 \times 2 = 8$ possibilities. But I am not working with a formal language. I am working with my practical understanding of a real problem, and just as this understanding allows me to formulate these dichotomies it may allow me to exclude some of these "theoretical" possibilities. In fact, I shall exclude $\bar{A}$ & B & C; it is too fantastic to imagine the wall being jacked up, its original foundation being removed, and the concrete floor being poured to extend beneath it.

So I have a frame Θ consisting of seven possibilities:

$\theta_1 = A$ & B & C: the concrete section was built first; the partition, which rests on it, was then the outside wall;

$\theta_2 = A$ & $\bar{B}$ & C: the concrete section was built first; the partition, which rests on a separate foundation, was then the outside wall;

$\theta_3 = \bar{A}$ & $\bar{B}$ & C: the oak section was built first; the partition, which rests on a separate foundation, was then the outside wall;

$\theta_4 = A$ & B & $\bar{C}$: the concrete section was built first, and the original outside wall rested on it; so does the present partition, which has replaced that original wall;

$\theta_5 = A$ & $\bar{B}$ & $\bar{C}$: the concrete section was built first, and the original outside wall rested on it; the present partition is on a separate foundation;

$\theta_6 = \bar{A}$ & B & $\bar{C}$: the oak section was built first, and the original outside wall rested on it; the present partition is on the concrete section;

$\theta_7 = \bar{A}$ & $\bar{B}$ & $\bar{C}$: the oak section was built first, and the original outside wall rested on it; the present partition is on a separate foundation.

And having laid this frame Θ out, I can now think of A, B, and C as subsets of Θ: $A = \{\theta_1,\theta_2,\theta_4,\theta_5\}$, $B = \{\theta_1,\theta_4,\theta_6\}$, and $C = \{\theta_1,\theta_2,\theta_3\}$. And I can write $\{\theta_1\} = A \cap B \cap C$, etc.

Notice that I have made explicit in my list of the seven possibilities some assumptions that I had not mentioned before. The description

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

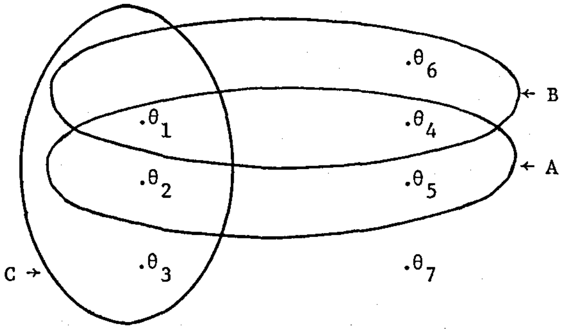

Figure 2. The Possibilities

of the possibility θ_4 makes explicit, for example, my exclusion of the possibility that the original outside wall rested on a foundation that was removed when the oak section was added. No doubt even my linguistic description of the seven possibilities leaves many aspects of my practical understanding implicit; there may well be fantastic scenarios that are consistent with the bare form of the words I have written but which are so inconsistent with my conception of these possibilities that I would reject them out of hand once they were drawn to my attention. This does not necessarily mean that I have expressed myself poorly; it is just that concrete conceptions in concrete circumstances are never fully captured by linguistic descriptions.

It is often held that the objects of probability judgments are linguistic objects: either sentences or else classes of logically equivalent sentences, these equivalence classes being called "propositions". Here I am asking the reader to take a slightly broader view. The objects of our probability judgments are possibilities and sets of possibilities, and it is required only that we have adequate means of distinguishing among, sorting, and grouping these possibilities. These means may be purely mental, or they may involve records and computing machines, but it is not required that the possibilities themselves should be sentences in any language, natural or symbolic. Thus we may use the word "proposition" without any linguistic connotations--it refers simply to the intuitive content of the assertion that the truth is in a given subset of possibilities.

Ian Hacking [5] has convincingly argued against the view that probability judgments have equivalence classes of sentences as their objects, on the grounds that we may not know whether given sentences are logically equivalent. We may not know, for example, that a given mathematical assertion is true and provable and hence logically equivalent to any other sentence which is true by virtue of its logical form. "Hence," he writes, "we must cast about for other objects for personal probability. Sentences are the obvious choice." But perhaps it is not necessary to cast about for so formal an object. Perhaps it is best to

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

say, as Hacking does, that the objects of our "personal probabilities" are "personal possibilities", and leave it at that. We can always undertake, surely, to give a linguistic description of each of these possibilities; but there is no reason to pretend that we have fixed on any particular linguistic description or that any such description would be fully adequate.

Notice that my frame of discernment Θ is itself very much a product of my evidence. It was only in the process of obtaining the evidence that I even thought about the possibility of a separate concrete foundation. And many of the implicit assumptions that help define my possibilities also result from my evidence. It might be argued against many of these assumptions that they themselves are based only on probabilities, and that they should not necessarily be taken for granted as a preliminary to making careful probability judgments. This is true, and further thought about my evidence may in fact lead me to isolate and question some of these assumptions. But when we make probability judgments, we must always, I think, take for granted some things that may later be called into question.

4. Making the Probability Judgments

In introducing our example I listed four items of evidence. These items do seem to interact only in terms of our frame of discernment, and it even seems reasonable to further decompose the first item by distinguishing between the presumption that the partition is probably the original outside wall and the uncertain perception that it rests on the concrete floor. This gives us five items of evidence:

E_0: The partition is probably the original outside wall. (For it would have been a pointless expense not to retain the outside wall as a partition.)

E_1: The partition appears to rest on the concrete floor.

E_2: The footing under the oak floor seems inadequate for an outside wall.

E_3: Stubs in the attic suggest a roof over the concrete section.

E_4: My neighbor's testimony as to the original use of the house suggests a concrete floor.

It is, of course, merely a judgment that these items of evidence interact only in terms of the frame Θ--a judgment of the same character as the probability judgments I am about to make. Another person might suspect that my assessment of the footing under the oak floor and my assessment of the stubs in the attic are both based on some basic misunderstanding I have about houses. But in my judgment my uncertainties concerning these two matters are entirely distinct.

How do these five items of evidence bear on the frame Θ? It seems to me that each of them bears on Θ in a very simple way--each supports a single subset of Θ. The question, in each case, is how strongly it supports that subset. In the theory of belief functions, this "degree

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

of support" is equated with the degree of belief (or the proportion of my probability) that I would commit to that subset on the basis of the given item of evidence alone, in the absence of the other evidence.

Let us consider each item of evidence in turn.

E_0. My general understanding of the motives and circumstances involved in enlarging houses warrants, standing alone, a high degree of belief that the present partition is the original outside wall--i.e., a high degree of belief in the subset $C = \{\theta_1, \theta_2, \theta_3\}$. How high? The number .8 seems conservative but reasonable. (This does not mean that I think the original wall is replaced in as many as .2 of the cases where a house is enlarged. I would be hard pressed to estimate how often such walls are replaced; but if I had to make a guess, I would guess much lower than .2. The number .8 is merely the proportion of my probability that I wish to commit on the basis of my imperfect understanding.)

E_1. It looked to me as if the partition rested on the concrete floor--I couldn't see a crack. How clear was this perception? I think it warrants a degree of belief of .98 in the proposition $B = \{\theta_1, \theta_4, \theta_6\}$, which says that the partition does rest on the concrete floor. I might try to explain this by saying that the perception was so clear and convincing that at least 98% of "such perceptions" must surely be right. But this is only a strained metaphor; in truth the category of "perceptions such as this one" has only a verbal reality. The number .98 is ultimately simply a probability judgment.

E_2. The footing under the oak floor seems slight. Surely you would not set an outside wall on so slight a footing. But you might if you were skimping and knew better than I what you could get away with. Still, I would be willing to commit half of my belief that no outside wall was ever supported by this floor. This means a degree of belief of .5 for the set of possibilities $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$.

E_3. The stubs of 2" × 4" lumber that I saw in the attic are attached directly above the present partition, and it is hard for me to imagine what else they might be but remnants of rafters for a roof that covered only the concrete section and was anchored to a wall where this partition now is. Of course, it was a while before I came up with this idea; and then I let my imagination rest. So I am not entirely certain that I am not overlooking some other possibility. So I am inclined to be cautious: I shall represent this evidence by awarding a degree of belief of .8 to the set of possibilities $\{\theta_1, \theta_2, \theta_4\}$.

E_4. I have only limited confidence in my neighbor's testimony, and my speculation based on it must be discounted yet further. Say I give

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

Table 1

Evidence	Focus	Degree of Support
E_0	$C = \{\theta_1,\theta_2,\theta_3\}$	.8
E_1	$B = \{\theta_1,\theta_4,\theta_6\}$	.98
E_2	$\{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5\}$	.5
E_3	$\{\theta_1,\theta_2,\theta_4\}$	.8
E_4	$A = \{\theta_1,\theta_2,\theta_4,\theta_5\}$	.08

* * * * *

credence .2 to the testimony and .4 to the chain of speculation; this yields a degree of belief of $.2 \times .4 = .08$ in the proposition $A = \{\theta_1,\theta_2,\theta_4,\theta_5\}$, that the concrete section was built first.

So for each of the five items of evidence I have identified a subset of Θ that the evidence directly supports (its "focus") and made a judgment as to the strength of this support. In each case the result is a "simple support function"--a belief function that commits a certain amount of my belief to the focus without commiting the rest to anything in particular. In the case of the evidence E_0, for example, I have commited .8 of my belief to the focus $C = \{\theta_1,\theta_2,\theta_3\}$ (= the partition is the original outside wall) without commiting the other .2 to $\bar{C} = \{\theta_4,\theta_5,\theta_6,\theta_7\}$ or, indeed, to anything except Θ itself. Of course, the .8 committed to C is also committed to any other subset containing C, and so my degree of belief, on this evidence, for an arbitrary subset X of Θ is

$$Bel_0(X) = \begin{cases} 0 & \text{if} \quad C \not\subset X \\ .8 & \text{if} \quad C \subset X \neq \Theta \\ 1 & \text{if} \quad X = \Theta \end{cases} .$$

For a further discussion of simple support function, see [8], p. 75.

It is intuitively clear that these five items of evidence together provide strong support for the subset $A = \{\theta_1,\theta_2,\theta_4,\theta_5\}$, the proposition that the concrete section was built first. Items E_3 and E_4 each support A by themselves. And further support is provided by the interaction of the other three items. Items E_1 and E_2, for example, interact to support $\{\theta_1,\theta_4,\theta_6\} \cap \{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5\} = \{\theta_1,\theta_4\}$, which is a subset of A. Dempster's rule of combination is designed to reflect such interaction, so we should expect it to produce a high degree of belief for A.

When we do combine the five simple support functions by Dempster's rule, we obtain the belief function whose "basic probability numbers" are given in Table 2. These numbers represent the amounts of belief "exactly committed" to the various subsets of Θ. Only the "focal elements", the subsets that have positive belief exactly committed to them, are listed; the basic probability numbers for other subsets are zero. The basic probability numbers sum to one; this is the measure of our total belief. To obtain the total amount of belief committed to a given subset X, we must add the basic probability numbers for all its subsets. Thus the total belief committed to $A = \{\theta_1,\theta_2,\theta_4,\theta_5\}$ is the sum of the basic probability numbers marked with a check in Table 2: .978288, or about .98. (The six significant figures are shown in Table 2 only to show the full calculation; surely no more than the two significant figures in the final answer Bel(A) = .98 would be meaningful.) And the total committed to $\bar{A} = \{\theta_3,\theta_6,\theta_7\}$ is zero; we may express this by saying that A has plausibility one.

Since it does not involve conflicting evidence, this example does not fully illustrate Dempster's rule. When there are conflicting items of evidence, so that every possibility in Θ has some evidence against it, there may not be a proper subset of Θ that has plausibility one.

* * * * *

Table 2

Focal Element	Basic Probability Number	
$\{\theta_1\}$	.784000	✓
$\{\theta_1,\theta_4\}$	.177968	✓
$B = \{\theta_1,\theta_4,\theta_6\}$	.018032	
$\{\theta_1,\theta_2\}$	.013056	✓
$\{\theta_1,\theta_2,\theta_4\}$	.003200	✓
$C = \{\theta_1,\theta_2,\theta_3\}$	.002944	
$\{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5\}$	.000368	
$\Theta = \{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\theta_7\}$	.000368	
$A = \{\theta_1,\theta_2,\theta_4,\theta_5\}$	.000064	✓

* * * * *

5. Conditioning on the Evidence

In the global form in which it is usually defended, the Bayesian theory of probability offers a startling contrast to the understanding of evidence developed in the preceding pages. In this global theory, it is assumed that our thinking before obtaining new evidence is in

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

terms of a frame of discernment Θ so detailed that it is sure to discern both that evidence and the propositions on which that evidence will be brought to bear. And we are told we should have an additive probability distribution P over Θ giving our "prior" degrees of belief--i.e., our degrees of belief before obtaining the evidence. Assessment of the evidence then comes down to "conditioning" P.

Let us spell this out in more detail. When we say that Θ discerns the evidence, we mean that the evidence corresponds to a subset E of Θ--i.e., that insofar as it bears on Θ, the evidence amounts to conclusive evidence that the truth is in E. This means that the evidence really does no more than rule out the possibilities in $\bar{E}$. As de Finetti ([2], p. 141) puts it, experience "acts always and only" by "suppressing the alternatives that turn out to be no longer possible." Before obtaining the evidence we cannot know, of course, which subset E of Θ it will correspond to in this way, but it is the Bayesian assumption that it will correspond to some subset in this way. It is also assumed that any proposition on which we will want to bring this evidence to bear will correspond to the truth being in some subset A of this same frame of discernment Θ. And we are told to take the evidence E into account by changing our degree of belief in such a proposition from P(A) to the "conditional probability"

$$P(A|E) = \frac{P(A \cap E)}{P(E)} \quad . \tag{1}$$

(See, for example, Lindley [6], p. 11.)

This is a very striking picture. But our example illustrates quite well the perplexity that often arises when we try to fit this picture to a problem from actual experience. The picture simply does not fit, for in real problems we do not anticipate our evidence before obtaining it. We do not have the required global frame of discernment.

In the matter of my house, it is simply fact that I did not anticipate the possibility of obtaining the sorts of evidence that I finally did obtain. My inspection of the attic, for example, was undertaken before I had even realized that the two parts of the house might have been built at different times; I merely wished to understand whether the wall between them was a bearing wall and whether it could be removed. I did not anticipate seeing the odd ends of 2" × 4" lumber that I now believe to be remnants of rafters; and even when I did see them, it took me some time to realize what I was seeing. Similarly, I started chatting with my neighbor without foreseeing what he might tell me, and I wriggled under the oak floor to attend to the plumbing without foreseeing that I might encounter so makeshift a footing.

Does the fact that we do not usually anticipate our evidence really threaten the rationale or the generality of the Bayesian theory? Most of the Bayesians with whom I have discussed the question deny that it does. They argue along the following lines:

> Certainly I do not in practice foresee all the possibilities as to how my evidence might turn out. But after I have obtained the evidence I want to assess it coherently. So I must think about how it would be assessed by a person wise enough or lucky enough to have anticipated the possibility of just such evidence. I imagine myself in the shoes of such a person, one who has anticipated the possibility of just this evidence but who is otherwise limited to the experience and wisdom that I had before obtaining it. I realize that this person needs to have an additive probability distribution in order to be coherent, so I put myself to constructing one that fits my experience and wisdom. And I realize that coherence also requires this person to assess the evidence by conditioning his probability distribution, so I condition the one I have constructed.

According to this view, use of the Bayesian theory to assess given evidence E may involve imagining that we have not yet obtained E but that we have a frame Θ that discerns E. The trick is to imagine this vividly enough that we can construct an additive probability distribution P over Θ.

In the simplest case, where there is only a single proposition A whose posterior probability we wish to assess, this exercise of imagination may be limited to answering two questions: "If somehow the possibility of E had been suggested to us before it happened, then what degree of belief would we have had that it would happen? And what degree of belief would we have had that it would happen <u>and</u> that A is true?" Our answers to these questions can be called $P(E)$ and $P(A \cap E)$, respectively; they are all that is needed to calculate $P(A|E)$ by (1), and giving them amounts to constructing an additive probability distribution P over a frame Θ that discerns only three possibilities: $A \cap E$, $\bar{A} \cap E$, and $\bar{E}$. In many cases we may prefer a more detailed frame, either because we want to assess the posterior probability of several propositions or simply because we want to think things through more carefully; and then our exercise of imagination in constructing P will have to be more extensive.

It is no objection to the Bayesian theory that it requires us to exercise our imagination. But why should we exercise our imagination in this particular way? Do the Bayesian "coherence arguments", which are often considered the fundamental rationale of the Bayesian theory, really give us any reason to do so? More importantly, can we expect an attempt to exercise our imagination in this way to work? Can we expect that we will be able to answer the questions asked, and that our answers will capture the message of the evidence?

6. The Coherence Arguments

Most contemporary Bayesians contend that the Bayesian theory is normative--a rational thinker <u>ought</u> to have degrees of belief given by an additive probability distribution, and he <u>ought</u> to assess new evidence by conditioning that distribution. The arguments on which these

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

contentions are based all involve, in one way or another, ideas about preferences among gambles. But since these arguments are associated with the slogan that non-Bayesian thinking is "incoherent", it is convenient to call them the "coherence arguments".

Whether the coherence arguments are persuasive under any circumstances is a vexed question, and I shall not treat it here. But I do wish to make a point I think indisputable: These arguments are not intelligible unless it is assumed that one has anticipated the evidence to be conditioned on. And so they cannot be counted as arguments for a Bayesian treatment of unanticipated evidence.

Consider the arguments for the additivity of our degrees of belief over a frame of discernment Θ. We consider gambles over Θ, such a gamble being an agreement that we are to pay or be paid a sum of money once it is learned which possibility θ in Θ is the truth, the sum and direction of payment depending on the θ. Any total ordering of such gambles that satisfies certain conditions corresponds to an additive probability distribution over Θ, and this fact is exploited either by arguing that the preferences of any thoughtful person should satisfy axioms that entail such a total ordering (e.g., Savage [7]), or else by imagining that we are forced to offer to take one side or the other of every gamble, in which case Dutch book can be made against us unless our choices of sides are given by such a total ordering (e.g., de Finetti [1]). These arguments all turn on various devices designed to make us feel that we must indeed express a reasoned preference between every pair of gambles. ("You must decide," etc.) But what force can these devices have when we can only conceive of the frame Θ retrospectively, and we are asked to construct entirely hypothetical preferences--preferences "that we would have had if...?"

The arguments for the Bayesian rule of conditioning, formula (1), are even more clearly tied up with the assumption that Θ is envisaged in advance. Consider, for example, the argument based on de Finetti's <u>definition</u> of conditional probability as the value of a "called-off bet". This argument runs as follows: Say we accept the ideas about ordering gambles over Θ that are set forth in the arguments for additivity and thus agree to define "degree of belief" in terms of the values of these gambles; we agree that having a degree of belief of $1/(1+x)$ in A means being indifferent towards a bet for A at odds x:1. (This is a gamble where we obtain x if $\theta \in A$ and lose 1 if $\theta \notin A$.) De Finetti ([1], p. 109) suggests that we similarly define our "conditional degree of belief $P(A|E)$" to be $1/(1+y)$, where y:1 are the odds at which we are indifferent towards a bet on A that is called off if E <u>is</u> false. (This is a gamble where we obtain y if $\theta \in A \cap E$, lose 1 if $\theta \in \bar{A} \cap E$, and obtain zero if $\theta \in \bar{E}$.) Formula (1) then follows from the properties of the total ordering on gambles. And so if we suppose that our attitude towards such a called-off bet should be the same as our attitude towards a simple bet on A after we obtain conclusive evidence for E, then (1) becomes a representation of how our beliefs are changed by such new evidence. But what is the point of all this when "E" has not even been conceived of until after it is known to be true?

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

(Notice the assumption that our attitude towards the called-off bet should be the same before and after we have obtained conclusive evidence for E. This assumption does not have any claim to self-evidence when viewed from outside the Bayesian framework. It would not be true, for example, if we evaluated gambles in terms of the "upper and lower expectations" calculated from a belief function. Both Freedman and Purves [4] and Teller [10] have given "Dutch book" arguments for the Bayesian rule of conditioning that do not depend on this assumption. But these arguments require us to have even greater prior knowledge as to what our evidence will be like. They assume that there is a partition of Θ such that E is an element of that partition and such that we know beforehand that the evidence will amount to conclusive evidence for some element of the partition.)

7. How Probable is the Evidence?

As I suggested at the end of Section 5 above, the important question about the idea of constructing, after the fact, a probability distribution over a frame that discerns the evidence is whether this idea will work--whether the required act of imagination can be carried out in detail in such a way as to organize effectively the insight and experience available for assessing the evidence. It is easy enough for me to give the name E to my evidence and to "imagine" in an abstract way that I had foreseen the possibility of E. And so it seems intelligible to ask about P(E) and $P(A \cap E)$--prior degrees of belief that I would have had, had the possibility of E occurred to me. But there is no reason to presume that my attempt to refer E to my prior experience and understanding in this way will always succeed. Whether it will succeed in any particular case will depend, surely, both on how my prior experience and understanding are organized and on how my understanding of the evidence E is organized.

In my judgment, the attempt to assess P(E) can succeed only when E is very simple or highly stylized. In moderately complicated and realistic cases, such as the matter of my house, we will be bewildered by the question of how probable our evidence itself is.

One aspect of our bewilderment is that the probability of the evidence E seems to depend on the specificity of E; the more contingent detail we count as part of E, the smaller P(E) must be. And usually we do not have a very clear idea how specific E is. Just exactly what did I see in the shallow crawl space under the oak floor? How precisely did I note the number of blocks, the size and character of each, and the distances between them? How specific is "my evidence" on these points? This seems to be a hopelessly ill-posed problem. On the one hand, the concrete facts available to my inspection were no doubt indefinitely detailed and specific. But this indefinite detail and specificity is not the evidence I actually took into account--it is merely what was there (so I hypothesize) for my senses to probe. On the other hand, this probing was not sufficiently self-reflective to permit a conscious judgment as to how much detail I did take into account in arriving at my judgments of fact and probability (my

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

judgment that there were only a few supports and that these were probably inadequate to support an outside wall). Thus there is no way for me to say how specific "my evidence" was.

It is conceivable that this puzzle about the specificity of the evidence E might not arise. Perhaps there are cases where an intuitive judgment about the probability P(E) is directly accessible to our introspection, and in such cases this intuitive judgment would itself define the degree of specificity of E. But I am unaware of any cases where Bayesians claim such direct intuitive insight. So what are we to do?

There is a great temptation, I think, to try to redefine what is meant by "our evidence E" in some way that would permit a relatively explicit understanding of its degree of specificity and an assessment of P(E) on the basis of this understanding. Usually this means taking "E" to be a more or less formal or verbal description of the evidence, picking out its relevant features and specifying explicitly the specificity with which these have been observed. (One assumes, to use the language of philosophers, that the evidence consists of "observation sentences". In plain English: that it can be put in words.) This is a trap. For there is no reason to expect that our attempts to explicitly and consciously describe the evidence will capture just those features of the evidence and just that degree of specificity that influenced our judgment.

What is evidence? The dictionary defines evidence as "something that tends to prove." And, as this definition suggests, the range of what can count as evidence is fully as broad as the English word "thing". The scene under the oak floor of my house, my neighbor's hesitant behavior as he discussed the history of my house, a library full of books, a statistician's "data", even my memory itself--all these things can count as evidence. But how and how forcefully they count as evidence depends not just on their nature in themselves but also on a complex, interactive, and contingent process of judgment. We never mentally grasp our evidence all at once--we must always explore it, or "sample" from it. And we do not usually sample "facts" or "observation sentences". The exploration of our evidence by our senses produces neurological signals, surely, but there is no reason to think that these signals have meanings so independent of their context and often contingent destinations that they can be regarded as coded "observation sentences". And though our thought may sometimes be sufficiently self-conscious to allow us to say something about what features of the evidence have inspired the suspicions, conclusions and probability judgements that result from this exploration, we can seldom articulate elementary facts which we have discerned and fully understood and which account for these final judgments. Nor could an omniscient observer of our thinking necessarily articulate such elementary facts. Such an observer's omniscience would enable him to outdo our introspection in pinpointing the features of the evidence that informed or inspired us, but this does not mean that the observer could articulate these features as "facts" or "observation sentences" without thereby noting more than we took into account or even correcting mistakes that we made.

In practice, we can usually describe fairly well the general nature of our evidence, but we cannot put into words all the details that determine its force. Consider, for example, my elderly neighbor's testimony. He hesitated and did not seem completely certain. He admits that his memory is no longer as reliable as it once was. But how hesitant was he? How uncertain? How unreliable? Perhaps I can answer these questions with a number: I can say, as I did in Section 4 above, that his testimony seems to me to merit a credence of .2. But I am at a loss to capture the grounds of this judgment in a list of "facts" that I observed about his behavior.

I conclude that the idea of constructing, after the fact, a global probability distribution that assigns a probability P(E) to the evidence E is unworkable. We have no direct intuitions about P(E). And it will often be impossible to formulate an explicit understanding of the specificity of E on which to base a judgment of P(E).

Practical Bayesians accept, I believe, the impossibility of constructing global probability distributions even after the fact. They do not try to evaluate P(E). Instead they think about other probabilities and conditional probabilities. They think about $P(A|E)$, and they think about formulae such as Bayes' theorem, which relate $P(A|E)$ to other conditional probabilities. We shall now turn to this practical version of the Bayesian theory. But as we do so we should bear in mind the lesson of the preceding argument: in the assessment of evidence, formulae such as Bayes' theorem cannot derive any legitimacy from their status as theorems about a global probability distribution, for there is no such global probability distribution--nor can one be constructed. As a tool for assessing evidence, Bayes' theorem must stand on its own; it has no more a priori justification than Dempster's rule and the other tools of the theory of belief functions.

8. The Practical Bayesian Theory

Consider the simplest form of Bayes' theorem, which is easily derived from (1):

$$P(A|E) = \frac{P(A)P(E|A)}{P(A)P(E|A) + P(\bar{A})P(E|\bar{A})} . \qquad (2)$$

The quantities on the right-hand side of (2) all have intuitive interpretations. P(A) is the probability we think the proposition A would merit on the basis of what we knew before obtaining the evidence E. $P(E|A)$ is the probability we would have assigned to the evidence E had we known A to be true and had someone put into our heads the possibility of such evidence; similarly, $P(E|\bar{A})$ is the probability we would have assigned E had we known A to be false. If these quantities were directly accessible to our intuition, then we might use (2) as a practical tool for the assessment of evidence, even while rejecting the

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

idea of a global probability distribution.

We see immediately, though, that $P(E|A)$ and $P(E|\bar{A})$ are not accessible to our intuition. These conditional probabilities depend on the specificity of E in exactly the same way as the unconditional probability $P(E)$ does, and hence raise the same problems. But (2) is easily transformed into

$$\frac{P(A|E)}{P(\bar{A}|E)} = \frac{P(A)}{P(\bar{A})} \cdot \frac{P(E|A)}{P(E|\bar{A})} , \tag{3}$$

and this formula makes it clear that we need not assess the absolute magnitude of these conditional probabilities--we need only assess their ratio. This ratio does not seem to depend on the specificity of E, so it may be meaningful to us, and even accessible to our intuition. In our example, for instance, we may be willing to say that the occurrence of the evidence E would have seemed many times more likely to us if we had known the concrete section to have been built first (A) than if we had known the oak section to have been built first ($\bar{A}$).

But how many times more likely? In our example, as in most, the complexity of the evidence E makes it difficult to think about even this question directly, and so we must decompose E and the ratio $P(E|A)/P(E|\bar{A})$ into simpler units.

A first step is to decompose E into simpler items of evidence, say $E_1, E_2, \ldots, E_k$. Recall that we are thinking of E as a subset of some global frame of discernment over which P is defined. (We have concluded that such a global frame is a fiction, but we must work within this fiction to develop our formulae.) We must similarly think of $E_1, E_2, \ldots, E_k$ as subsets, subsets such that $E = E_1 \cap E_2 \cap \ldots \cap E_k$. Then we can use the rules of the additive probability calculus to decompose $P(E|A)$ and $P(E|\bar{A})$ in terms of the E_i, obtaining

$$\frac{P(A|E)}{P(\bar{A}|E)} = \frac{P(A)}{P(\bar{A})} \cdot \frac{P(E_1|A)}{P(E_1|\bar{A})} \cdot \frac{P(E_2|A \cap E_1)}{P(E_2|\bar{A} \cap E_1} \cdots \frac{P(E_k|A \cap E_1 \cap \ldots \cap E_{k-1})}{P(E_k|A \cap E_1 \cap \ldots \cap E_{k-1})} . \tag{4}$$

If the E_i are conditionally independent given A and given $\bar{A}$ (with respect to the global probability distribution P), then this simplifies to

$$\frac{P(A|E)}{P(\bar{A}|E)} = \frac{P(A)}{P(\bar{A})} \cdot \frac{P(E_1|A)}{P(E_1|\bar{A})} \cdot \frac{P(E_2|A)}{P(E_2|\bar{A})} \cdots \frac{P(E_k|A)}{P(E_k|\bar{A})} . \tag{5}$$

Thus the problem of intuitive assessment is reduced to assessing the "prior probability" $P(A)$ and the k "likelihood ratios" $P(E_i|A)/P(E_i|\bar{A})$.

Unfortunately, this first step in decomposing the problem may not go far enough. Typically, as in the example of my house, the conditional probabilities in (4) will be rather remote from intuitive assessment, and it will not be plausible to regard the E_i as conditionally independent given A and $\bar{A}$. (In the matter of my house, E_1 and E_3, for instance, do not seem to be conditionally independent given A; under the assumption that A is true, the evidence E_1 would strengthen the case for the partition being the original outside wall and hence increase the probability of E_3.) The problem is that A and $\bar{A}$ may not specify enough.

So we make yet another decomposition; we break A into subhypotheses $H_1, H_2, \ldots, H_m$, and we break $\bar{A}$ into subhypotheses $H_{m+1}, H_{m+2}, \ldots, H_n$. In other words, we choose a partition $H_1, H_2, \ldots, H_n$ of the global frame of discernment such that $A = \cup_{r=1}^{m} H_i$ and $\bar{A} = \cup_{r=m+1}^{n} H_i$. And we try to choose this partition so that we can plausibly assess all the ratios $P(E_i|H_r)/P(E_i|H_s)$ and so that the E_i can be plausibly regarded as conditionally independent given each H_r. If we can find such a partition, then the same principles that led us to (5) yield

$$\frac{P(H_r|E)}{P(H_s|E)} = \frac{P(H_r)}{P(H_s)} \cdot \frac{P(E_1|H_r)}{P(E_1|H_s)} \cdot \frac{P(E_2|H_r)}{P(E_2|H_s)} \cdots \frac{P(E_k|H_r)}{P(E_k|H_s)} \qquad (6)$$

for all r and s between 1 and n. And once we have assessed the quantities on the right-hand sides of these equations, we can calculate the $P(H_r|E)$ and thence $P(A|E) = \Sigma_{r=1}^{m} P(H_r|E)$.

The equations (6) embody the practical version of the Bayesian theory. It is a version that must look to success in practice for its justification. I have sketched the usual derivation of (6), which depends at every step on the existence of a global probability distribution. But I wish to stress again that this global distribution is a fiction, and that the derivation is only a metaphorical justification for the equations. Their real justification must come from their fit with our intuitive understanding of evidence in practical examples.

In the global version of the Bayesian theory the assessment of evidence seems easy, almost mindless--all that is required is that we condition our global probability distribution on the evidence. But in the practical version, as embodied by (6), the assessment of evidence becomes, just as in the theory of belief functions, an art. Its tasks include not only the assessment of ratios such as those on the right-hand side of (6), but also the identification of a set of hypotheses $H_1, H_2, \ldots, H_n$ and the choice of a decomposition $E_1, E_2, \ldots, E_k$. The art

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

is to make these choices so that the ratios lend themselves to intuitive assessment, seem to capture the intuitive thrust of the evidence, and are "independent". This requires, in general, both imagination and a certain amount of plausible conjecture.

What do we mean by the requirement that the different items of evidence be "independent" on each of the hypotheses H_r? Bayesians are accustomed to answering this question in terms of the global probability distribution: independence means that a certain multiplicative relation just happens to be true of this distribution. But once we have acknowledged the fictitiousness of this global distribution, an answer in terms of our intuitive understanding of the evidence is required. The answer is, I think, that the different items of evidence must be independent in exactly the same sense as required for Dempster's rule of combination: The frame of discernment defined by $H_1,\ldots,H_n$ must discern the interaction of the evidence. That is to say, any relevant conjecture favored or disfavored by the interaction of different items of evidence must be specified by these hypotheses.

It would be wrong, then, to claim that the practical Bayesian theory has any criterion for judging "whether two items of evidence are independent" that is not available to the theory of belief functions. Users of both theories must make such judgments on the basis of their understanding of what the evidence supports and how it interacts.

There are, of course, differences between the two theories' treatments of independence. One such difference arises when we are unable to find a frame with respect to which a given item of evidence can be further decomposed into non-interacting components. In the theory of belief functions we must take this as the stopping point for decomposition: we must assess degrees of support provided by the item of evidence directly, without further decomposition. But in the practical Bayesian theory there seems to be another possibility: We may use a decomposition where there is interaction and take that interaction into account by trying to make sense of the ratios in (4). It is sometimes claimed that this possibility makes the practical Bayesian theory much more powerful than the theory of belief functions. This may be true. But to demonstrate that it is true, one would need to exhibit examples where decomposition is necessary (i.e., direct assessment is too difficult), a non-interacting decomposition cannot be obtained by a slight shift in viewpoint, and yet the ratios in (4) are intuitively meaningful. I have not seen such examples. The examples usually discussed in this connection involve either statistical models, where decomposition is not necessary, or genuine global distributions, where (4) is a theorem--i.e., where one is using genuine conditioning rather than the practical Bayesian theory. (Recall that genuine conditioning is subsummed within the theory of belief functions as a special case of Dempster's rule of combination.)

Another difference arises because of the Bayesian notion of "prior evidence". In the theory of belief functions, all the items of evidence are on the same footing: each E_i determines a belief function, and all these are combined. But in the Bayesian theory, some of the E_i are assessed by assessing likelihood ratios $P(E_i|H_r)/P(E_i|H_s)$, while others must go into assessing the prior probabilities $P(H_r)$. The choice of which evidence to regard as "prior evidence" and which evidence to assess in terms of likelihood ratios seems to be a delicate aspect of the Bayesian art.

9. Making the Bayesian Probability Judgments

I shall now give one possible Bayesian analysis of the evidence concerning my house. Since practical Bayesian analysis is an art, there can be no pretense that this analysis is the best possible Bayesian analysis of this evidence. But I hope it is a fair example of what can be done.

I have already formulated, for the analysis using belief functions, a decomposition $E_0, E_1, \ldots, E_4$ of the evidence and a frame of discernment $\Theta = \{\theta_1, \ldots, \theta_7\}$ that discerns the interaction of these items of evidence. So I am ready to try to apply (6), putting these θ's in the role of the H's written there.

First I assess my prior probabilities for the θ's. There is at least one aspect of my evidence that can conveniently be thought of as contributing to such prior beliefs: the presumption that those who enlarged the house would have retained the outside wall as a partition rather than replacing it. This is the evidence I have called E_0, and it supports $C = \{\theta_1, \theta_2, \theta_3\}$. In the analysis using belief functions, I awarded C degree of belief .8 on the basis of that evidence. But that was a conservative, non-additive judgment. Now I must speak the language of additive probabilities, where an .8 degree of belief for one thing means a .2 degree of belief for its opposite. And it is awkward to say that I have anything like a .2 prior degree of belief that the wall was replaced. So I shall instead give $C = \{\theta_1, \theta_2, \theta_3\}$ a prior degree of belief of .96, and $\overline{C} = \{\theta_4, \theta_5, \theta_6, \theta_7\}$ a prior degree of belief of .04. This is about the best compromise I can make between my desire to assess the strength of the presumption for C conservatively and my reluctance to give significant positive belief to $\overline{C}$.

I must still distribute the .96 among the three possibilities in C and the .04 among the four possibilities in $\overline{C}$. Here I face a common Bayesian conundrum--how to distribute prior belief over various hypotheses when our prior opinions amount more or less to neutrality--how, in short, to reconcile ignorance with prior probabilities. We usually resort to symmetries, confusing and conflicting as these may be. Accordingly, I divide the probability evenly in both cases, giving the prior probabilities shown in Figure 3. The resulting prior

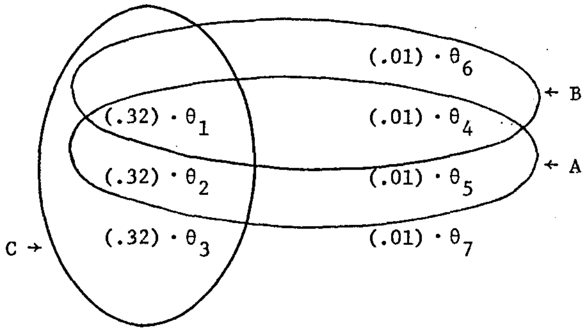

Figure 3. Prior Probabilities

* * * * *

probabilities for A and B are .66 and .34, respectively, and these figures do not reflect my initial neutrality towards these propositions. But what can I do?

To complete the Bayesian analysis, I must assess likelihood ratios $P(E_i|\theta_r)/P(E_i|\theta_s)$ for each of the remaining four items of evidence: E_1, E_2, E_3, and E_4. Let us consider each of these items in turn.

First E_1, my evidence that there is no crack between the concrete floor and the concrete under the partition. My failure to find any such crack seems much more likely under the hypotheses θ_1, θ_4, or θ_6, which declare the partition to be on the concrete floor, than under the other hypotheses. How much more likely? It seems to capture the force of the evidence to say that it is about 100 times more likely, so I set $P(E_1|\theta_r)/P(E_1|\theta_s) = 100$ for r = 1,4,6, and s = 2,3,5,7. I indicate this in Table 3 by giving $P(E_1|\theta_1)$ as 100 K_1 and $P(E_1|\theta_2)$ as K_1, etc; K_1 is a constant I need not specify.

As I say, it seems to capture the force of the evidence to set $P(E_1|\theta_r)/P(E_1|\theta_s) = 100$. But what question is really being answered by this number? The force of the evidence seems to be a matter of how certain I am that there is no division in the concrete, not just a matter of how much more likely I think it that I would not perceive such a division under the one hypothesis than under the other. Perhaps then, the question should be how much more likely I would be to fail to find a crack under the one hypothesis than the other, given that I looked as hard as I did, that I had only a narrow spot between rugs to examine, that there was just that much glue and debris in my way as there was, etc.--given, in short, all the things that influenced my judgment when I said how certain I was. But here I face the difficulty I discussed in Section 7 above; I do not have such masterful knowledge of all the

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

Table 3

Hypothesis	$P(\theta_r)$	$P(E_1\|\theta_r)$	$P(E_2\|\theta_r)$	$P(E_3\|\theta_r)$	$P(E_4\|\theta_r)$	$P(\theta_r\|E)$
θ_1	.32	100 K_1	$2K_2$	100 K_3	$1.5K_4$	.960223
θ_2	.32	K_1	$2K_2$	100 K_3	$1.5K_4$	.009602
θ_3	.32	K_1	$2K_2$	K_3	K_4	.000064
θ_4	.01	100 K_1	$2K_2$	100 K_3	$1.5K_4$	.030007
θ_5	.01	K_1	$2K_2$	K_3	$1.5K_4$	.000003
θ_6	.01	100 K_1	K_2	K_3	K_4	.000100
θ_7	.01	K_1	K_2	K_3	K_4	.000001

* * * * *

influences on my judgment that I can sit back and think about what is likely under all these influences. And so I rather feel that the number 100 is more an intuitive judgment of the strength of the evidence for $B = \{\theta_1,\theta_4,\theta_6\}$ than an honest assessment of a "likelihood ratio".

The same problem arises with my other three items of evidence. In each case I can use the "likelihood ratio" as a vehicle for expressing my judgment of the strength of the evidence and introducing this judgment into the Bayesian analysis; but when I do this, I do not feel that I am actually answering the question that the ratio ostensibly asks.

Consider E_2, my impression that the footing under the oak floor is inadequate to support a bearing wall. This supports $\{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5\}$. But how do I think about $P(E_2|\theta_1)/P(E_2|\theta_6)$, say? Should I think about *how many times less likely I think it that I should have gained this* impression had the floor in fact supported a bearing wall than had it not? I find this question convoluted and beside the point. I think it is rather likely that I would have found the footing inadequate to support a bearing wall had it never done so, and I think it unlikely that I would have found it inadequate had it done so; so if I thought in these terms, I would make $P(E_2|\theta_1)/P(E_2|\theta_6)$ rather large. But this is awkward, for I do not think of E_2 as strong evidence. It is weak because I am not confident about my own understanding of what constitutes a barely adequate footing. Can I represent this lack of confidence by lowering $P(E_2|\theta_1)$ and raising $P(E_2|\theta_6)$? This seems odd to me, but it is all I can do. So I set $P(E_2|\theta_1)/P(E_2|\theta_6) = 2$, as indicated in Table 3.

Next E_3, the stubs of 2"×4" lumber that I saw in the attic. This evidence favors $\{\theta_1,\theta_2,\theta_4\}$. In the analysis using belief functions I assessed this evidence conservatively, on the grounds that there might be some other explanation of the stubs that I had not thought of. I find it flatly impossible to be so conservative when I think in terms of likelihood ratios. For I cannot see my way to making "something I have not thought of" into a non-negligible probability for such 2" × 4" stubs being there given that they are not remnants of rafters. So I set $P(E_3|\theta_1)/P(E_3|\theta_3)$ equal to 100, as indicated in Table 3.

Finally E_4, my speculation based on my neighbor's testimony. This is very weak evidence. For what it is worth it favors $A = \{\theta_1,\theta_2,\theta_4,\theta_5\}$. Can I represent this by saying that such testimony, with all its hesitations, would be more likely on the hypothesis that the concrete section had been built first than on the hypothesis that the oak section had been built first? This seems silly to me, but it is what is required. Let us say one and a half times more likely.

So I have assessed my prior probabilities and my likelihood ratios. Now I can calculate my posterior probabilities $P(\theta_r|E)$ by the formulae

$$\frac{P(\theta_r|E)}{P(\theta_s|E)} = \frac{P(\theta_r)}{P(\theta_s)} \cdot \frac{P(E_1|\theta_r)}{P(E_1|\theta_s)} \cdot \frac{P(E_2|\theta_r)}{P(E_2|\theta_s)} \cdot \frac{P(E_3|\theta_r)}{P(E_3|\theta_s)} \cdot \frac{P(E_4|\theta_r)}{P(E_4|\theta_s)} \cdot$$

The results are shown in the last column of Table 3 and in Figure 4. Notice that θ_1 emerges with a very high probability. And the total probability for A is $P(A|E) = P(\theta_1|E) + P(\theta_2|E) + P(\theta_4|E) + P(\theta_5|E) = .999835$.

As the reader will have noted, this Bayesian analysis indicates much stronger degrees of belief for θ_1 and A than the analysis using belief functions did. Here I have obtained $P(\theta_1|E) = .96$ and $P(A|E) = .9998$, whereas the final belief function Bel in Section 4 above gave $Bel(\{\theta_1\}) = .78$ and $Bel(A) = .98$. This is mainly due to the much less

* * * * *

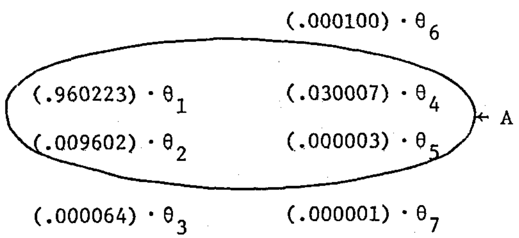

Figure 4. Posterior Probabilities

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press

conservative assessment of E_0 and E_3 in the Bayesian analysis. Another contrast is in the treatment of θ_4; we have $P(\theta_4|E) = .03$, as opposed to $Bel(\{\theta_4\}) = 0$.

I have made it clear, I am sure, that I do not find the Bayesian analysis satisfactory. I am frustrated by my inability to assess some of the probabilities conservatively, I am troubled by the arbitrariness of the prior probabilities, and I find the likelihood ratios a convoluted and awkward vehicle for expressing the force of the various items of evidence. The acuteness of these problems in this example may be more the fault of the artist than of the Bayesian art, but the problems are, I think, typical of practical Bayesian analysis.

10. Conclusion

In the course of this paper I have tried to develop three general theses:

- The practical version of the Bayesian theory must stand or fall on the basis of its ability to effectively organize practical probability judgments. It has no more a priori justification than the theory of belief functions.

- The judgment, in the practical Bayesian theory, that two items of evidence are independent conditionally on a frame Θ is essentially the same as the judgment, in the theory of belief functions, that Θ discerns the interaction of the two items of evidence.

- There are some examples, at least, where the theory of belief functions asks for more natural judgments and does a better job of organizing those judgments than the practical Bayesian theory.

I do not mean to argue that Bayesian conditioning and Bayes' theorem are never useful. After all, conditioning is a special case of Dempster's rule of conditioning. (See [8], p. 67.) But it is wrong to claim that the usefulness of conditioning is completely general. And it is misguided to defend this claim with rhetoric about "coherence".

Note

[1] I wish to thank Amos Tversky, who forced me to try harder to understand the implications of the Bayesian demand for something to condition on, Don Davis, who told me to look around in the attic, and Paul Mostert, who made me think about chain foundations. I have also benefited from conversations with Dennis Lindley, Terry Shafer, and Joe Van Zandt. My research for the paper was partially supported by allocation 3315-x038 from the General Research Fund of the University of Kansas and by grant MCS 78-01887 from the National Science Foundation.

References

[1] De Finetti, Bruno. "Foresight: Its Logical Laws, Its Subjective Sources." In Studies in Subjective Probability. Edited by Henry E. Kyburg, Jr., and Howard E. Smokler. New York: Wiley, 1964. Pages 93-158. (Originally published as "La Prévision: Ses lois logigue, ses sources subjectives." Annales de L'Institut Henri Poincaré 7(1937): 1-68.)

[2] ----------------. Theory of Probability, Volume 1. New York: Wiley, 1974.

[3] Dempster, A.P. "A Generalization of Bayesian Inference." Journal of the Royal Statistical Society, Series B 30(1968): 205-247.

[4] Freedman, David A., and Purves, Roger A. "Bayes' Methods for Bookies." Annals of Mathematical Statistics 40(1969): 1177-1186.

[5] Hacking, Ian. "Slightly More Realistic Personal Probability." Philosophy of Science 34(1967): 311-325.

[6] Lindley, Dennis. Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 1: Probability. Cambridge: Cambridge University Press, 1965.

[7] Savage, L.J. The Foundations of Statistics. New York: Wiley, 1954.

[8] Shafer, Glenn. A Mathematical Theory of Evidence. Princeton: Princeton University Press, 1976.

[9] -------------. "Non Additive Probabilities in the Work of Bernoulli and Lambert." Archive for History of Exact Sciences 19 (1978): 309-370.

[10] Teller, Paul. "Conditionalization and Observation." Synthese 26(1973): 218-258.

https://doi.org/10.1086/psaprocbienmeetp.1978.2.192482 Published online by Cambridge University Press