Bayes or Bust?
A Critical Examination of Bayesian Confirmation Theory

John Earman

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Times Roman by Asco Trade Typesetting Ltd., Hong Kong, and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Earman, John.
Bayes or bust? : a critical examination of Bayesian confirmation theory / John Earman.
p. cm.
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-05046-3
1. Science—Philosophy. 2. Science—Methodology. 3. Bayesian statistical decision theory. I. Title.
Q175.E23 1992
502.8—dc20 91-35623
CIP

For Peter Hempel, teacher extraordinaire,
And Grover Maxwell, a gentleman who left his mark on an ungentlemanly world

2 The Machinery of Modern Bayesianism

I. J. Good once quipped that there are more forms of Bayesianism than there are actual Bayesians. While the ever growing popularity of Bayesianism may have invalidated the letter of this quip, its core message is still sound: there are many rooms to the mansion that Bayes helped to build. No attempt will be made here to systematically survey all of this real estate. Bayesians of whatever persuasion can speak for themselves; indeed, they do speak for themselves, often *ad nauseam*. My focus will be kept on issues concerning the testing and confirmation of scientific hypotheses and theories, typically of a nonstatistical kind. Bayesian personalism will be the starting point for most of my investigations. Issues in Bayesian decision theory and technical issues in Bayesian statistics will be largely ignored, although from time to time technicalia will intrude.

1 The Elements of Modern Bayesianism

Bayesians of all stripes are united in the convictions that qualitative approaches to confirmation, such as hypotheticodeductivism and Hempel's instance confirmation (see chapter 3), are hopeless and that an adequate accounting of the way evidence bears on hypotheses and theories must be quantitative. The form of Bayesianism I will track here follows in Thomas Bayes's footsteps by implementing the quantitative approach in terms of degrees of belief regimented according to the principles of the probability calculus. The form of probability theory needed for applications to issues of confirmation will be presented in section 2. Bayes, as we saw in chapter 1, exploited the connection between degrees of belief and betting behavior in an attempt to justify the principles of probability. Modern Bayesians follow suit with their Dutch-book arguments, which will be examined in sections 3 and 4.

Bayesians are also united on the importance of Bayes's theorem, a result that Bayes himself never stated in modern form. If H, K, and E are respectively the hypothesis at issue, the background knowledge, and the new evidence, then one form of Bayes's theorem states that

$$\Pr(H/E \,\&\, K) = \frac{\Pr(H/K) \times \Pr(E/H \,\&\, K)}{\Pr(E/K)}. \tag{2.1}$$

If $\{H_i\}$, $i = 1, 2, \ldots$, is a set of mutually exclusive and exhaustive hypotheses, the principle of total probability allows (2.1) to be rewritten as

$$\Pr(H_i/E \,\&\, K) = \frac{\Pr(H_i/K) \times \Pr(E/H_i \,\&\, K)}{\sum_j \Pr(E/H_j \,\&\, K) \times \Pr(H_j/K)}. \tag{2.2}$$

In Bayesian accounts of confirmation, the explanations of confirmational virtues are couched largely in terms of the factors on the right hand sides of (2.1) and (2.2): $\Pr(H/K)$, the *prior probability* of H; $\Pr(E/H \,\&\, K)$, the *likelihood* of E on H and K; and $\Pr(E/K)$, the *prior likelihood* of E.

The forms of Bayesianism to be examined here also share the tenet that learning from experience is to be modeled as conditionalization. The rule of *strict conditionalization* says that if it is learned for sure that E and if E is the strongest such proposition, then the probability functions $\Pr_{\text{old}}$ and $\Pr_{\text{new}}$, representing respectively degrees of belief prior to and after acquisition of the new knowledge, are related by

$$\Pr_{\text{new}}(\cdot) = \Pr_{\text{old}}(\cdot/E). \tag{SC}$$

Bayes's proposition 5 can, as we saw in chapter 1, be regarded as an attempt to justify this rule. From the point of view of strict conditionalization, Bayes's theorem (2.1) makes explicit how the acquisition of new evidence impacts on previous degrees of belief to produce new degrees of belief.[1]

A more sophisticated form of conditionalization that allows for uncertain learning is due to Richard Jeffrey (1983b). If we observe a jelly bean by dim and flickering candle light, we will rarely come away with certain knowledge of the color of the bean, but our probabilities will have changed.[2] We may have gone, for example, from complete ignorance as to whether the bean is red, yellow, or green (as represented by probabilities of 1/3 for each) to, say, a probability of 2/3 for red and a probability of 1/6 each for yellow and green. To generalize and formalize, let $\{E_i\}$, $i = 1, 2, \ldots$, be a partition of the probability space. Intuitively, the belief change that takes place is supposed to be generated by the way in which the experience bears on this partition (e.g., the color partition in the above example). The belief change then accords with *Jeffrey conditionalization* just in case

$$\Pr_{\text{new}}(A) = \sum_i \Pr_{\text{old}}(A/E_i) \times \Pr_{\text{new}}(E_i) \quad \text{for all } A. \tag{JC}$$

Strict conditionalization is the special case where the new probability of one of the elements of the partition is one. An application of total probability shows that (JC) obtains under the condition of *rigidity*:

$$\Pr_{\text{new}}(A/E_i) = \Pr_{\text{old}}(A/E_i) \quad \text{for all } A \text{ and all } i \tag{R}$$

Arguably, (R) should apply in the jelly bean case when we look but don't get to touch, smell, or taste the bean, so that any change in our degrees of belief about the sweetness, scent, or texture of the bean should be due entirely to changes in our degrees of belief about its color.[3]

In chapter 1 we saw that Bayes's essay contained a tension between personalism (probability as personal degree of belief) and objectivism (probability as uniquely determined rational degree of belief). The tension survives in modern Bayesianism. The *pure personalists*, as represented by de Finetti and his followers, recognize the axioms of probability as the only synchronic constraints on degrees of belief. Some personalists have also refused to recognize any diachronic constraints, but it turns out that the Dutch-book arguments used to justify the probability axioms can also be used to justify rules of conditionalization (see, however, section 6 below). *Tempered personalists* would add further constraints, such as Lewis's principal principle to be discussed below in section 7, or Shimony's (1970) injunction on the members of a scientific community to assign a nonzero prior to any hypothesis seriously proposed by a fellow member of the community. *Objectivists*, such as Harold Jeffreys (1961, 1973), carry the tempering of priors to the extreme by proposing principles to uniquely fix these numbers. Thomas Bayes himself seems to have fallen into this camp, at least with respect to the problem treated in his founding essay.

The implications of these differing forms of Bayesianism for confirmation theory will be discussed in later chapters. The present chapter concentrates on an elementary exposition of the common core of all forms of Bayesian personalism.

2 The Probability Axioms

Since propositions are the object of belief and since probability is being interpreted as degree of belief, probabilities will be assigned to objects that express propositions, namely sentences. More specifically, let $\mathscr{A}$ be a collection of sentences. The content and structure of $\mathscr{A}$ will vary from context to context, but at a minimum it is assumed that $\mathscr{A}$ is closed under finite truth-functional combinations. Then a *probability function* Pr is a map from $\mathscr{A}$ to $\mathbb{R}$ satisfying at least the following restrictions:

$$\Pr(A) \geqslant 0 \quad \text{for any } A \in \mathscr{A} \tag{A1}$$

$$\Pr(A) = 1 \quad \text{if} \models A \tag{A2}$$

$$\Pr(A \vee B) = \Pr(A) + \Pr(B) \quad \text{if} \models \neg(A \,\&\, B) \tag{A3}$$

Here $\models A$ means that A is valid in the sense that A is true in all models or all possible worlds.[4] Again, the content and structure of the models or possible worlds will depend upon the context. I assume at a minimum that $\mathscr{A}$ respects propositional logic.[5] In this case (A1) to (A3) suffice to prove many of the familiar principles of probability, including the following:

$$\Pr(\neg A) = 1 - \Pr(A) \tag{P1}$$

$$\Pr(A) = \Pr(B) \quad \text{if} \models A \leftrightarrow B \tag{P2}$$

$$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \,\&\, B) \tag{P3}$$

$$\Pr(A) \leqslant \Pr(B) \quad \text{if } A \models B \tag{P4}$$

Here $A \models B$ means that A semantically implies B in the sense that B is true in every model or possible world in which A is true.

Conditional probability may be introduced as a defined concept:

Definition If $\Pr(B) \neq 0$, then $\Pr(A/B) \equiv \Pr(A \,\&\, B)/\Pr(B)$.

Bayes's theorem is now a simple consequence of this definition. An alternative approach takes conditional probability $\Pr(\cdot/\cdot)$ as primitive and defines the associated unconditional probability $\Pr(\cdot)$ as $\Pr(\cdot/N)$, where N is a necessary truth (i.e., $\models N$). The advantage of this approach is that $\Pr(A/B)$ can be defined even when $\Pr(B) = 0$. Conditional probability is discussed in more detail in appendix 1 to this chapter.

Some of the applications to be considered in later chapters also assume a principle of continuity.

C If $A_i \in \mathscr{A}$, $i = 1, 2, \ldots$, are such that $A_{n+1} \models A_n$ for each n and $\{A_1, A_2, \ldots\}$ is inconsistent (i.e., the A_i are not all true in any model or possible world), then $\lim_{n\to\infty} \Pr(A_n) = 0$.

Actually, the axiom I will use most often is a weaker principle that applies to first-order predicate logic. Let 'P' be a monadic predicate and let $a_1, a_2, \ldots$ be a countably infinite sequence of individual constants. The principle added as an additional axiom asserts that

$$\Pr((\forall i)Pa_i) = \lim_{n\to\infty} \Pr\left(\mathop{\&}_{i \leqslant n} Pa_i\right), \tag{A4}$$

where $\&_{i \leqslant n} Pa_i$ stands for $Pa_1 \,\&\, Pa_2 \,\&\, \ldots \,\&\, Pa_n$. If we require that $(\forall i)Pa_i \models Pa_n$ for every n and that $\{\neg(\forall i)Pa_i, Pa_1, Pa_2, \ldots\}$ be inconsistent, then (A4) is shown to be a consequence of (C) by one's taking $A_n \equiv (\&_{i \leqslant n} Pa_i \,\&\, \neg(\forall i)Pa_i)$. It also follows that $\Pr((\exists i)Pa_i) = \lim_{n\to\infty} \Pr(\bigvee_{i \leqslant n} Pa_i)$, where $\bigvee_{i \leqslant n} Pa_i$ stands for $Pa_1 \vee Pa_2 \vee \ldots \vee Pa_n$. Axiom (A4) can be regarded as an extension of the finite additivity principles (A3) and (P3) to countable additivity.

In a manner of speaking, "half" of (A4) is already a consequence of (A1) through (A3). Since $(\forall i)Pa_i \models \&_{i \leqslant n} Pa_i$, it follows by (P4) that $\Pr((\forall i)Pa_i) \leqslant \Pr(\&_{i \leqslant n} Pa_i)$. Moreover, $\Pr(Pa_1)$, $\Pr(Pa_1 \,\&\, Pa_2)$, $\Pr(Pa_1 \,\&\, Pa_2 \,\&\, Pa_3), \ldots$ is a monotone decreasing sequence bounded from below (by (A1)), and so it must have a limit. Thus $\Pr((\forall i)Pa_i) \leqslant \lim_{n\to\infty} \Pr(\&_{i \leqslant n} Pa_i)$. To turn the '$\leqslant$', into an '$=$', as required by (A4), requires a new substantive assumption.

Continuity or countable additivity does not come without intuitive cost. Consider a denumerably infinite list $H_1, H_2, \ldots$ of pairwise incompatible and mutually exhaustive hypotheses. One might think that it should at least be possible to treat these hypotheses in an evenhanded manner by assigning them all the same probability. But this we cannot do consistently with (C), since (C) implies that $\sum_{i=1}^{\infty} \Pr(H_i) = 1$. Continuity thus forces us to play favorites. (Sticking to finite additivity would allow for a draconian evenhandedness in the form $\Pr(H_i) = 0$ for all i.) On the other hand, abandoning countable additivity leads to results that Bayesians and non-Bayesians alike find repugnant. Some of these results will be discussed in appendix 1.

A different nomenclature is presupposed when mathematicians and statisticians speak of probability. For them, a probability space is a triple $(\Omega, \mathscr{F}, \mathscr{P}r)$. Ω, a set of elements, is called the sample space; $\mathscr{F}$, a field of subsets of Ω, is the collection of measurable sets; and $\mathscr{P}r$ is a nonnegative (finitely or countably additive) function from $\mathscr{F}$ to $\mathbb{R}$. (Here countable additivity means that if $B_i \in \mathscr{F}$, $i = 1, 2, \ldots$ are pairwise disjoint, then $\mathscr{P}r(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mathscr{P}r(B_i)$.) As is discussed in detail in chapter 6, one can move from the Bayesian personalist conception of probability to the mathematical conception by taking Ω to be the set of models of the language of $\mathscr{A}$, $\mathscr{F}$ to be a field generated by sets of models of the form mod(A) for a sentence $A \in \mathscr{A}$, and $\mathscr{P}\!r$ to be a measure satisfying $\mathscr{P}\!r(A) = \mathscr{P}\!r(\mathrm{mod}(A))$. One can also move in the opposite direction, although an awkwardness occurs when $\mathscr{F}$ is a σ field (see appendix 2) and probabilities qua degrees of belief are assigned to sentences in a standard first-order language, for then not every member of $\mathscr{F}$ will correspond to a sentence, since these languages do not allow infinite conjunctions or disjunctions. We can often smooth over this awkwardness by taking limits of probabilities of finite conjunctions or disjunctions.

3 Dutch Book and the Axioms of Probability

Rather than simply assuming that degrees of belief are regimented by the principles of probability, one could try to exploit the interpretation of probability as degree of belief as a means of getting a justification for the probability axioms. We saw in chapter 1 that Thomas Bayes took this tack by using the connection between degrees of belief and betting behavior. Ramsey (1931) and de Finetti (1937) followed a related tack with their Dutch-book strategy, although they were apparently unaware of the details of Bayes's work, which contains, as we have seen in chapter 1, intimations of Dutch book. The presentation given here follows Shimony 1955.

By a *bet* on $A \in \mathscr{A}$ let us understand a contractual arrangement between a bettor and a bookie by which the bettor agrees to pay the bookie the amount $\$b$ if A turns out to be false and the bookie agrees to pay the bettor $\$a$ if A turns out to be true. The sum $\$(a + b)$ is called the *stakes* of the bet, and the ratio b/a is called the bettor's *odds*. If Pr is the bettor's degree-of-belief function, the expected monetary value of the bet for him is $\$a \times \Pr(A) - \$b \times \Pr(\neg A)$. The bet is said to be *fair* (respectively, *favorable*, *unfavorable*) to the bettor according as the expected value is zero (respectively, positive, negative). Using the negation principle (P1), the condition for a fair bet comes to $\Pr(A) = b/(a + b)$. This ratio is called the bettor's *fair betting quotient*.

The idea of the Dutch-book argument is to turn this construction around to produce a justification of the probability axioms: assume that degree of belief functions as a fair betting quotient and then show that something very nasty will happen if the degrees of belief fail to conform to the probability axioms. Thus if $\Pr(A) = r$ is your degree of belief in A, then (the story goes) you should be willing to bet on A on the terms in table 2.1. S is allowed to be either positive or negative, which means that you are

Table 2.1
Terms for betting on A

	Pay	Collect	Net
A false	rS	0	$-rS$
A true	rS	S	$(1 - r)S$

Note: S stands for the stakes.

required to accept either end of the bet. If you do enter such an arrangement, the nasty thing that threatens is Dutch book, a finite series of bets such that no matter what happens, your net is negative (a violation of what is called *coherence* for degrees of belief). The *Dutch-book theorem* shows that if any one of the axioms (A1) to (A3) is violated, then Dutch book can be made. The *converse Dutch-book theorem* shows that if (A1) through (A3) are satisfied, then Dutch book cannot be made in a finite series of bets. This converse is crucial to the motivation for conforming degrees of belief to the principles of probability, for if such a conformity were not guarantee against Dutch book, the threat of Dutch book would not be a very effective inducement to conformity. Only the proof of the Dutch-book theorem will be sketched here. The interested reader can consult Kemeny 1955 and Lehman 1955 for the converse.

To establish that (A1) is necessary to avoid Dutch book, suppose that $\Pr(A) = r < 0$. Choose $S < 0$ and note that the net is negative whether or not A is true. Similarly, if $\Pr(A) = r > 1$, choosing $S > 0$ leads to a loss, come what may. We can now establish that (A2) is necessary to avoid Dutch book. For suppose that $\Pr(A) = r \neq 1$ even though $\models A$. By the previous results, $0 \leqslant r \leqslant 1$. Choosing $S < 0$ then leads to a loss in case A is true, which is the only possible case. Finally, to show the necessity of (A3), suppose that $\models \neg(A \,\&\, B)$ and consider a series of three bets: one on A with a betting quotient $\Pr(A) = r_1$ at stakes S_1, one on B with a betting quotient $\Pr(B) = r_2$ at stakes S_2, and one on $A \vee B$ with a betting quotient $\Pr(A \vee B) = r_3$ at stakes S_3. There are three possible cases to consider (table 2.2). The theory of linear equations then shows that the stakes can be chosen so that the nets are all negative unless $r_3 = r_1 + r_2$, i.e., unless (A3) holds.

If regarded as a definition, the formula given in section 1 for conditional probability does not stand in need of a justification. But as in de Finetti 1937, the notion of the conditional probability of B on A can be introduced

Table 2.2
Net payoffs for the three bets taken together

	Net
A true, *B* false	$(1 - r_1)S_1 - r_2S_2 + (1 - r_3)S_3$
A false, *B* true	$-r_1S_1 + (1 - r_2)S_2 + (1 - r_3)S_3$
A false, *B* false	$-r_1S_1 - r_2S_2 - r_3S_3$

as a primitive and then operationalized in terms of a bet on *B* conditional on *A*, the terms of which specify that if *A* obtains, a standard unconditional bet on *B* is in effect, whereas if *A* fails, the bet is called off. Then (the story goes) Pr(*B*/*A*) should be the agent's critical odds for this conditional bet. The agent is now offered three bets: a standard bet on *A*, a standard bet on *B* & *A*, and a bet on *B* conditional on *A*. It is left as an exercise to show that unless $\Pr(B/A) \times \Pr(A) = \Pr(B \,\&\, A)$, stakes can be chosen for the three bets so that the agent has a sure net loss. This argument does *not* justify the rule of conditionalization, which requires a different argument (see section 5 below).

The Dutch-book justification for continuity is not so pretty, and this is perhaps one of the reasons it plays no role in the Bayesianism of Ramsey, de Finetti, and Savage.[6] To Dutch-book a violation of (C) or (A4), which is not also a violation of (A1) through (A3), requires laying an infinite series of bets. But if I were to risk the same finite amount, no matter how small, on each of these bets, then I would have to have an infinite bankroll, an impossible dream. And if the dream should come true, I would not care one whit about losing a finite or even an infinite sum if, as can always be arranged, I have an infinite amount left over. To remedy this defect, we can imagine that the bettor accepts an infinite series of fair bets but that the total amount he risks is finite; e.g., he risks \$(1/2) on the first bet, \$(1/4) on the second, \$(1/8) on the third, and so on. With this setup, Adams (1961) shows that a sure loss results from a violation of the general continuity axiom (C) (see also Spielman 1977).

4 Difficulties with the Dutch-Book Argument

Qualms about the Dutch-book justification of the probability axioms are so numerous and diverse that it is hard to classify them. For future reference I note that when the requirement of logical omniscience is dropped, as it must be for realistic agents, the situation becomes more complicated; this matter is discussed in chapter 5. For the present context, which takes logical omniscience for granted, I begin with three miscellaneous qualms. First, the Dutch-book construction for countable additivity involves, in Ernest Adams's words, "extremely unrealistic systems" (1961, p. 8). For those who insist that degrees of belief must be operationalized in terms of economic transactions, this constitutes a reason to reject countable additivity. (Thus it is not surprising that countable additivity plays no role in de Finetti's personalism.) But for those of us who reject operationalism and behaviorism and insist that countable additivity is needed, the difficulty is a shortcoming of the Dutch-book construction. Second, the requirement that the agent be willing to take either side of the bet (i.e., the stakes *S* may be either positive or negative) may not be satisfied by actual gamblers, and in any case it already assumes the negation principle.[7] Third, a Bayesianism that appeals to both Dutch book and strict conditionalization is on a collision course with itself. The use of strict conditionalization leads to situations where $\Pr(A) = 1$ although $\not\models A$. As a result, something almost as bad as Dutch book befalls the conditionalizer; namely, she is committed to betting on the contingent proposition *A* at maximal odds, which means that in no possible outcome can she have a positive gain and in some possible outcome she has a loss (a violation of what is called *strict coherence*). It is too facile to say in response that this is a good reason for abandoning strict conditionalization in favor of Jeffrey conditionalization or some other rule for belief change; for all the results about merger of opinion and convergence to certainty so highly touted in the Bayesian literature depend on strict conditionalization (see chapter 6).

A more basic worry harkens back to Bayes's insistence that probability as a betting quotient be attached to "events," i.e., decidable propositions (see chapter 1). Bets on the outcome of the Kentucky Derby are one thing, bets on scientific hypotheses are quite another. A hypothesis with the quantifier structure $(\exists x)(\forall y)Rxy$ can be neither verified nor falsified by finite means. Thus a bet on such a hypothesis turns on a contingency that can never be known for certainty to hold or to fail, and so the parties to the bet have no sure way to settle the matter. To try to settle the bet by appeal to the probable truth or falsity of the hypothesis runs afoul of the fact that the parties can and often do disagree on whether the hypothesis is probably true. But if the bet is never paid off, fear of being bilked disappears.

The response to this worry might be that bookies wearing wooden shoes, money pumps, etc. are just window dressing. The underlying assumption is that degrees of belief are manifested in preferences over the kinds of bets described in section 3. This assumption granted, the Dutch-book construction stripped of its decoration shows that the failure of degrees of belief to conform to the probability calculus results in a structural inconsistency in the individual's preferences. Suppose that the individual is nonsatiated in that she prefers more money to less. Then if this person violates (A1) or (A2), the Dutch-book construction reveals that she is literally inconsistent with herself, since she prefers the certainty of handing over some $\$\varepsilon > 0$ to the status quo, despite her professed nonsatiation. In the case of (A3) the argument is more involved, since it appeals to another principle, "the package principle"; to wit, a person's preferences are inconsistent if there is a finite series of bets such that she regards each as preferable to the status quo while at the same time she regards the status quo as preferable to the package of bets. If this hypothetical agent violates (A3), we proceed to construct a finite series of bets each of which she finds favorable. By the package principle, she should then find the package favorable. But the package is shown to be equivalent to handing over $\$\varepsilon > 0$, which contradicts nonsatiation. Note that on this reading the Dutch-book construction does not justify strict coherence, i.e., the requirement that $\Pr(A) = 1$ only if $\models A$, which I take to be a mark in favor of this reading.

Schick (1986) has questioned the normative status of the package principle. Its plausibility, he argues, rests on accepting the notion of value additivity, which holds that the value of the package of bets is the sum of the values of the individual bets. But, Schick claims, an agent who refuses to conform her degrees of belief to the probability axioms may read the Dutch-book construction as a reason to reject value additivity. Schick's objection may not at first seem very moving, but it gains force in the context of the sequential decision making that comes into play in the attempted diachronic Dutch-book justification for conditionalization (see section 6).

Although the above reconstrual of the Dutch-book construction is a step forward, it is still too closely tied to the behavioristic identification of belief with dispositions to place bets. Once it is admitted that betting behavior is only indicative of, and not constitutive of, underlying belief states, it must also be admitted that belief and behavior are mediated by many factors and that these factors can weaken to the breaking point the simpleminded linkage assumed in the Dutch-book construction. In poker, for example, betting high may be a good way to scare off the other players and win the pot (see Borel 1924). And generally, a knowledge of the tendencies of opponents may make it advisable to post odds that differ from one's true probabilities (see Adams and Rosenkrantz 1980).[8]

Two responses can be made to this complaint. First, one can drop the Dutch-book approach in favor of a justification of the probability axioms that focuses directly on the nature of belief and the cognitive aims of inquiry and eschews altogether preferences for goodies, monetary or otherwise. Some candidates for such a justification will be examined in the next section. Second, one can continue to push the Dutch-book approach by taking into account in a more systematic manner the preference structure of the agent. I will follow this theme in the remainder of this section.

The opening melody of this theme is that the Dutch-book construction rests on the assumption that utility is linear with money, or equivalently, that agents are risk neutral, an assumption known to be false for many if not most real-world agents.[9] To illustrate the complications that can arise in trying to use betting behavior to elicit degrees of belief for such real-world agents, let us analyze from the point of view of expected-utility theory the elicitation device Bayes himself used. Let $\$q$ be the maximum amount the agent is willing to pay for a contract that awards $\$r$ if A is true and $0 otherwise. If U is the agent's utility function and $\Pr(dw/A)$ and $\Pr(dw/\neg A)$ are the agent's conditional probability distributions for wealth exclusive of the contract prize, then a little algebra shows that the expected-utility hypothesis implies that the agent's degree of belief in A is

$$1\Big/\left[1+\frac{\int (U(w+r-q)-U(w))\Pr(dw/A)}{\int (U(w)-U(w-q))\Pr(dw/\neg A)}\right]$$

(see Kadane and Winkler 1987). If the agent is risk neutral, i.e., if U is linear, then the degree of belief is seen to be equal to q/r, as Bayes thought. If $\Pr(dw/A) = \Pr(dw/\neg A)$ (i.e., the agent's wealth apart from the contract payoff is not probabilistically dependent on A) but the agent is not risk neutral, then $\Pr(A)$ will differ from q/r: if the agent is risk-averse, q/r will understate $\Pr(A)$, while if she is risk-positive, q/r will overstate $\Pr(A)$. And if $\Pr(dw/A) \neq \Pr(dw/\neg A)$, q/r is an even more distorted measure of $\Pr(A)$.

The moral is that the direct elicitation of degrees of belief by betting behavior is doomed to failure. Degrees of belief and utilities have to be elicited in concert. In the standard developments of this concerted elicitation the aim is to show that preferences satisfying (what are taken to be) rationality constraints can be represented in terms of expected utility, with the probabilities being uniquely determined and the utilities determined up to positive linear transformations. But the alleged rationality constraints are open to challenge (see, for example, the paradoxes in Allais 1953 and Ellsberg 1961). Moreover, when the utilities are dependent not just on the prizes but also on the propositions whose utilities are being elicited, then the probabilities may not be uniquely determined (see Schervish, Seidenfeld, and Kadane 1990 and Seidenfeld, Schervish, and Kadane 1990). Here I must break off the discussion, since I have strayed beyond the scope of this work.

5 Non-Dutch-Book Justifications of the Probability Axioms

Aside from a fear of being bilked by Dutch bookies, there are a number of other motivations for conforming degrees of belief to the probability calculus, three of which will be mentioned here.

The first is articulated by Rosenkrantz (1981), who follows de Finetti (1972). Consider a partition $\{H_i\}$, $i = 1, 2, \ldots, N$, and an agent who distributes her degrees of belief x_i over the H_i in accord with the constraint that $0 \leqslant x_i \leqslant 1$ but not necessarily obeying the condition $\sum_i x_i = 1$, as would be the case if she obeyed the probability calculus. Suppose that when H_j is the true hypothesis, the inaccuracy of her degrees of belief is measured by the least-squares function

$$I(x; H_j) \equiv x_1^2 + \cdots + x_{j-1}^2 + (1 - x_j)^2 + x_{j+1}^2 + \cdots + x_N^2 \tag{2.3}$$

If the x_i do not sum to 1, there is an alternative set of degrees of belief y_i that do sum to 1 and that dominate the x_i in the sense that $I(y; H_j) < I(x; H_j)$, whatever the value of j. This conclusion continues to hold when (2.3) is generalized to a weighted least-squares measure where the weights reflect judgments of how far the false alternatives are from the true hypothesis. If the conclusion could be further generalized to any "reasonable" measure of inaccuracy, we would be entitled to draw the moral that failure to obey the axioms of probability undermines the goal of accuracy. A discussion of what conditions constitute a reasonable measure of inaccuracy, together with a review of results and conjectures about the sought after generalization, are found in Rosenkrantz 1981 (see also Lindley 1982).

A second kind of justification is best construed as directed at well-tempered personalists who aim at rational degrees of belief. It can be found in various versions in Aczél 1966; Cox 1946, 1961; Good 1950; and also in Shimony 1970, the version I will report here. It works on the concept of conditional probability. Let $\Pr(H/E)$ be a real-valued function defined on pairs of sentences (H, E), where H is a member of a nonempty set $\mathscr{A}$ of sentences closed under truth functional operations and E is a member of the noncontradictory elements $\mathscr{B}^0$ of $\mathscr{B} \subseteq \mathscr{A}$ (see appendix 1). It is further supposed that $\Pr(\cdot/\cdot)$ satisfies the following six conditions:

C1 $\Pr(H/E) = \Pr(H'/E')$ if $\models H \leftrightarrow H'$ and $\models E \leftrightarrow E'$.

C2 For any $E \in \mathscr{B}^0$, there is an r_0 such that for any contradiction $C \in \mathscr{A}$ and any $H \in \mathscr{A}$, $\Pr(C/E) = r_0 \leqslant \Pr(H/E)$.

C3 There is an r_1 such that for all E, $F \in \mathscr{B}^0$, $\Pr(E/E) = \Pr(F/F) = r_1 > r_0$.

C4 $\Pr(H \,\&\, E/E) = \Pr(H/E)$.

C5 For any E, $F \in \mathscr{B}^0$, there is a function f_E such that $\Pr(H \,\&\, F/E) = f_E(\Pr(H/F \,\&\, E), \Pr(F/E))$.

C6 For any $E \in \mathscr{B}^0$, there is a continuous and monotone increasing function g_E in both variables such that if $E \models \neg(H \,\&\, J)$, then $\Pr(H \vee J/E) = g_E(\Pr(H/E), \Pr(J/E))$.

Then there exists a continuous and monotone increasing function h such that $h(r_0) = 0$, $h(r_1) = 1$, and $\widehat{\Pr}(H/E) = h(\Pr(H/E))$ satisfies the standard axioms for conditional probability.

The usefulness of this technical result for the justification of the probability axioms depends on the persuasiveness of two further assumptions: first, that (C1) through (C6) should be satisfied for any rational conditional degree of belief function and, second, that if Pr is a suitable measure of rational degree of belief, then so is any monotone function of Pr, which leaves us free to choose a Pr that satisfies the standard axioms. Neither of these assumptions recommends itself with overwhelming force.

A third mode of justification starts from Carnap's (1950) remark that rational degrees of belief can, in some instances, be construed as estimates of relative frequencies. Thus if H is of the form 'Pa', my degree of belief in H may be interpreted as my estimate of the relative frequency of individuals with the property designated by 'P' in some appropriate reference class.[10] If my personal probabilities for propositions of this form are not to be precluded a priori from being accurate estimates of frequencies, they must fulfill the standard probability axioms, since frequencies do (see van Fraassen 1983a and Shimony 1988).

Although attractively straightforward, such frequency-driven justifications have their limitations. As a result of calculation or of consulting theories like quantum mechanics, my degree of belief in H may be an irrational number. If 'frequency' means finite frequency, i.e., the ratio of the number of individuals that have the property to the total number of individuals in the (finite) reference class, then I am automatically precluded from having an exactly accurate estimate. Limiting relative frequencies in infinite sequences do not share this shortcoming, but such frequencies can lead via the continuity axiom to a conflict with other probability assignments we may want to make. Thus, for example, my estimate of the limiting relative frequency for events such as Pa_i may be 0 for each i, in which case I set $\Pr(\bigvee_{i \leqslant n} Pa_i) = 0$ for every n. But at the same time I may be convinced that at least one of the individuals must be a 'P', which contradicts (A4). More generally, for the multiply quantified hypotheses encountered in the advanced sciences, there is no obvious or natural way in which one's degree of belief can be regarded as an estimate of relative frequency in either the finite or limiting sense. Of course, I can calibrate my degree of belief in H with frequencies by finding an H' such that $\Pr(H) = \Pr(H')$ and such that $\Pr(H')$ does have a natural interpretation as an estimate of a frequency. But without further restrictions, there is no guarantee that the probabilities assigned to the class of hypotheses so calibrated will satisfy the probability axioms.

Although Dutch book and the other methods of justification investigated in this section are all subject to limitations and objections, collectively they provide powerful persuasion for conforming degrees of belief to the probability calculus.

6 Justifications for Conditionalization

Dutch-book justifications can be given for both strict conditionalization (Teller 1973, 1976) and Jeffrey conditionalization (Skyrms 1987).[11] To consider the former, suppose without any real loss of generality that upon learning E the agent shifts from $\Pr_{\text{old}}$ to $\Pr_{\text{new}}$, where $y = \Pr_{\text{old}}(A/E) - \Pr_{\text{new}}(A) > 0$ and $x = \Pr_{\text{old}}(A/E) > 0$. The diachronic Dutch bookie first sells the agent three bets b_1: [\$1; A & E], b_2: [\$$x$; $\neg E$], and b_3: [\$$y$; E], at what the agent computes to be their fair values. (Recall that [\$$z$; C] stands for the contract that pays \$$z$ if C obtains and \$0 otherwise.) If E proves to be false, the agent has a net loss of \$$y\Pr_{\text{old}}(E)$. On the other hand, if E turns out to be true, the bookie buys back from the agent the bet b_4: [\$1; A] for its then expected value to the agent (\$$\Pr_{\text{new}}(A)$ = \$$(\Pr_{\text{old}}(A/E) - y)$). The agent then has a net loss of \$$y\Pr_{\text{old}}(E)$, regardless of whether A obtains.

We can assess this argument for conditionalization in the light of the distinction drawn above in section 4 between two readings of the Dutch-book construction. If the central concern is to escape being systematically bilked by a bookie, there is a simple solution that doesn't commit you to conditionalization: don't publicly announce your strategy for changing belief in the face of new evidence. If you are worried about clairvoyant bookies who can read your mind, then don't make up your mind in advance; just wait to see what evidence comes in and then wing it. (This is, in fact, what many of us do in practice.) This will make you proof against systematic bilking, save by those bookies who have the ability to foresee your future belief states. But from such precognitive bookies not even good Bayesian conditionalizers are safe. Of course, if you do not conditionalize, there will be a hypothetical lucky bookie who by chance rather than system hits on a series of bets that guarantees you a net loss, but then even if you do conditionalize, there will be a hypothetical lucky bookie who takes you for a loss.

On the more pristine reading of the original synchronic Dutch-book construction, the bookies in wooden shoes were only window dressing, and what was really being revealed (so the story went) was a structural inconsistency in the preferences of an agent who did not conform her degrees of belief to the probability calculus. In applying this reading to the diachronic setting, we need to divide cases. Consider first the case of an agent who eschews preset rules for changing degrees of belief. In this instance it is hard to see how the charge of inconsistency can legitimately be leveled. For how can such an agent's preferences over bets at t_1 be inconsistent with her preferences over bets at t_2 any more than her preferences over wines at t_1 can be inconsistent with her preferences over wines at t_2? Perhaps in response it will be urged that without melding together preferences at different times to form an integrated whole, it wouldn't be proper to speak of an enduring agent. That is certainly true, but surely the requirements for personal identity over time cannot be taken to entail rationality constraints—and conditionalization is allegedly such a constraint—since a person who behaves irrationally does not cease to be a person.

The agent who has adopted a rule for belief change is more open to the charge of inconsistency, since she has already committed herself at t_1 to what her preferences over bets will be at t_2. It would then seem that we can apply at t_1 the package principle introduced in the discussion of synchronic Dutch book: if an agent prefers each of a finite series of bets to the status quo, then she also prefers the package of bets to the status quo. To make this principle yield the desired consequence in the present setting, 'prefer' must be taken to mean prefer when the decision is viewed as an isolated one, which is the tacit understanding in effect when the critical odds for a bet on A are used to elicit the agent's degree of belief in A. But an agent who is not a conditionalizer can satisfy the package principle by taking 'prefer' to mean prefer when the decision to accept or reject the bet is placed in the context of a sequential decision problem. If we view the diachronic Dutch-book construction as a sequential decision process, the decision tree looks as in figure 2.1. The principles of rational decision making require that at decision node 1 the agent face up to what she knows about what her preferences will be at node 2, should she get there (see Seidenfeld 1988). She knows that at node 2 the tiniest premium will lead her to prefer to sell back to the bookie the bet on A, and she sees that in

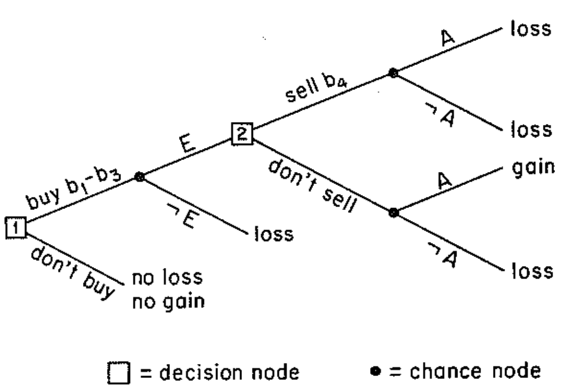

Figure 2.1
Diachronic Dutch book on a decision tree

the decision context this choice leads to a sure loss. She sees also that she gets to node 2 if at node 1 she chooses to buy b_1 to b_3 and E obtains, and further that if she chooses to buy b_1 to b_3 and E fails, she incurs a sure net loss. Thus, all things considered, she sees that buying b_1 to b_3 is unfavorable. It is on just these grounds that Maher (1992) maintains that the diachronic-Dutch-book argument is fallacious (see also Levi 1987).

To the extent that these decision-theoretic considerations are effective in undermining the diachronic-Dutch-book justification for conditionalization, they also bring into question the Dutch-book justification for the axioms of probability. In essence, the decision-theoretic message is to look before you leap. Such advice is just as valid in the synchronic setting as in the diachronic or multitemporal setting. And in the former setting, the advice clashes with the package principle needed in the argument for the principle of additivity of the probabilities of exclusive alternatives, which brings us full circle back to Schick's (1986) objection to Dutch-book arguments. The circle leaves me in an unsettled position. I agree, for example, that if I adopted a rule of belief change other than conditionalization and if I were cagey enough to draw up the decision tree for diachronic Dutch book, then I would refuse to accept the initial bets. But since I regard each of these bets as fair, should I not therefore recognize that there is something amiss in my opinion/preference structure? Grounds for a definitive answer do not exist, or if they do, I do not know of them.

A different and more modest justification for conditionalization has been given by Teller (1976), who argues that there are specifiable circumstances under which it can be maintained that if any change in belief is reasonable, then such a change must be via conditionalization. To identify some of these circumstances, Teller proves the following formal result (see also Teller and Fine 1975). Suppose that $Pr_{old}(E) > 0$ and that the agent's domain $\mathscr{A}$ of beliefs is *full* in the sense that for any number q and any $A \in \mathscr{A}$ such that $Pr_{old}(A) = r$ and $0 \leqslant q \leqslant r$ there is a $B \in \mathscr{A}$ such that $B \models A$ and $Pr_{old}(B) = q$. Suppose further that $Pr_{new}(\cdot)$ is such that $Pr_{new}(E) = 1$ and that for all $A, B \in \mathscr{A}$ such that $A \models E$ and $B \models E$, if $Pr_{old}(A) = Pr_{old}(B)$, then $Pr_{new}(A) = Pr_{new}(B)$. Then $Pr_{new}(\cdot) = Pr_{old}(\cdot/E)$.

As can easily be verified under the assumption that $Pr_{old}(E) = 0$, Teller's crucial condition $C(E)$ is equivalent to $C'(E)$:

$C(E)$ For all $A, B \in \mathscr{A}$ such that $A \models E$ and $B \models E$, if $Pr_{old}(A) = Pr_{old}(B)$, then $Pr_{new}(A) = Pr_{new}(B)$.

$C'(E)$ For all $A, B \in \mathscr{A}$ (whether or not they entail E), if $\Pr_{old}(A/E) = \Pr_{old}(B/E)$, then $\Pr_{new}(A) = \Pr_{new}(B)$.

There are clear cases where we want to impose $C(E)$ or $C'(E)$ for at least some A and B. Thus, let A be the proposition that Dancer will win the Derby, B the proposition that Prancer will win the Derby, and E the proposition that one or the other has won. Suppose that an agent is initially equally confident of A and B. She now learns precisely that E—that and no more. It would seem that, in accord with $C(E)$, she would be unreasonable in these circumstances to adjust her degrees of belief so that Dancer is now preferred to Prancer (or vice versa). But to invoke the formal result, we need to extend the argument to all pairs of initially equally probable propositions entailing E. It is hard to see how this can be done for any particular $\mathscr{A}$ that is sufficiently rich without using reasoning that would apply equally to any $\mathscr{A}$ and would thus abandon the modesty of the approach.

The basis for an immodest justification can perhaps be found in van Fraassen's (1989) result that under the assumption of the fullness of $\mathscr{A}$, $C(E)$ is implied by the requirement that the new probability of any proposition $A \in \mathscr{A}$ is a function solely of the evidence E and the old probability of A. It is well to note, however, that van Fraassen himself would not take such a justification to imply that conditionalization is necessary for rationality, since in his view rationality does not require that belief change follows a preset rule (see van Fraassen 1989 and 1990).

A different motivation for Jeffrey conditionalization starts from the idea that one should make as small a change as possible in one's overall system of beliefs compatible with the shift in those beliefs directly affected by the learning experience. Consider a probability function Pr on $\mathscr{A}$, thought of as giving the probabilities prior to making an observation. Let $\{E_i\}$ be a partition, intended as the locus of belief change, and let Pr* be a measure on $\{E_i\}$ such that $\Pr^*(E_i) > 0$ and $\sum_i \Pr^*(E_i) = 1$, intended to give the new probabilities of the E_i after observation. One would like to extend Pr* to a probability measure Pr** on $\mathscr{A}$ in such a way that Pr** makes as minimal a change as possible in Pr. Relative to several natural distance measures, the probability obtained by Jeffrey conditionalization fits the bill, although for some distance measures it may not do so uniquely (see Diaconis and Zabell 1982).

When the effect of observation is not so simple as to be localizable in a single partition, the method for updating probabilities becomes problematic. Suppose that one's experience results in new degrees of belief for each of the partitions $\{E_i\}$ and $\{F_j\}$. It is not guaranteed a priori that these degrees of belief are mutually coherent in the sense that they are extendible to a full probability on $\mathscr{A}$. A necessary and sufficient condition for the existence of such an extension is supplied by Diaconis and Zabell (1982). Assuming coherence, one could proceed to produce a new probability function by successive Jeffrey conditionalizations on the two partitions. But the order of conditioning may matter. If we denote the results of Jeffrey conditionalizing on $\{E_i\}$ (respectively $\{F_j\}$) by $\Pr_E(\cdot)$ ($\Pr_F(\cdot)$), then the order does not matter in that $\Pr_{EF}(\cdot) = \Pr_{FE}(\cdot)$ just in case $\Pr_F(E_i) = \Pr(E_i)$ and $\Pr_E(F_j) = \Pr(F_j)$ for all i and j.[12] The interested reader is referred to Diaconis and Zabell 1982 and van Fraassen 1989 for more discussion of these and related matters.

While the cumulative weight of the various justifications for conditionalization seems impressive, it should be noted that the starting assumptions of strict and Jeffrey conditionalization are left untouched. The former assumes that learning experiences have a precise propositional content in the sense that there is a proposition E that captures everything learned in the experience, while the latter assumes that if there is no precise propositional content, still the resulting belief changes can be localized to a partition. One or the other of these assumptions is surely correct for an interesting range of cases, but it is doubtful that they apply across the board. And where the doubt is realized, the present form of Bayesianism is silent.

In the remainder of this book I will concentrate on cases where strict conditionalization applies.

7 Lewis's Principal Principle

What David Lewis (1980, 1986) calls the principal principle (PP) may be viewed both as a rationality constraint on personal probabilities and as an implicit definition of objective probabilities. To paraphrase Lewis, (PP) requires that if $\Pr(\cdot)$ is a rational degree of belief function, A a proposition asserting that some specified event occurs at time t (e.g., a given coin lands heads up when flipped at t), A_p the proposition that asserts that the chance or objective probability at time t of A's holding is p, and E any proposition compatible with A that is admissible at t, then $\Pr(A/A_p \,\&\, E) = p$. Admissibility is, as Lewis notes, a tricky notion. But for present purposes it suffices to focus on one category of evidence that should be admissible in the intended sense, namely, any proposition E about matters of particular historical fact up to time t (e.g., information about the outcomes of past flips of the coin).

A glance at Bayes's calculations reported in chapter 1 is enough to establish that the Reverend Thomas himself used a version of (PP). Some of the mathematical niceties of Bayes's application of (PP) will be taken up in chapter 4, but these will be ignored in the present chapter to simplify the discussion.

Some early critics of probabilistic epistemology worried that the standard probability apparatus doesn't suffice to capture the full force of uncertain judgments. Consider two cases of partial knowledge. In the first I know literally nothing about a coin, save that it is two-sided and has a head and a tail. In the second I learn that 10,000 flips have produced 5,023 heads. If A is the proposition that the next flip will be heads, then in each of the two cases my degree of belief conditional on the total available evidence will presumably be (roughly) .5. But in the second case the "weight" of the evidence seems much greater, and consequently, my degree of belief is much firmer. The worry is that two numbers are needed to characterize my belief state, one describing my degree of belief, the other describing the weight of the evidence. But by using (PP), we can show that information about weight is already encoded in the standard probabilities. If we assume for sake of convenience that p can take on only discrete values p_i, we can write

$$\Pr(A/E) = \sum_i \Pr(A/A_{p_i} \,\&\, E) \times \Pr(A_{p_i}/E)$$

$$= \sum_i p_i \times \Pr(A_{p_i}/E),$$

where the first equality uses the principle of total probability and the second follows by (PP). The probability of A on E is thus the first moment of the distribution $\Pr(A_{p_i}/E)$. One would expect this distribution to look like figure 2.2a in the first hypothesized case and like figure 2.2b in the second. Thus, as E. T. Jaynes (1959) suggests, at least part of what is meant by 'weight of evidence' can be explicated in terms of the concentration of the $\Pr(A_p/E)$ distribution. This sense of weight is connected to the notion of firmness or resiliency, since presumably the greater the weight, the more new information about the outcomes of additional coin flips that is needed to significantly alter $\Pr(A_{p_i}/E)$ and thus $\Pr(A/E)$.

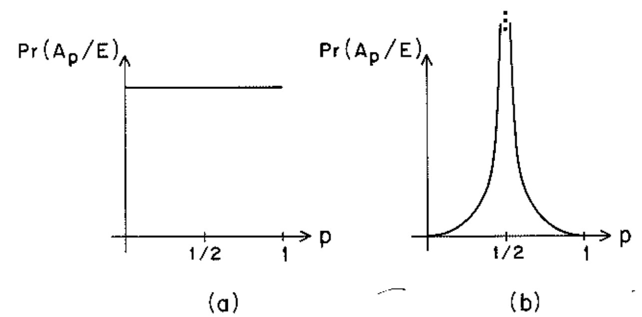

Figure 2.2
The distribution of personal probability of the objective probability

Principle (PP) is also (jokingly) referred to as Miller's principle because David Miller (1966) claimed to show that the principle is inconsistent. There is no need to review Miller's attack here, since Jeffrey (1970) and Howson and Urbach (1989) have successfully parried the attack. But it is worth reviewing van Fraassen's assessment of (PP), since if he is correct, (PP) has a less lofty status than it might seem to have at first glance.

Van Fraassen writes,

> The intuition that Miller's Principle is a requirement of rationality firmly links its credentials to a certain view of ourselves—namely, that we are finite, temporally conditioned rational beings. We have no crystal balls, and no way to gather information about the future which goes beyond the facts which have become settled to date. If we thought instead that Miller's Principle must apply to all possible and conceivable rational beings, we would have to conclude that omniscience implies determinism. (1989, p. 196)

The argument proceeds by supposing that there is a rational agent who is omniscient. For that agent, $\Pr(A) = 1$ or 0, according as A is true or false for any proposition A. So $\Pr(A_{p*}) = 1$ for some unique $0 \leqslant p^* \leqslant 1$. But then by (PP), we get $\Pr(A/A_{p*}) = p^* = \Pr(A \,\&\, A_{p*})/\Pr(A_{p*}) = \Pr(A)$. Hence p^* is 1 or 0, according as A is true or false, which is determinism.

There is pressure, however, to argue the other way around. The objective chance of a specified outcome (e.g., the reflection of a photon by a half-silvered mirror) is $0 < p^* < 1$. This I know because quantum mechanics (QM) tells me so. Therefore, if I am rational, I shouldn't assign 1 or 0 as my degree of belief in the outcome. So objective chance is incompatible with rational omniscience about the future, and the scope of (PP) does after all include all rational agents. A potential difficulty with this line is that it might seem to conflict with a reliability conception of knowledge. Thus, suppose that a person is able to correctly predict the future time after time. Wouldn't we eventually be willing to say that this person knows what the future holds? Correct prediction in itself is not a sure indicator of the relevant sort of reliability, for it is consistent with lucky guessing. What is needed for knowledge is the existence of a belief-forming mechanism that reliably yields certainty, a probability of 1 or 0, on the events in question. But such a mechanism is arguably inconsistent with the seemingly irreducible nontrivial probabilities involved in quantum events. Indeed, several of the no-go results for hidden-variable interpretations of QM are not so much proofs that no deterministic mechanism underlies QM as they are demonstrations of the inconsistency of treating quantum-mechanical magnitudes as if they had simultaneously determinate values.

In other cases, such as classical statistical mechanics, we want to maintain both determinateness and determinism on the microscopic level and yet speak of objective chances of events defined on the macroscopic level. For example, if a gas is initially confined to one half of a container by a partition and the partition is removed, then the chance is overwhelmingly great that in a time short by macroscopic standards the gas molecules, insofar as macroscopic measurements will be able to ascertain, will become evenly distributed over the entire container. In assigning personal probabilities, it would be irrational to ignore such teachings of statistical mechanics. But this judgment, as opposed to the parallel judgment in the QM case, rests, as van Fraassen says, on our view of ourselves as temporally bounded agents who have no crystal balls for reading the future. And it also rests on our view of ourselves as being bounded in other ways as well, in particular, as being unable to discern the current exact microstate of the gas. This second limitation means, in effect, that the admissible evidence for (PP) is more circumscribed than originally announced, which is an indication that the probabilities provided by classical statistical mechanics are not wholly objective. This is not an unwelcome conclusion, since it is generally acknowledged that these probabilities are partly physical and partly epistemic.

Principle (PP) also has the apparent virtue that when combined with the law of large numbers, it explains how we can come to learn the values of objective chance parameters. Consider a coin-flipping case with independent and identically distributed (IID) trials and an objective chance p for

heads. Starting from p, we can construct (as explained in appendix 2) a measure $\mathscr{P}_\ell$ on subsets of the collection of all possible outcomes of an infinite repetition of this chance experiment, and we can prove that the $\mathscr{P}_\ell$ measure of the set of all infinite sequences of flips in which the relative frequency of heads converges to p is 1 (the strong law of large numbers). It follows that if one starts by assigning a nonzero prior to the hypothesis that the objective chance of heads is p, obeys (PP), and updates probabilities by conditionalizing on the outcomes of repeated coin flips, then in almost every infinite repetition of the experiment (i.e., except for a set of $\mathscr{P}_\ell$-measure 0) one's personal probability will converge to 1 on the said hypothesis in the limit as the number of flips goes to infinity (see chapter 4).

The mathematics here is impeccable, but the metaphysics remains murky. If we think of p as something like a single-case propensity, then the original application of (PP) has a plausible ring to it. Moreover, given the assumption of IID trials, the objective probability in n trials of getting m heads is

$$\binom{n}{m} p^m (1-p)^{n-m}.$$

It follows that as $n \to \infty$ the objective probability goes to 0 that the relative frequency of heads differs from p by any specified $\varepsilon > 0$ (a form of the weak law of large numbers). So by applying (PP) at each stage of this reasoning, we can conclude that our personal probability goes to certainty that the frequency of heads comes within any desired $\varepsilon > 0$ of the true objective probability. But to get a personal-probability analogue of the strong form of the large numbers, we need to operate with the measure $\mathscr{P}_\ell$ on the collection of infinite repetitions, and it is not immediately apparent why $\mathscr{P}_\ell$ should function as an objective probability in the relevant sense of (PP), that is, so as to underwrite the conclusion that our personal probability ought to be one that in this infinite repetition of the experiment the limit of the relative frequency of heads will equal the objective probability of heads. The original (PP) can be defended on the grounds that it is constitutive of what is meant by objective probability. But we can only get away with such a move once.

Rather than start with single-case probabilities and then build the measure $\mathscr{P}_\ell$ on sets of infinite sequences, J. L. Doob (1941) proposed that we take as basic the measure $\mathscr{P}_\ell$, identify the event (say) of a coin's landing heads on the 35th flip with the set of all infinite sequences that yield heads in the 35th place, and then take the $\mathscr{P}\!r$ measure of this collection to be the probability of heads on the 35th flip (and thus by the IID assumption, the probability of heads on any trial). But what is now lacking is the conviction that this probability value functions enough like a single-case propensity so as to underwrite (PP) as applied to a particular, concrete flip.

It remains to be seen whether the difficulty here is trying to tell us something about the strong law of large numbers or about (PP) or both.

8 Descriptive versus Normative Interpretations of Bayesianism

Is Bayesianism to be regarded as descriptive of actual reasoning, or does it rather fix the pathways that "correct" or "rational" inductive reasoning must follow? Bayes's arguments for the probability axioms and the modern descendants of these arguments, the Dutch-book construction, certainly presuppose a normative aim, as do the discussions of rules of conditionalization and Lewis's (PP). This is just as well, since it is currently a matter of lively controversy as to whether actual agents can be represented as obeying the Bayesian constraints on belief and the allied decision rule of maximizing expected utility (see Kahneman, Slovic, and Tversky 1982).

Eschewing the descriptive in favor of the normative does not erase all difficulties. For 'ought' is commonly taken to imply 'can', but actual inductive agents can't, since they lack the logical and computational powers required to meet the Bayesian norms. The response that Bayesian norms should be regarded as goals toward which we should strive even if we always fall short is idle puffery unless it is specified how we can take steps to bring us closer to the goals. To make the complaint concrete, note that in a rich language, agents who are computationally bounded may fail to satisfy probability axiom (A2). This failure is not a mere inadvertence that can easily be corrected, since by their very nature these agents fall short of the logical omniscience that requires recognition of all logical truths in the domain of Pr. Thus a realistic Bayesianism must somehow make room for logical learning. And it is in this regard that one must agree with Good (1977) that probability qua degree of belief can change not only as the result of observation and experiment but also as a result of calculation and pure thought. This matter will surface again in chapter 5 in the discussion of the problem of old evidence.

Actual agents also fall short of logical omniscience by being unable to parse all the possibilities, and this inability can skew degrees of belief.

The probability calculus requires that the degrees of belief assigned to Einstein's general theory of relativity (GTR) and its negation sum to one. But when Einstein first proposed GTR, physicists had only the dimmest idea of what was contained in the portion of possibility space denoted by $\neg$GTR, and thus their assessments of the probability of GTR were ill informed in the worst way. 'Explore the space of possibilites' is an empty injunction unless accompanied by practical guidelines. Although I have no general prescription to offer in this regard, I will offer in chapter 7 some examples of how the exploration has been conducted in some actual and challenging cases. It should be noted, however, that such an exploration cannot be undertaken in an orthodox Bayesian fashion, for the recognition of heretofore obscured possibilities is typically accompanied by belief changes, and it is hardly possible to account for all of these changes by conditionalization, whether of the strict or Jeffrey form. This matter will be taken up in chapters 7 and 8.

Finally, the considerations of chapter 9 raise a new and different challenge to the normative status of Bayesianism by showing that the structural constraints it imposes on degrees of belief entail a substantive knowledge of a kind that most scientists would not regard as appropriate to bring to a domain of inquiry.

9 Prior Probabilities

The topic of priors will come up again and again in the chapters below. While it would not be productive to anticipate in advance all of the nuances of the discussion, it may nevertheless be useful to outline the shape of one of the central issues. For the Bayesian apparatus to be relevant to scientific inference, it seems that what it needs to deliver are not mere subjective opinions but reasonable, rational, objective degrees of belief. Thence comes the challenge: How are prior probabilities to be assigned so as to make this delivery possible? (Note that the presupposition of this challenge is that the other factors involved in Bayes's theorem, the likelihoods, are unproblematic in a way that priors are not. While this may be true in some special cases, it is most certainly not true in general. But since this point only serves to complicate the matter at hand, I waive it for the time being.)

Three responses to the challenge are to be found in the Bayesian corpus. The first is that the assignment of priors is not a critical matter, because as the evidence accumulates, the differences in priors "wash out." Chapter 6 will examine in detail various theorems that are supposed to demonstrate this washout effect. In advance it is fair to say that the formal results apply only to the long run and leave unanswered the challenge as it applies to the short and medium runs.

The second response is to provide rules to fix the supposedly reasonable initial degrees of belief. In chapter 1 we met Thomas Bayes's attempt to justify the rule of a uniform prior distribution. We saw that, although ingenious, Bayes's attempt is problematic. Other rules for fixing priors suffer from similar difficulties. And generally, none of the rules cooked up so far are capable of coping with the wealth of information that typically bears on the assignment of priors.

The third response is that while it may be hopeless to state and justify precise rules for assigning numerically exact priors, still there are plausibility considerations that can be used to guide the assignments. By way of concrete illustration, consider the recent controversy about AIDS transmission raised by Lorraine Day, a San Francisco surgeon. Day worried that surgeons might contract AIDS by inhaling the air-borne blood of infected patients. To protect against this risk, she urged her colleagues to wear space-suit-like outfits when using high speed drills and saws that create a fine mist of blood droplets. Critics responded that it is implausible that Day's hypothetical transmission mechanism poses any serious risk, since the AIDS virus should remain suspended in the blood droplets and since these droplets are typically too large to pass through the openings of standard surgical masks. In Bayesian jargon, the critics are urging that these plausibility considerations justify assigning a low prior to Day's hypothesis.

This third response does point to an important aspect of actual scientific reasoning, but at the same time it opens the Bayesians to a new challenge, which Fredrick Suppe has put in the form of a dilemma:

> If standard inductive logic [i.e., Bayesianism] is intended to provide an analysis of that plausibility reasoning, then we have a vicious regress where each iteration of the Bayesian method requires a logically prior application; hence it is impossible to ever get the Bayesian method going. Hence standard inductive logic is an inadequate model of scientific reasoning about evidence and the evaluation of hypotheses. If, on the other hand, standard inductive logic does not provide an analysis of that plausibility reasoning, standard inductive logic is a critically incomplete, hence an inadequate model of scientific reasoning about evidence and the evaluation of hypotheses. (1989, p. 399)

Although some Bayesians have tried to seize one or the other of the horns of this dilemma, it seems to me that the only escape is between the horns.[13] That is, Bayesians must hold that the appeal to plausibility arguments does not commit them to the existence of a logically prior sort of reasoning: plausibility assessment. Plausibility arguments serve to marshall the relevant considerations in a perspicuous form, yet the assessment of these considerations comes with the assignment of priors. But, of course, this escape succeeds only by reactivating the original challenge. The upshot seems to be that some form of the washout solution had better work not just for the long run but also for the short and medium runs as well.

The matter of plausibility arguments also serves to bring to the surface one of the lingering doubts that many philosophers have about Bayesianism. The worry is that the Bayesian apparatus is just a kind of tally device used to represent a more fundamental sort of reasoning whose essence does not lie in the assignment of little numbers to propositions in accord with the probability axioms. The only effective way to assuage this worry is to examine the many attempts to capture scientific reasoning in non-Bayesian terms and to detail how each of these attempts fails. Some of this work will be done in chapter 3.

10 Conclusion

In the next two chapters I will assume that Bayesians are armed with the probability calculus, including countable additivity if it should prove helpful, and also with whatever form of conditionalization seems appropriate to the context. How this arsenal is deployed to attack problems in confirmation theory will be the subject of discussion.

Appendix 1: Conditional Probability

Conditional probability

Suppose that $\mathscr{B} \subseteq \mathscr{A}$ and let $\mathscr{B}^0$ stand for the noncontradictory elements of $\mathscr{B}$. Then a conditional probability $\Pr(\cdot/\cdot)$ is a function from $\mathscr{A} \times \mathscr{B}^0$ to $\mathbb{R}$ satisfying the following:

CP1 $\Pr(\cdot/B)$ is an unconditional probability on $\mathscr{A}$ for any $B \in \mathscr{B}^0$.

CP2 $\Pr(B/B) = 1$ for any $B \in \mathscr{B}^0$.

CP3 $\Pr(A \,\&\, B/C) = \Pr(B/C) \times \Pr(A/B \,\&\, C)$ for any $A \in \mathscr{A}$, $B \in \mathscr{B}^0$, $C \in \mathscr{B}^0$, and $B \,\&\, C \in \mathscr{B}^0$.

$\Pr(\cdot/\cdot)$ is said to be *full* just in case $\mathscr{B} = \mathscr{A}$. It will be assumed here that we are dealing with full conditional probabilities, since any conditional probability can be extended to a full one. If $\models N$ for $N \in \mathscr{B}^0$, $\Pr(\cdot) \equiv \Pr(\cdot/N)$ is the unconditional probability associated with $\Pr(\cdot/\cdot)$. ($\Pr(\cdot)$ is independent of the choice of N.) It is easy to see that if $B \in \mathscr{B}^0$ is such that $\Pr(B) \neq 0$, $\Pr(A/B) = \Pr(A \,\&\, B)/\Pr(B)$ for any $A \in \mathscr{A}$.

Countable additivity, disintegrability, and conglomerability

A *partition* of the possibilities consists of a set $\{H_1, H_2, \ldots\}$ of statements $H_i \in \mathscr{A}$ that are pairwise exclusive and mutually exhaustive, i.e., $\{H_i, H_j\} \models P \,\&\, \neg P$ for $i \neq j$ and $\{\neg H_1, \neg H_2, \ldots\} \models P \,\&\, \neg P$. Let $\Pr(\cdot/\cdot)$ be a conditional probability and $\Pr(\cdot)$ its associated unconditional probability. Here $\Pr(\cdot)$ is said to be *countably additive* just in case the continuity condition (C) (p. 36) holds. This condition implies that for any partition $\{H_1, H_2, \ldots\}$, $\lim_{n\to\infty} \Pr(\bigvee_{i \leqslant n} H_i) = \sum_{i=1}^{\infty} \Pr(H_i) = 1$.

A countably additive $\Pr(\cdot)$ associated with the conditional probability $\Pr(\cdot/\cdot)$ has the property of *disintegrability*: for any $A \in \mathscr{A}$ and any partition $\{H_1, H_2, \ldots\}$, $\Pr(A) = \sum_{i=1}^{\infty} \Pr(A/H_i) \times \Pr(H_i)$. Disintegrability for a partition in turn entails *conglomerability*: if $k_1 \leqslant \Pr(A/H_i) \leqslant k_2$ for every i, then $k_1 \leqslant \Pr(A) \leqslant k_2$. The circle closes: conglomerability with respect to every countable partition implies countable additivity (Schervish, Seidenfeld, and Kadane 1984).

The failure of countable additivity and consequently of conglomerability can lead to very awkward situations, such as the failure of a natural principle of dominance, which demands that if an action O_1 is conditionally preferred to O_2 for each member of a partition $\{H_1, H_2, \ldots\}$, then O_1 is unconditionally preferred to O_2 (see Kadane, Schervish, and Seidenfeld 1986). However, nonconglomerability should not be allowed to become a bugaboo, since even when countable additivity holds, conglomerability can fail for uncountable partitions (see Kadane, Schervish, and Seidenfeld 1986).

The failure of disintegrability is also very awkward for Bayesian inference problems, since it means that the denominator of Bayes's theorem (2.1) cannot be written in the form given in (2.2); so, for example, the probability of the experimental outcome E cannot be assessed in terms of how well the alternative hypotheses H_i explain the outcome (as given by the likelihoods $\Pr(E/H_i \,\&\, K)$) and how antecedently probable the hypotheses are (as given by the priors $\Pr(H_i/K)$). This theoretical worry is of no practical importance if in all realistic cases inference involves only a finite number of H_i.

Finally, it should be noted that Thomas Bayes's own calculations (chapter 1) implicitly assumed a form of countable additivity (see chapter 4).

Appendix 2: Laws of Large Numbers

Recall that a finite field $\mathscr{F}_0$ over the set Ω is a collection of subsets of Ω that contains Ω and is closed under complementation and finite unions (and thus under finite intersections). A σ field $\mathscr{F}_\sigma$ is closed under countable unions (and thus under countable intersections). Each $\mathscr{F}_0$ generates a σ field $\mathscr{F}_\sigma$, namely, the smallest σ field containing $\mathscr{F}_0$. Let $\mathscr{pr}$ be a finitely additive probability measure on $\mathscr{F}_0$. (For any $A \in \mathscr{F}_0$, $\mathscr{pr}(A) \geqslant 0$; $\mathscr{pr}(\Omega) = 1$. And for any $A_1, A_2, \ldots, A_n \in \mathscr{A}$ such that $A_i \cap A_j = \varnothing$ for $i \neq j$, $\mathscr{pr}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathscr{pr}(A_i)$.) The function $\mathscr{pr}$ is said to be continuous from above at $\varnothing$ just in case if $A_i \in \mathscr{A}$, $i = 1, 2, \ldots$, are such that $A_{i+1} \subseteq A_i$ and $\bigcap_{i=1}^{\infty} A_i = \varnothing$, then $\mathscr{pr}(\bigcup_{i=1}^{n} A_i) \to 0$ as $n \to \infty$. (This implies a conditional form of countable additivity: *if* $\bigcup_{i=1}^{\infty} B_i \in \mathscr{A}$ and the B_i are pairwise disjoint, then $\mathscr{pr}(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mathscr{pr}(B_i)$.) Carathéodory's extension lemma shows that for such a $\mathscr{pr}$ there is a unique extension to a countably additive probability measure $\mathscr{P}r$ on the σ field generated by $\mathscr{F}_0$.

In the application to IID trials of coin flips, take Ω to be the collection of all one-sided infinite sequences of possible outcomes. (So a typical $\omega \in \Omega$ would be HTHHTHTTT....) Define a finite field of subsets of Ω by starting with the "cylinder sets," where a cylinder set is the set of all ω's that agree on the outcomes in a finite number of places. (A typical cylinder set would be the collection of all ω's that have heads in the 20th place and tails in the 801st place.) $\mathscr{F}_0$ is then the finite field consisting of the empty set and finite disjoint unions of the cylinder sets. A $\mathscr{pr}$ measure is defined on $\mathscr{F}_0$ by assigning probabilities to the cylinder sets in the natural way. For example, the measure of the set of all ω's having heads in the 32nd and 41st places and tails in the 33rd, 58th, and 105th places is $p^2(1-p)^3$, where p is the objective probability of heads. Since this $\mathscr{pr}$ is continuous from above at $\varnothing$, it follows from Carathéodory's lemma that there is a unique countably additive extension $\mathscr{P}r$ of $\mathscr{pr}$ to the σ field $\mathscr{F}_0$ generated by $\mathscr{F}_0$.

If we now let $j_n(\omega)$ stand for the number of heads in the first n trials of ω, the weak and strong form of the law of large numbers can be stated as follows:

WLLN The $\mathscr{P}\!r$ measure of the set of ω's for which $|(j_n(\omega)/n) - p| > \varepsilon$ approaches 0 as $n \to \infty$ for any $\varepsilon > 0$.

SLLN The $\mathscr{P}\!r$ measure of the set of all ω's such that $\lim_{n\to\infty}(j_n(\omega)/n) \neq p$ is 0.

To put (SLLN) in its positive form, the Pr probability is one that the limiting relative frequency of heads converges to p.

As indicated in section 7, a form of the weak law of large numbers can be formulated and proved without the help of countable additivity. Roughly, for any $\varepsilon > 0$, the probability (in the objective sense or in the degree-of-belief sense tempered by Lewis's principal principle) that the actually observed relative frequency of heads differs from p by more than ε goes to 0 as the number of flips goes to infinity. This form of the law of large numbers is to be found in the work of Bernoulli. The strong form of the law of large numbers, which requires countable additivity, was not proved until this century (see Billingsley 1979 for a proof).

3 Success Stories

The successes of the Bayesian approach to confirmation fall into two categories. First, there are the successes of Bayesianism in illuminating the virtues and pitfalls of various approaches to confirmation theory by providing a Bayesian rationale for what are regarded as sound methodological procedures and by revealing the infirmities of what are acknowledged as unsound procedures. The present chapter reviews some of these explanatory successes. Second, there are the successes in meeting a number of objections that have been hurled against Bayesianism. The following chapter discusses several of these successful defenses. Taken together, the combined success stories help to explain why many Bayesians display the confident complacency of true believers. Chapters 5 to 9 will challenge this complacency. But before turning to the challenges, let us give Bayesianism its due.

1 Qualitative Confirmation: The Hypotheticodeductive Method

When Carl Hempel published his seminal "Studies in the Logic of Confirmation" (1945), he saw his essay as a contribution to the logical empiricists' program of creating an inductive logic that would parallel and complement deductive logic. The program, he thought, was best carried out in three stages: the first stage would provide an explication of the qualitative concept of confirmation (as in '*E* confirms *H*'); the second stage would tackle the comparative concept (as in '*E* confirms *H* more than *E*′ confirms *H*′'); and the final stage would concern the quantitative concept (as in '*E* confirms *H* to degree *r*'). In hindsight it seems clear (at least to Bayesians) that it is best to proceed the other way around: start with the quantitative concept and use it to analyze the comparative and qualitative notions. The difficulties inherent in Hempel's own account of qualitative confirmation will be studied in section 2. This section will be devoted to the more venerable hypotheticodeductive (HD) method.

The basic idea of HD methodology is deceptively simple. From the hypothesis *H* at issue and accepted background knowledge *K*, one deduces a consequence *E* that can be checked by observation or experiment. If Nature affirms that *E* is indeed the case, then *H* is said to be HD-confirmed, while if Nature affirms $\neg E$, *H* is said to be HD-disconfirmed. The critics of HD have so battered this account of theory testing that it would be unseemly to administer any further whipping to what is very