

CHAPTER 8

BAYESIANISM

JAMES M. JOYCE

BAYESIANISM provides a unified theory of epistemic and practical rationality based on the *principle of mathematical expectation*.¹ In its epistemic guise it requires believers to obey the *laws of probability*. In its practical guise it asks agents to maximize *subjective expected utility* (see also Dreier, chap. 9, and Bicchieri, chap. 10, this volume).

This essay will be concerned primarily with Bayesian epistemology. Its first section defends the view that beliefs come in varying grades of strength. Section 2 explores the Bayesian requirement of *probabilistic consistency* for graded beliefs. Section 3 explicates Bayesian confirmation theory. Section 4 discusses the thesis that rational belief change proceeds via *conditioning*. Section 5 addresses the charge that Bayesianism engenders an untenable *subjectivism*.

1. GRADED BELIEFS AND CONDITIONAL BELIEFS

Bayesians maintain that any adequate epistemology must recognize that beliefs come in varying *gradations of strength*. They seek to replace the *categorical* notion of belief as an all-or-nothing attitude of accepting a proposition as true with a graded conception of belief as *level of confidence*. In general, a person's

level of confidence in a proposition X will correspond to the extent to which she is disposed to presuppose X 's truth in her theoretical and practical reasoning.²

There are two compelling reasons for thinking that opinions vary in strength. First, this is needed to make sense of decision making. To explain why Smith will bet on Stewball at 4-to-1 odds but not at even odds, we must suppose that she is not sure that Stewball will lose but is more confident that he will lose than that he will win. Second, since evidence comes in a wide variety of types and strengths, a person will be able to proportion her beliefs to her evidence only if her beliefs come in gradations. Someone who wants to know whether a coin will land heads when tossed might have as evidence any proposition of the form "the coin was fairly tossed a thousand times and n heads came up." Each value of n calls for a different doxastic attitude.

People also have graded *conditional* beliefs that express their confidence in the truth of propositions *on the supposition that other propositions are facts*. If a person has determinate unconditional beliefs in Y and $X \& Y$, and is not certain of $\neg Y$, then her belief in X conditional on Y will be a function of her unconditional beliefs in Y and $X \& Y$. That said, a person may hold a definite belief about X conditional on Y even when she has no determinate unconditional beliefs for $X \& Y$ and Y , or when she is certain that Y does *not* obtain. Conditional beliefs are best seen as *sui generis* judgments that cannot be *reduced* to unconditional beliefs.³

A crucial challenge for Bayesians is to explain how strengths of beliefs can be measured. It is often said that Bayesianism is committed to the existence of sharp, numerical *degrees of belief*. This requires each believer to have a *confidence measure* that, for each proposition X and condition Y , specifies a number $c_Y(X)$ that gauges her level of confidence in X conditional on Y . For fixed Y , $c_Y(\bullet)$ captures the person's degrees of belief on the supposition Y , and her unconditional beliefs are given by $c(\bullet) = c_T(\bullet)$ where T is any truth of logic. By convention, $c(X) = 1$ indicates complete certainty in X 's truth, while $c(X) = 0$ indicates complete incredulity.

The idea that people have numerically sharp degrees of belief has been widely criticized both for psychological implausibility and because it requires people to hold opinions far more definite than their evidence warrants.⁴ As a result, most Bayesians now grant that few graded beliefs can be precisely quantified. Instead, beliefs are represented variously by interval-valued probabilities, convex sets of confidence measures, or confidence orderings. The most general approach is to characterize a person's opinions using a system of descriptive constraints that might include any of the following sorts of statements:

- She is more confident in X than in Y .
- She believes Z to at least degree $\frac{1}{3}$ but at most degree $\frac{3}{4}$.

- She believes X conditional on Y more strongly than she believes Z conditional on W .

The person's graded beliefs can then be represented by the set Con of all confidence measures that satisfy the constraints. For this system, Con contains all measures: $c(X) > c(Y)$; $3/4 > c(Z) > 1/3$; $c_Y(X) > c_W(Z)$. In ideal cases the constraints will be so detailed that Con contains a single measure, but it will more commonly contain many measures. Facts about the person's opinions are given by properties that all Con 's elements share, for example, she can be said to believe X to degree x only when $c(X) = x$ for every $c \in Con$.

We can now state the first core tenet of Bayesian epistemology.

Thesis of Graded Belief. Any adequate epistemology must recognize that opinions come in varying gradations of strength. A person's graded beliefs can be represented using a set Con of confidence measures. Facts about her beliefs correspond to properties shared by all functions in Con .

2. THE REQUIREMENT OF PROBABILISTIC CONSISTENCY

The second core tenet of Bayesian epistemology requires that rational beliefs be consistent with the laws of probability. A probability function P assigns non-negative real numbers to propositions in such a way that

Normalization. $P(T) = 1$ for T any truth of logic.

*Additivity.*⁵ $P(X \vee Y) = P(X) + P(Y)$ if X and Y are logically incompatible.

A conditional probability assigns numbers to pairs of propositions X/Y in such a way that

Probability. $P(\bullet/Y)$ is a probability for every Y with $P(Y) > 0$.

Conditional Normalization. $P(Y/Y) = 1$.

Conditioning. $P(X/Y \& Z) \times P(Y/Z) = P(X \& Y/Z)$.⁶

Hereafter, "the laws of probability" will denote these five requirements.

Here are two useful consequences of these laws:

Logical Consequence. If X entails Y , then $P(Y) \geq P(X)$.

*Bayes's Theorem.*⁷ $P(X/Y) = [P(X) \div P(Y)] \times P(Y/X)$.

Logical Consequence ensures that probabilistic reasoning respects deductive logic. Bayes's Theorem relates the "direct" probability of X conditional on Y to the ratio of the unconditional probabilities of X and Y , and the so-called *inverse probability* (or "likelihood") of Y conditional on X .

Here is the basic Bayesian requirement of epistemic rationality:

Probabilistic Consistency. A rational subject's beliefs must conform to the laws of probability in the sense that at least one confidence measure that represents her beliefs must be a probability.

In other words, there must be a c in Con and a conditional probability such that $c_Y(X) = P(X/Y)$ whenever $P(X/Y)$ is defined.

The most common Bayesian rationale given for probabilistic consistency is the famous Dutch Book Argument (DBA) of Frank Ramsey (1931) and Bruno de Finetti ([1937] 1964). This argument purports to show that anyone whose beliefs violate the laws of probability is *practically irrational*. In broadest outline the reasoning runs thus:

Coherence. A practically rational agent will never freely perform any action when another act is certain to leave her better off in all possible circumstances.

Belief/Desire Psychology. A practically rational agent will always act in ways that she estimates will best satisfy her desires.

The EU-Thesis. A practically rational agent will estimate that an act best satisfies her desires if and only if that act maximizes her *subjective expected utility*.

Dutch Book Theorem. An agent who tries to maximize her subjective expected utility using beliefs that violate the laws of probability will freely perform an act that is sure to leave her worse off than some alternative act would in all circumstances.

Therefore, it is *practically irrational* to hold beliefs that violate the laws of probability.

While the DBA's first premise has generated little controversy, its second is often dismissed as bad psychology. A vast body of experimental evidence shows that, in addition to beliefs and desires, actions are affected by emotions, habits, decision-making heuristics, and judgmental biases.⁸ This is all undeniable, but those who see it as a problem for the DBA are confused about the Bayesian enterprise. Bayesians have always made it clear that they are offering a *normative* theory of *rational* behavior, *not* an empirical theory of actual behavior.⁹ Emotions, habits, and so on do *cause* actions, but the DBA does *not* rely on the belief/desire model as a causal theory of action. Rather, the model serves to determine which actions an agent has *sound reasons* to perform. Premise 2 makes no claims whatsoever about the psychological mechanisms that prompt actions. It says, rather, that what *makes* an act rational is that it bears the right relationship to the actor's beliefs and desires. When read this way, premise 2 has nothing to fear from empirical psychology.¹⁰

The EU-thesis is the most controversial premise of the DBA. As Ramsey (1931, 174) expressed it, "I suggest that we introduce as a law of psychology that behavior is governed by . . . mathematical expectation . . . If X is a proposition about which [an agent] is doubtful, any goods or bads for whose realization X is in his view a necessary and sufficient condition enter into his calculations multiplied by the same fraction, which is called the degree of his belief in X ." We will discuss the general EU-thesis shortly, but the DBA requires only a special case. Assume that: (a) the agent desires *only* money; (b) her desire for money does not vary with changes in her fortune; and (c) she is not averse to risk or uncertainty. The key insight of the DBA is that, if we ignore the fact that offering someone a bet on X can alter her opinions, *the EU-thesis entails that a person satisfying (a)–(c) will reveal the strengths of her beliefs in her betting behavior.*

Suppose we offer our agent a *wager* $W = [\$a \text{ if } X, \$b \text{ else}]$ in which she is *certain* that she will get $\$a$ if X is true and $\$b$ if X is false, and that X 's truth does not depend causally on W . The EU-thesis then entails that the agent's level of confidence in X will be revealed by the monetary value she puts on W . Her *fair price* for W is that sum of money $\$f$ at which she is indifferent between receiving a payment of $\$f$ or having W go into effect. When the agent has a definite degree of belief $c(X)$ for X she will value W by its *expected payoff*, so $f = \text{Exp}(W) = c(X) \times a + (1-c(X)) \times b$. Her fair price for W is then related to her degree of belief in X by the equation $c(X) = (f-b) \div (a-b)$.¹¹ So, if she is indifferent between $\$63.81$ and a bet that pays $\$100$ if it rains and $\$0$ if not, then she is confident to degree 0.6381 that it will rain.¹²

Once this connection between degrees of belief and fair prices is granted, the DBA becomes an exercise in mathematics. Given (a)–(c), the EU-thesis entails

that the agent will choose to swap any set of wagers for the sum of their fair prices, or swap any set of fair prices for its associated sequence of wagers. This "package principle"¹³ entails that a person who sets prices $\$0.25$, $\$0.25$, and $\$0.6$ on $W_X = [\$1 \text{ if } X, \$0 \text{ else}]$, $W_Y = [\$1 \text{ if } Y, \$0 \text{ else}]$ and $W_{X \vee Y} = [\$1 \text{ if } X \vee Y, \$0 \text{ else}]$, respectively, will exchange a portfolio containing $\$0.6$ and the first two wagers for one containing $\$0.5$ and the third. When X and Y are logically incompatible she buys nothing with her dime since the combination of W_X and W_Y is identical, in terms of payoffs, to $W_{X \vee Y}$. When an agent takes a self-defeating action of this sort, she is said to have made "Dutch book" against herself. Coherence entails that doing so is irrational.

The Dutch Book Theorem shows that susceptibility to Dutch books is the penalty for transgressing the laws of probability.

Dutch Book Theorem. Imagine an EU-maximizer who satisfies (a)–(c) and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

One proves this by showing that someone who violates a given law of probability is thereby committed to buying and selling a series of wagers whose net effect will be to cost her money come what may. For example, if one violated Additivity by having degrees of belief 0.25, 0.25, and 0.6 for X , Y , and $X \vee Y$, respectively, then one will be susceptible to the Dutch book just described. All other violations of the laws of probability have similar "Dutch book" justifications.

The DBA has three shortcomings. It assumes an agent who meets conditions (a)–(c), who sets a fair price on every wager she considers, and who maximizes expected utility. The first two restrictions are unrealistic, and the claim that practical rationality requires EU-maximization is controversial. An adequate justification of Probabilistic Consistency must both relax the first two restrictions and provide an independent justification for EU-maximization.

Fair prices can be avoided if rational agents are permitted to have incomplete or imprecise preferences that are *extendible* to completely precise preferences that avoid Dutch books.¹⁴ If one accepts this *requirement of coherent extendibility*, then for any rational agent there will be at least one, but usually many, complete sets of coherent fair prices that are consistent with her preferences. By the Dutch Book Theorem, each such system determines a probability function, and so the person's *Con* set will contain at least one probability.

While many Bayesians find this solution appealing, it would be better to avoid the detour through coherent extendibility altogether by showing that agents with imprecise or incomplete beliefs that violate the laws of probability are susceptible to Dutch books outright. While no general proof of this sort exists, one can go a long way toward the goal in some special cases. Suppose belief strengths can be

characterized *comparatively* by a relation $X .>. Y$ that holds when the agent is at least as confident in X as in Y . This relation need *not* be complete: neither $X .>. Y$ nor $Y .>. X$ needs to hold since the agent might have no determinate view about the relative likelihood X and Y . Even under these weak conditions, probabilistic inconsistency leaves the agent susceptible to Dutch book.¹⁵ This result can be generalized to apply to comparative conditional beliefs and to beliefs with “interval-valued” characterizations. Hence, under broad conditions, expected utility maximizers with probabilistically inconsistent beliefs, even imprecise ones, are susceptible to Dutch books.

To relax (a)–(c) and defend the EU-thesis, we must leave the DBA behind. What is needed is a *representation theorem* that justifies probabilistic consistency and expected utility maximization simultaneously, but without restricting the *content* of the agent’s desires.¹⁶ A number of such theorems can be found in the literature, but the one due to Leonard Savage ([1954] 1972) has attained the status of a “standard model.” Savage imagines an agent who uses her beliefs about the world’s possible *states* and her desires for *consequences* to form *preferences* among what he calls *acts*.¹⁷ Acts, states, and consequences are individuated so that (i) each act/state pair produces a *unique* consequence that settles every issue the agent cares about, and (ii) she is convinced that her behavior will make no causal difference to which state obtains. The agent is assumed to have a *preference ranking* over acts. For any acts A and B , she might strictly prefer A to B , be indifferent between them, or strictly prefer B to A .

In Savage’s framework, an *expected utility* is a function

$$Exp_{P,u}(A) = \sum_{S \in States} P(S) \times u(A, S)$$

where A is an act, P is a probability over states, and u is a *utility* function. $u(A, S)$ measures the degree to which the agent’s desires will be satisfied by the consequence produced by A and S . $Exp_{P,u}(A)$ is her estimate of the degree to which A is likely to produce consequences that satisfy her. Within Savage’s framework, we can state the basic Bayesian requirement of practical rationality as follows:

EU-coherence. There must be at least one probability P defined on states and one utility u for consequences that *represent* the agent’s preferences in the sense that, for any acts A and B , she strictly (weakly) prefers A to B only if $Exp_{P,u}(A)$ is greater than (as great as) $Exp_{P,u}(B)$.

To justify EU-coherence as the standard of practical reason, Savage imposes a system of axiomatic constraints on preferences and proves that satisfaction of these constraints guarantees the existence of the required P and u . Here are in-

formal analogues of Savage’s main axioms (where A, B, C are acts, and X and Y are disjunctions of states)

Trichotomy. The agent strictly prefers A to B , strictly prefers B to A , or is indifferent between them.

Transitivity. If the agent (strictly or weakly) prefers A to B and B to C , then she also prefers A to C (see also Sorensen, chap. 14, this volume).

“Sure Thing” Principle. If A and B produce the same consequences in every state consistent with $\neg X$, then the agent’s preference between the two acts depends only on their consequences when X obtains (see also Dreier, chap. 9, and Sorensen, chap. 14, this volume).

Wagers. For any consequences O_1 and O_2 , and any event X , there is an act [O_1 if X , O_2 else] that produces O_1 in any state that entails X and O_2 in any state that entails $\neg X$.

Savage’s P4. If the agent prefers [O_1 if X , O_2 else] to [O_1 if Y , O_2 else] when O_1 is more desirable than O_2 , then she will also prefer [O_1^* if X , O_2^* else] to [O_1^* if Y , O_2^* else] for any other outcomes such that O_1^* is more desirable than O_2^* .

Savage showed that these axioms, along with a few others, guarantee the existence of a unique probability P and a utility u , unique up to the arbitrary choice of a unit and zero-point, whose associated expectation represents the agent’s preferences.¹⁸

Many Bayesians use this result to justify probabilistic consistency. P4 is the lynchpin of their case. Savage uses P4 to “define” what it means for the agent to be more confident in X than in Y :

A practically rational agent *believes X more strongly than she believes Y* if and only if she strictly prefers [O_1 if X , O_2 else] to [O_1 if Y , O_2 else] for some (hence *any*, by P4) outcomes with O_1 more desirable than O_2 .

Savage treated this as a “definition” because, like many social scientists of his day, he believed that legitimate objects of scientific study must be operationally defined. Strengths of beliefs are defined in terms of preferences, which are operationally defined in terms of overt choices. There are many well-known problems with this outmoded behaviorist methodology, but we need not endorse it to recognize that Savage’s principle correctly captures the relationship between prefer-

ences and belief strengths (albeit not as a matter of definition). If O_1 is preferred to O_2 , then the agent has a *good reason* for preferring $[O_1 \text{ if } X, O_2 \text{ else}]$ to $[O_1 \text{ if } Y, O_2 \text{ else}]$ exactly if she is more confident in X than in Y . Given this, P4 entails that the agent's beliefs are represented by the probability in any (P, u) pair that represents her preferences. Probabilistic Consistency is thus derived as a consequence of the theory of practical rationality embodied in Savage's axioms.

There is a vast literature dedicated to determining whether or not these axioms really are requirements of practical rationality. Savage's defenders seek to show that violating them leads agents to choose means that are necessarily insufficient for their ends. Critics try to refute these arguments, typically by arguing that they distort rational attitudes toward risk or uncertainty. The "Sure Thing" Principle has been especially controversial, but Transitivity has been questioned as well. We cannot hope to even scratch the surface of these issues here and will leave interested readers to consult the literature.¹⁹

There is, however, a broader worry about both the DBA and the representation theorem approaches. These arguments can show only that it is *practically* irrational to violate the laws of probability. Some critics have wondered why this should indicate anything at all about the *epistemological status* of probabilistically inconsistent beliefs. Ralph Kennedy and Charles Chihara (1979, 30) put the point concisely: "The factors that are supposed to make it irrational to have a [probabilistically inconsistent] set of beliefs . . . are irrelevant, epistemologically, to the truth of the propositions in question. The fact (if it is a fact) that one will be bound to lose money unless one's degrees of belief [obey the laws of probability] just isn't epistemologically relevant to the truth of those beliefs."²⁰

In responding to such worries many Bayesians invoke a form of *pragmatism*. Ramsey writes that the strength of a belief "is a causal property of it, which we can express vaguely as the extent to which we are prepared to act on it" (1931, 169). De Finetti thinks it "trivial and obvious . . . that the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on that event" ([1937] 1964, 101). Savage, as we saw, simply *defines* the strengths of beliefs in terms of characteristic patterns of preferences for actions. Anybody who conceives of beliefs this way, as mere causes for actions, will resist the suggestion that there is any specifically "epistemological" way of evaluating them. There is nothing more to the rationality of beliefs, they will say, than their propensity to produce practically rational actions.

Other Bayesians are loath to endorse such an uncompromising pragmatism. In addition to their role in producing actions, beliefs respond to evidence, serve as premises in theoretical reasoning, and are judged in terms of the truth or falsity of their constituent propositions. Given all this, it seems a mistake to tie the rationality of beliefs too closely to that of actions or preferences. Some who feel this way adopt a "depragmatizing" strategy and reinterpret the DBA and representation theorems so as to expose the specifically *epistemic* costs of probabilistic

inconsistency. Others offer new sorts of arguments that directly bring out the epistemic shortcomings of probabilistically inconsistent beliefs.

An early "depragmatizer," Brian Skyrms, writes that "what is basic [in the DBA or representation theorems] is the consistency condition that you evaluate a betting arrangement independently of how it is described (e.g., as a bet on $X \vee Y$ or as a system of bets consisting of a bet on X and a bet on Y) . . . The cunning bettor is simply a dramatic device—the Dutch book a striking corollary—to emphasize the underlying issue of coherence" (1984a, 21–22).²¹ In other words, the cognitive flaw associated with probabilistic inconsistency is a susceptibility to what psychologists call *framing effects*, in which a *single* option is evaluated differently when presented under different guises.²² To illustrate, suppose that a person is both more confident in X than in Y and more confident in $Y \vee Z$ than in $X \vee Z$ where Z is incompatible with X and Y . Savage's axioms entail that, for any outcomes with O_1 strictly preferred to O_2 , the agent will prefer *Package-A* = $\{[O_1 \text{ if } X, O_2 \text{ else}], [O_1 \text{ if } Z, O_2 \text{ else}]\}$ to *Package-B* = $\{[O_1 \text{ if } Y, O_2 \text{ else}], [O_1 \text{ if } Z, O_2 \text{ else}]\}$ but prefer *Wager-B* = $[O_1 \text{ if } Y \vee Z, O_2 \text{ else}]$ to *Wager-A* = $[O_1 \text{ if } X \vee Z, O_2 \text{ else}]$. This is inconsistent because the packages are just the corresponding wagers redescribed.

While this diagnosis is correct as far as it goes, it does not reveal what is wrong with the agent's *beliefs*, since the inconsistency holds among *preferences*.²³ Some depragmatizers hope to go further by showing that inconsistent preferences are invariably accompanied by inconsistent beliefs, specifically by inconsistent beliefs about the values of prospects.²⁴ In an attempt to "depragmatize" the DBA, Colin Howson and Peter Urbach *define* a person's degree of belief in X as the betting odds she *believes* to be fair, and go on to emphasize that "believing certain odds fair does not in any way imply that you will accept bets at those odds or even at any greater odds. To believe odds to be fair is to make an intellectual judgment, not . . . to possess a disposition to accept particular bets when they are offered" (1989, 57). Thus, someone who strictly (weakly) prefers A to B is thereby committed to the belief that it would be advantageous (fair) to have A rather than B . By the same token, a person who is more confident in X than in Y is committed to believing that *Package-A* will be more advantageous to her than *Package-B*. Given this, Howson and Urbach argue the DBA shows that probabilistically inconsistent beliefs lead to *logically* inconsistent beliefs about the values of wagers. For example, an agent who is more confident of X than Y and more confident of $Y \vee Z$ than $X \vee Z$ will inconsistently believe that it is in her interest both to have *Package-A* over *Package-B* and to have *Wager-B* over *Wager-A*.²⁵

Unfortunately, the Howson/Urbach approach is still infected with pragmatism. There is an inferential gap between the probabilistic inconsistency of beliefs about nonevaluative propositions and the logical inconsistency of value judgments about prospects involving these propositions.²⁶ When beliefs justify value judgments, the sort of "justification" at work is not purely epistemic; it invariably

hinges on substantive principles of *practical* rationality. For example, the fact that the agent is more confident in X than in Y only “justifies” the judgment that Package-A is more desirable than Package-B if Savage’s Sure Thing Principle is assumed. In reality, one cannot *deduce* inconsistencies in an agent’s beliefs from inconsistencies in her preferences since any such inference will be mediated by the (nonlogical) principles of *practical* reasoning that relate beliefs to preferences. Whatever defects there are in the agent’s beliefs are still, at root, pragmatic.

A truly epistemic rationale for Probabilistic Consistency must explain how violations of the laws of probabilities impede the *accuracy* of beliefs. Bas van Fraassen and Abner Shimony have argued, in different ways, that violating these laws can lead people to make poorly *calibrated* estimates of relative frequency.²⁷ Additivity requires a person who assigns a degree of belief to each proposition in a set $X = \{X_1, X_2, \dots, X_n\}$ to estimate the frequency of truths in X as $[c(X_1) + c(X_2) + \dots + c(X_n)] \div n$. One evaluates the accuracy of such estimates using a quantity called the *calibration index*.²⁸ For each $1 > a > 0$, let X_a be the set of propositions to which c assigns value a , let n_a be the cardinality of X_a , and let $\alpha(X_a)$ be the proportion of truths in X_a . Then c ’s calibration index is given by

$$\text{Cal}(c) = \sum_a n_a \times (\alpha(X_a) - a)^2.$$

c is *well-calibrated* to the extent that it minimizes this quantity. When c is perfectly calibrated, half the propositions assigned probability $1/2$ are true, two-fifths of those assigned probability $2/5$ are true, and so on. Van Fraassen and Shimony each use the calibration index to measure the “fit” between beliefs and the world, and each shows that (under fairly restrictive conditions) beliefs that violate the laws of probability are necessarily less well calibrated than they could otherwise be. These are interesting results, but they do not justify probabilistic consistency because *calibration simply is not a reasonable measure of the “fit” between graded beliefs and the world*. To state just one problem, Joe can be better calibrated than Jane even though Jane *always* believes truths more strongly, and falsehoods less strongly, than Joe does.²⁹

Another approach, pursued in Joyce 1998, relates probabilistic consistency directly to the *accuracy* of graded beliefs. The strategy here involves laying down a set of axiomatic constraints that any reasonable gauge of accuracy for confidence measures should satisfy, and then showing that probabilistically inconsistent measures are always less accurate than they need to be. The two most important constraints are

Normality. If c ’s values are always at least as close to the actual truth-values of propositions as c^* ’s are, then c is at least as accurate as c^* .

Convexity. If c and c^* are equally accurate (and not identical), then their *even mixture* $[c + c^*] \div 2$ is strictly more accurate than either.

Normality connects accuracy to truth. *Convexity* penalizes overconfidence (so that, e.g., believing X and $\neg X$ both to degree $1/2$ is better than believing them both to degree 1). On the basis of these axioms, and a few others, it can be proved that if c violates the laws of probability then there is a probability function c^+ that is strictly more accurate than c *under every logically consistent assignment of truth-values to propositions*. To the extent that one accepts the axioms,³⁰ this shows that the demand for probabilistic consistency follows from the purely epistemic requirement to hold beliefs that accurately represent the world.

3. BAYESIAN CONFIRMATION THEORY

The most influential aspect of Bayesian epistemology is its theory of *evidential support*. Bayesians reject the idea that evidential relations can be characterized in an objective, belief-independent manner; evidence is always relativized to a person and her opinions. On this view, a person’s *total, nonincremental* evidence regarding a hypothesis H is directly reflected in her level of confidence in H .³¹ This evidence derives from two sources: (a) the person’s own subjective “prior” opinions about the intuitive plausibility of H and other propositions, and (b) any new knowledge she has acquired via learning. She is more confident in H than in $\neg H$ exactly if the totality of her prior and learned evidence tells more strongly in favor of H than against it. Similarly, her level of confidence in H conditional on E reflects her *total* evidence for H when E added to her stock of knowledge. The *disparity* between her unconditional level of confidence in H and her level of confidence in H given E then captures the amount of *additional* evidence that E provides for H . This justifies the following qualitative account of *incremental* evidential support:

E *confirms* (*disconfirms*, is *irrelevant* to) H for a person whose beliefs are represented by the measures in Con if and only if $c(H/E)$ exceeds (is exceeded by, equal to) $c(H)$ for every $c \in Con$.

Bayesians use this analysis to provide explanations of important truisms about evidential relationships and facts about scientific practice.³² Here are two examples that both follow from Bayes’s Theorem:

Prediction Principle. If a person is more confident in E conditional on H than conditional on $\neg H$, then E confirms H for her.

Surprise Principle. If a person is equally confident in E and E^* conditional on H , then E confirms H more strongly for her than E^* does (or disconfirms it less strongly) if and only if she is *less* confident of E than of E^* .

The prediction principle provides a Bayesian rationale for the *hypothetico-deductive model of confirmation*. On the H-D model, hypotheses are incrementally confirmed by any evidence statements they *entail*. Bayesians are able to make sense of this insight even though they reject the idea that evidential relations can be characterized independent of belief. When H entails E , the Prediction Principle entails that E (incrementally) confirms H for anyone *who does not already reject H or accept E* . Consequently, the results of experiments that fit the H-D paradigm will be deemed evidentially relevant by *anyone* who has not yet made up his mind about the data or the hypothesis. While the *degree* of confirmation will vary across people, every rational person will agree that the hypothesis is supported by the data to some degree.

The surprise principle explains why *unexpected* evidence seems to have more “confirming potential” than evidence that is already known. On the Bayesian view, a person’s appraisal of E ’s evidential import for H varies *inversely* with her confidence in E when $c(E/H)$ is held fixed. In particular, if H entails E , so that $c(E/H) = 1$, then E confirms H more strongly the less likely it is.

Some see this as a double-edged sword. According to Clark Glymour (1980, chap. 3), it entails that “old” evidence about which a person is already certain (or very highly confident) cannot provide any (much) support for any hypothesis. When $c(E)$ is close to 1, $C(H)$ will be close to $c(H/E)$, and this detracts from E ’s confirming power. As Glymour emphasizes, however, many splendid pieces of scientific reasoning involve adducing “old” evidence to support novel hypotheses. Many aspects of this *Problem of Old Evidence* pose challenges for Bayesianism, but the most serious is to explain how a person who is certain or almost certain of E , and is fully aware of all the relevant logical relationships between H and E , can see E as evidence for H .³³

Such an explanation can be given once we realize that the Bayesian tent has room for more than one notion of evidential support. The point is easiest to see when put in quantitative terms. Many Bayesians use the *difference measure* $d(H, E) = c(H/E) - c(H)$ to capture the *degree* to which E incrementally supports H . Since this goes to zero as $c(E)$ approaches one the Problem of Old Evidence clearly arises for d . There is, however, a closely related function, $d^*(H, E) = c(H/E) - c(H/-E)$, that lacks this undesirable feature: as long as $c(H/E)$ and $c(H/-E)$ are defined, $c(E)$ can range from zero to one without affecting the value of $d^*(H, E)$. The suggestion, made in Joyce 1999 and Christensen 1999, is that d^* captures the sense of confirmation at issue in old evidence cases. It compares total evidence that a person will have for H if E is added to her stock of knowledge with the total evidence she will have for H if $-E$ is added. Since this comparison does *not* depend

on her level of confidence in E , it can remain fixed even as her beliefs about E and H change. Thus, a person who is certain of E , and who is fully aware of all the relevant logical relationships between H and E , can still see E as evidence for H in the sense of regarding H as more likely if E is true than if E is false.

It must be emphasized that d^* is *not* being proposed as a replacement for d .³⁴ Rather, d remains the basic measure of incremental confirmation, and d^* captures one of its *components*. Writing $d(H, E) = c(-E) \times d^*(H, E)$ makes it clear that d^* is a part of d that does *not* depend on E ’s prior probability. There are many questions about confirmation that d can be used to answer, but d^* is useful for answering others. We should use d when we want to say how much E supports H relative to a common background of knowledge concerning E . d^* is useful when we are talking about E ’s confirming power *across* states of knowledge about E in which $c(H/E)$ and $c(H/-E)$ are fixed. (This turns out to be a fairly common case.) Hence, there is no *conflict* between d and d^* : they measure different things and are used to answer different questions.

There are other measures of confirmation as well. The most important is the *likelihood ratio* $c(E/H) \div c(E/-H)$, which expresses the disparity between $c(H/E)$ and $c(H)$ as the ratio of H ’s *odds* given E to its unconditional *odds*.³⁵ This measure is useful when there is consensus about how strongly H and $-H$ predict E , but none about the probabilities of H and E . Here too it is best to be ecumenical. There is no single right way to measure confirmation; different measures are suited for different pragmatic purposes. To tell the *whole* story about a person’s evidential situation vis-à-vis H and E , we would need to describe her *entire* system of beliefs. Fortunately, the whole story rarely interests us. We use different senses and measures of incremental confirmation to illuminate those parts of it that do.

That said, it is possible to isolate an element that is common to any theory of evidence that deserves the name *Bayesian*. The core tenet of any such theory is the observation, encapsulated in Bayes’s Theorem and the Prediction Principle, that *E ’s confirming potential relative to H is enhanced to the extent that H makes E more probable*. It follows from this that the ability of evidence to *discriminate* among competing hypotheses is limited by the relative degree to which these hypotheses probabilify the evidence. Imagine a believer who has more *total* evidence for H than for H^* , so that $c(H) > c(H^*)$. What would it take to reverse this situation by adding E to her knowledge, so that $c(H^*/E) > c(H/E)$? Bayes’s Theorem tells us that E can reverse the “balance of evidence” only if H^* predicts E ’s truth more strongly than H ’s does. This leads to the following:

Discrimination Principle. If a person initially has more total evidence for H than for H^* , and if H predicts E at least as strongly as H^* does, then the person will have more total evidence for H than for H^* after E is added to her stock of knowledge.

Note that E can never reverse the balance of total evidence between H and H^* when both hypotheses entail E . This observation will play a central role in our discussion of the Bayesians' account of learning.

Like the rest of Bayesian confirmation theory, the Discrimination Principle rests on the assumption that a person's total evidence is a combination of her "prior" views about the intuitive plausibility of hypotheses and information she has acquired via learning. So, to fully understand Bayesianism, we need some appreciation of its conception of "prior" opinions and its theory of learning. We begin with the latter.

4. THE BAYESIAN THEORY OF LEARNING

Bayesians see learning as a process of *belief revision* in which "prior" beliefs are replaced by "posterior" beliefs that incorporate new information. For ease of exposition we will assume an ideal learner whose prior and posterior opinions can be represented by probability functions c_0 and c_1 . Bayesian learning proceeds in two stages: one causal, one inferential. First, the person has a *learning experience* in which perception, intuition, memory, or some other *noninferential causal process* immediately alters some subset of her beliefs. Second, she uses information acquired via this experience, in conjunction with other things she knows, to revise the rest of her opinions. We will focus on experiences whose sole *immediate* effect is to alter the person's level of confidence in a proposition E ,³⁶ so that c_1 is constrained to be such that $c_1(E) \neq c_0(E)$. Every other change in the learner's posterior beliefs is due to a (probabilistic) inference from this basic one. The challenge is to explain how her *prior* opinions can justify the choice of a specific posterior c_1 from among the many that might meet the constraint.

In the simplest learning experiences, where the person becomes *certain* of E , Bayesians advocate the following rule of belief revision:

Simple Conditioning. If a person with a "prior" such that $0 < c_0(E) < 1$ undergoes a learning experience whose only immediate effect is to raise E 's probability to one, then $c_1(H) = c_0(H/E)$ for any proposition H .

The effect of this is to set probabilities of hypotheses inconsistent with E to zero and to increase probabilities of hypotheses that entail E uniformly by a factor of $1/c_0(E)$.

Simple conditioning requires a learner to become *certain* of E 's truth. As Richard Jeffrey (1983, 164–69) has long argued, however, our evidence is typically

too vague and imprecise to justify such "dogmatism." More realistic learning experiences are modeled by the following rule:

Jeffrey Conditioning. If a person with a "prior" such that $0 < c_0(E) < 1$ undergoes a learning experience whose only immediate effect is to set $c_1(E) = q$, then $c_1(H) = q \times c_0(H/E) + (1-q) \times c_0(H/\neg E)$ for any proposition H .

This holds probabilities conditional on E and $\neg E$ fixed, so that $c_1(H/E) = c_0(H/E)$ and $c_1(H/\neg E) = c_0(H/\neg E)$, and it multiplies the probability of each hypothesis that entails E (or $\neg E$) by a factor of $q/c_0(E)$ (or $(1-q)/c_0(\neg E)$). Note that Jeffrey Conditioning reduces to Simple Conditioning when $q = 1$.

Many justifications have been offered for the conditioning rules. The most contentious are the *diachronic Dutch book arguments*. Like their synchronic counterparts, these are *pragmatic* arguments designed to show that failing to condition on one's evidence can lead to practical incoherence. Here, though, the self-defeating choices are made at *different* times. David Lewis offered a Dutch book rationale for Simple Conditioning, Brad Armendt generalized it to cover Jeffrey Conditioning, and Bas van Fraassen and Michael Goldstein each showed that both forms of conditioning are instances of the *Reflection Principle*, which also has a Dutch book rationale.³⁷ Our discussion will focus on Reflection since it is the most general principle of the bunch.

Reflection requires a person's degree of belief in H at time t_0 to agree with her t_0 -*expectation* of her time t_1 degree of belief in H . If " $c_1(H) = x$ " says the person's degree of belief in H at time t_1 is x , then Reflection requires that $c_0(H/c_1(H) = x) = x$ and therefore that

$$c_0(H) = \sum_x c_0(c_1(H) = x) \times x.^{38}$$

This captures what is at issue in diachronic Dutch Book arguments, since, as van Fraassen and Goldstein each showed, a person is invulnerable to diachronic Dutch book *if and only if* she satisfies Reflection.

As many authors have recognized,³⁹ however, there is nothing irrational, *per se*, about violating Reflection or leaving oneself open to diachronic Dutch books. In fact, practical rationality *requires* it when one undergoes an *antilearning experience*, that is, a belief change one takes to be *unreliable*. Suppose a miserly expected utility maximizer has inadvertently ingested a drug that will soon make her highly confident of the claim, H , that she will be able to recite the Gettysburg Address backward on command. She now regards H as unlikely, and does not take the fact that she will soon believe it to be any indication of its truth. What should she do? Since she knows that she will soon pay close to \$1 to buy the wager [\$1 if H , \$0 else], the best way for her to offset her future idiocy is to hedge her bets now by selling the wager for as much as she can get above its current

(low) fair price. Even though she is sure to lose money in the aggregate, this is perfectly rational. The irrational thing would be to allow the beliefs of her future, idiot self to guide her present actions.

It is irrational to violate Reflection *only* when one regards the belief change one is about to undergo as a genuine *learning experience* that is likely to improve the overall accuracy of one's beliefs. But, what is it to *regard* a belief change as a learning experience? As Brian Skyrms (1993) has argued, Reflection itself provides the answer. A person sees a prospective change in her belief about H as increasing accuracy exactly if she sees her future degree of belief in H as a reliable indicator of H 's truth-value. When the evidential connection is perfect, her beliefs will satisfy $c_0(H/c_1(H) = x) = x$ for all x . So, *by definition*, a person regards a belief change as a genuine learning experience only if it satisfies Reflection. This makes Reflection (or invulnerability to diachronic Dutch books) entirely useless as a *justification* for the conditioning rules. Indeed, when *properly* formulated (as here), these rules take the shift from $c_0(E)$ to $c_1(E)$ *at face value as a learning experience*.

A further problem with diachronic Dutch book arguments is their *pragmatic* character. Even if conditioning *pays*, antipragmatists will still wonder whether it leaves learners better off from a purely epistemic perspective. One can argue that it does by showing that conditioning rules are "epistemically conservative" in that they produce a posterior that *departs minimally* from the prior while taking the evidence into account. The best results along these lines are due to Persi Diaconis and Sandy Zabell (1982), who show that, on a variety of ways of measuring differences between probabilities, simple and Jeffrey conditioning *uniquely* minimize change subject to the constraints imposed by experience. A more general approach, still grounded in epistemic conservatism, justifies the conditioning rules by showing that they alone keep central features of the prior intact. Consider the property

Ordinal Invariance. If $c_1(E) > 0$ then, for any hypotheses H and H^* , $c_1(H \& E) > c_1(H^* \& E)$ iff $c_0(H \& E) > c_0(H^* \& E)$.

This requires a person who has a learning experience involving E to retain her views about the *comparative* probability of propositions that entail E . It can be shown that *only* the conditioning rules generally meet this condition.⁴⁰ Thus, we will have a sound epistemic rationale for conditioning if we can make a convincing case for Ordinal Invariance.

The Discrimination Principle of section 3 gives us the resources to do so. It entails that a person who initially has more total evidence for $H \& E$ than for $H^* \& E$ will still have more total evidence for $H \& E$ after undergoing a learning experience that alters E 's probability. As already noted, a rational believer will invest more prior (posterior) confidence in $H \& E$ than in $H^* \& E$ just in case her total evidence for the former at t_0 (at t_1) exceeds her total evidence for the

latter. It follows directly from Discrimination that c_1 will rank $H \& E$ above $H^* \& E$ if and only if c_0 does. Hence, Discrimination implies Ordinal Invariance. Simple and Jeffrey conditioning are thus justified as the only belief revision rules that are consistent with this most basic tenet of the Bayesian theory of evidence.

5. PRIOR PROBABILITIES: THE CHARGE OF SUBJECTIVISM

The most persistent objection to Bayesianism is that it engenders an untenable *subjectivism* in which all manner of preposterous beliefs and ludicrous inferences are immunized from criticism. If the constraints on rational opinion begin and end with the laws of probability and the conditioning rules, then Bayesianism allows a person to draw almost any conclusion on the basis of almost any evidence as long as she starts out with suitable prior beliefs. There are, for example, probabilistically consistent priors that make the existence of statues on Easter Island evidence for the conclusion that Martians built the Pyramids, or that count peyote-induced belief changes as learning experiences. By allowing such absurdities, Bayesianism seems to say that what it makes sense to believe, and what counts as evidence for what, is "all just a matter of opinion." In Bayesian epistemology, it appears, just about anything goes.

Bayesians have addressed this charge in a variety of ways. "Personalists," like Savage and de Finetti, bite the bullet and deny that there are any constraints on beliefs that outrun the laws of probability and the conditioning rules. They seek to blunt the force of the "anything goes" objection by arguing that the effect of a person's priors will tend to diminish as she acquires increasingly more evidence about the world. Here is a famous statement of the view: "If observations are precise . . . then the form and properties of the prior distribution have negligible influence on the posterior distribution. From a practical point of view, then, the untrammelled subjectivity of opinion . . . ceases to apply as soon as much data becomes available. More generally, two people with widely divergent prior opinions but reasonably open minds will be forced into arbitrarily close agreement about future observations by a sufficient amount of data."⁴¹ This "merger of opinion" is what stands in for objectivity in the personalist picture: objectivity is *intersubjective agreement* in the long run.

The theoretical basis of such claims is found in a set of mathematical results, the "washing out" theorems, which show that people who begin with different priors will tend to reach consensus as they acquire increasingly more data. Sup-

pose we have two subjects with priors c and c^* that assign H an intermediate probability. Imagine further that there is an infinite sequence of evidence statements $\{E_1, E_2, E_3, \dots\}$ such that for each j :

- (a) c and c^* assign each finite data sequence $D_j = \pm E_1 \& \pm E_2 \& \dots \& \pm E_j$ an intermediate probability.⁴²
- (b) At time- j each subject has an experience that sets E_j 's probability to 1 or to 0.
- (c) Both subjects regard these as *learning* experiences.
- (d) Both subjects condition on the evidence they receive, so that for each j and each finite data sequence D_j , $c_j(H) = c(H/D_j)$ and $c_j^*(H) = c^*(H/D_j)$.

(a)–(d) entail that the probabilities each subject assigns to H at successive times form a *martingale sequence* in which each term is the expected value of its successor, so that $c_j(H) = \sum_x c_j(c_{j+1}(H) = x) \times x$. The *Doob Martingale Convergence Theorem* then entails that, with probability one, $c_j(H)$ and $c_j^*(H)$ each converge to a definite limit.⁴³

Establishing that these limits coincide requires an additional assumption. There are two kinds of conditions that will do the job. First, the evidence can be so informative that it forces *any* rational believer to the same conclusion about H . Either of the following clauses will accomplish this:

- (e) Each possible data sequence $\pm E_1 \& \pm E_2 \& \pm E_3, \dots$ entails H or $\neg H$.
- (e*) Each data sequence determines an *objective probability* for H .

On the basis of Coherence alone, (e) guarantees that $c_j(H)$ and $c_j^*(H)$ will converge to H 's truth-value (as specified by the data). Similarly, if we combine (e*) with what David Lewis (1980) calls the "Principal Principle," which requires rational believers to align their degrees of belief with known objective probabilities,⁴⁴ then $c_j(H)$ and $c_j^*(H)$ will converge to H 's objective probability (as specified by the data).

For purposes of allaying fears about subjectivism, these ways of obtaining agreement in the limit are wholly impotent. Priors only get "washed out" because the data is so incredibly informative that, as a matter of logic, it makes each subject's antecedent beliefs *irrelevant* to her final view. Indeed, a "washing out" result that uses (e) (or [e*]) amounts to little more than the claim that two subjects will come to agree about H 's truth-value (or probability) if each ultimately learns H 's truth-value (or its objective probability). If all learning situations were like this, there would be no need for prior probabilities at all.

To obtain washing-out theorems that respond to subjectivist worries, we must imagine that nothing in the laws of logic or probability forces subjects to assign a specific probability to H conditional on each data sequence. We must, rather,

derive joint convergence from commonalities among the priors alone. One way to do this, pioneered by Savage ([1954] 1972, 46–50), is to have subjects agree that the evidence statements are *statistically independent and identically distributed* conditional on H and $\neg H$.

- (e**) For some constants x and y and all j and k , $c(E_k/H) = c(E_k/H \& E_j) = x$ and $c(E_k/\neg H) = c(E_k/\neg H \& E_j) = y$. The same holds with c replaced by c^* .

Savage showed that under these conditions $c_j(H)$ and $c_j^*(H)$ converge to the same value with probability one (according to both c and c^*).

There is less here than meets the eye. (e**) requires an *exceptional* amount of initial consensus between the parties. In effect, both must start out treating H like a chance hypotheses, so that $H = "x$ is the objective probability of every $E_j,"$ and each must see the E_j as describing independent trials of a chance process. While this does happen in rare, well-behaved cases (e.g., coin flipping), it typically fails. Once again, the assumptions needed to ensure convergence to a common value severely limit the theorem's value as a response to the charge of subjectivism.

Indeed, it is hard to see how any result based on (a)–(d) can rebut subjectivism. These assumptions require *substantial* initial agreement among the subjects. (a) asks them to agree about which data sequences have a chance of occurring. (c) requires them to see the belief changes they are about to undergo as learning experiences. If they disagree about these things, they will *not* tend toward consensus as the evidence accumulates simply because they will not agree about what counts as "the evidence." The basic problem with using "washing out" results to refute subjectivism is simple: no agreement in, no agreement out!

Another approach, championed by so-called *objective* Bayesians, seeks to avoid subjectivism by restricting prior probabilities. The best-developed view of this sort is that of the physicist E. T. Jaynes, who writes, "The most elementary requirement of consistency demands that two persons with the same relevant prior information should assign the same prior probabilities. Personalistic doctrine makes no attempt to meet this requirement."⁴⁵ According to Jaynes, any well-posed problem of inductive inference is defined by certain *objective constraints*, often deducible from physical theory or symmetry principles, which fix *expected values* for various quantities. An *acceptable* prior must yield the required expectations. If the expected value of the toss of a die is constrained to be 3.5, say, then $c(1) \times 1 + c(2) \times 2 + \dots + c(6) \times 6 = 3.5$ will hold for every acceptable prior c . To choose the correct prior from among the many that are acceptable, Jaynes advocates *entropy maximization*. Relative to a partition $\{X_1, X_2, \dots, X_k\}$ of mutually exclusive, collectively exhaustive propositions, the entropy in a probability c is defined as

$$\text{Entropy}(c) = \sum_j c(X_j) \times \log(c(X_j)).$$

Entropy(c) measures the inverse of the amount of *information* (about the X_i) that c encapsulates. For the die above it is easy to show that the uniform distribution, in which each side comes up with probability $1/6$, uniquely maximizes entropy. By choosing the (unique) probability that maximizes entropy, Jaynes argues, we respect the constraints but otherwise make the fewest possible additional assumptions.

While maximizing entropy is a fine method for finding a probability that meets specific constraints, the idea that it "objectifies" Bayesianism is illusory. As Colin Howson and Peter Urbach (1989, 289) point out, the choice of one system of constraints rather than another is a subjective matter par excellence: "No prior probability . . . expresses merely the available factual data; it inevitably expresses some sort of opinion about the possibilities consistent with the data. Even a uniform probability distribution is [uniform] only relative to some partition of these possibilities: we can always find another with respect to which the distribution is as biased as you like—or don't like. Jaynes's objective priors do not exist." Even if we all agree that the expected value of a toss of a die is 3.5, we will be justified in settling on the uniform distribution only if we also agree that this is our *only* relevant piece of information. This is a *subjective* judgment, as is any other of the form "precisely *these* constraints characterize our knowledge." If you think that odd tosses are more likely than even ones, and I think the reverse, then we would both be wrong to ignore our beliefs and settle on the uniform distribution. We might agree with Jaynes that "consistency demands that two persons with the same relevant prior information should assign the same prior probabilities," but we will disagree about which "prior information" counts as relevant and even about what this information is. Appealing to a physical theory will not help matters unless we are already convinced of it, which, again, is a matter for our respective background beliefs. Jaynes's program provides no answer to the charge of subjectivism.

Perhaps Bayesians should admit that there is more to epistemic rationality than the laws of probability and the conditioning rules. These provide an *internalist logic* of rational belief: they tell us whether a person's beliefs cohere with one another, what she counts as evidence for what, and how she should revise her opinions in light of what she regards as learning experiences. The whole Bayesian apparatus, in other words, is appropriate for describing and criticizing a person's *own reasons* for believing what she does.

Any internalist view is going to suffer from "garbage in, garbage out" problems. If a person starts out with inaccurate priors that assign low probabilities to truths and high ones to falsehoods, or if she has false views about which belief changes are learning experiences, then her subsequent beliefs will be inaccurate and unreliable as well. Sophisticated Bayesians should concede this and grant that there are further *externalist* principles of epistemic rationality that can be used to evaluate opinions. These principles will not assess a person's beliefs on the basis

of *her own* view of things—that is, they will not take her prior opinions or her views about learning at face value as the personalists do. Rather, they will consider the actual accuracy of her priors and the actual reliability of the processes she treats as learning experiences. Ramsey saw this right from the start. In addition to being concerned with *rational* belief, which answers to internalist Bayesian norms, he sought a theory of *reasonable* belief that would assess a person's doxastic attitudes and habits on the basis of the actual accuracy of the beliefs they generate.⁴⁶ Personalist Bayesians have largely ignored this aspect of Ramsey's thought, but they would be wise to give it more heed.

One should not think of these externalist approaches *replacing* Bayesianism. A complete account of epistemic rationality will necessarily involve both internalist considerations that concern a believer's reasons and externalist considerations having to do with the accuracy of her opinions and the reliability of her belief-forming processes. While Bayesians have little to say about the latter issues, they offer us a detailed, systematic, and exceedingly plausible account of the former. For all its shortcomings, *Bayesianism remains without peer as a theory of epistemic reasons and reasoning*. As long as we use it for this purpose it will serve us well.

NOTES

1. Classic sources are Ramsey 1931, de Finetti (1937) 1964, and Savage (1954) 1972.
2. The strength of a belief should not be confused with any *feeling of conviction*. As Ramsey (1932, 169) noted, the beliefs we hold most strongly are often associated with no feelings whatever. Moreover, people who *feel* convinced of propositions sometimes reason and act as if they are false. Nor is a graded belief a categorical belief about an *objective probability*. A person can be *certain* that the coin he is about to toss is either two-headed or two-tailed, and yet be maximally uncertain about which possibility obtains. He then believes "the coin will come up heads" to degree $1/2$ even though he knows that $1/2$ is *not* its objective probability. In general, a person's degree of confidence in a proposition is her *subjective expectation* of its objective probability.
3. See Rényi 1955, Harper 1976, Spohn 1986, McGee 1994, van Fraassen 1995, and Hammond 1994.
4. See Levi 1980, 85–91, and Kaplan 1996, 27–31.
5. P must also be *countably* additive. The issues surrounding countable additivity are too involved to pursue here. See Seidenfeld and Schervish 1983 and Kaplan 1996, 32–36.
6. Conditioning entails that $P(X/Y) = P(X \& Y) \div P(Y)$ when $P(Y) > 0$.
7. Bayes (1764) 1958.
8. See Osherson 1995 and Shafir and Tversky 1995.
9. Ramsey is especially lucid on this point in 1931, 173.
10. It need not be any part of Bayesianism that acting on the basis of habits, emo-

tions, and so on is irrational. If an act bears the right relationship to the actor's beliefs and desires, then it is rational however it is caused. When emotions or habits tend to lead (nonaccidentally) to rational actions, Bayesians should encourage emotional or habitual decision making.

11. The value of $c(X)$ is independent of the particular choice of a and b . For any $a^* > b^*$, the fair price f^* of $W^* = [\$a^* \text{ if } X, \$b^* \text{ else}]$ will be such that $c(X) = (f^* - b^*) \div (a^* - b^*)$.
12. Conditional beliefs are reflected in fair prices of bets that get "called off" when the condition fail to obtain.
13. Some commentators incorrectly portray the "package principle" as an added, hidden premise in the DBA. Actually, it follows from the EU-thesis and (a)-(c).
14. See Kaplan 1996, 23-31; Jeffrey 1992, 82-85; Joyce 1999, 43-45.
15. The proof of this claim, which is beyond the scope of this paper, relies on technical results found in Kraft, Pratt, and Seidenberg 1959 and Scott 1964.
16. Most Bayesians regard the DBA as a "toy model" whose rhetorical purpose is to point beyond itself to these more serious representation results. See Skyrms 1984b, Jeffrey 1992, Maher 1993, Kaplan 1996, and Joyce 1999, among others. This has *always* been the prevailing wisdom among Bayesians. Indeed, Ramsey only mentions the Dutch Book Theorem in passing *after* having proved a representation theorem.
17. Despite this terminology, Savage's acts are *not* best understood as events the agent can directly control. See Joyce 1999, 61-62, and 107.
18. To establish uniqueness Savage imposed a variety of constraints (like Trichotomy and Wagers) that require extremely *rich* systems of preferences. Most Bayesians now agree that these *richness conditions* far exceed what is demanded by practical rationality. Accordingly, Savage's result is best thought of as guaranteeing the existence of a (large) family of probability/utility pairs, all of whose associated expectations represent the agent's preferences.
19. See Broome 1991 and Maher 1993 for overviews.
20. For similar worries, see Rosenkrantz 1981, 214; Joyce 1998, 584-86; and Christensen 2001, 356-64.
21. Skyrms claims, rightly, that Ramsey too saw this as the deep flaw that incoherent preferences serve to indicate. Armendt (1993, 3) defends a similar view.
22. See Tversky and Kahneman 1981.
23. See Joyce 1998, 586.
24. See Howson and Urbach 1989, Christensen 1996, Hellman 1997.
25. Christensen (2001) seeks to reinterpret representation theorems in an analogous way.
26. Compare Maher 1997.
27. See van Fraassen 1983 and Shimony 1988.
28. For useful discussion of the calibration index, see Murphy 1973.
29. See Joyce 1998, 494-95. For other problems, see Seidenfeld 1985.
30. Maher (2002) expresses reservations about *Convexity*, and champions a nonconvex accuracy gauge.
31. While Bayesians rarely discuss *total* evidence, their theory of incremental evidence makes sense only when it is based on an account of total evidence like the one given here.
32. For useful discussion, see Earman 1992, chap. 3.

33. Other aspects of the problem concern how to characterize the knowledge of logical relationships within the Bayesian framework, and how to handle genuinely novel hypotheses. See Joyce 1999, 204, and Earman 1992, 120-35, for relevant discussion.
34. This runs contrary to the line taken in Eells 2000.
35. H 's odds are $c(H) \div c(-H)$, and its odds conditional on E are $c(H/E) \div c(-H/E)$. Odds talk can be translated into probability talk using the mapping $c(H) = \text{odds}(H) \div (1 + \text{odds}(H))$.
36. More complicated experiences might alter her levels of confidence in each element of a set of statements, or alter the conditional probability of some proposition given another, or directly fix the value of a random variable.
37. Lewis's result is reported and discussed in Teller 1973. Armendt 1980, van Fraassen 1984, and Goldstein 1983 contain the generalizations.
38. As van Fraassen notes, it is important to understand that $c_1(H)$ is a *nonrigid* designator since the identity $c_1(H) = x$ is otherwise trivial.
39. Levi 1987; Christensen 1991; Maher 1993, 106-20.
40. Joyce 1999, 195-96.
41. Edwards, Lindeman, and Savage, 1963, 201. Also, see Suppes 1966, 204.
42. $\pm E$ may be E or $\neg E$.
43. Doob 1971.
44. Of course, personalists, like de Finetti and Savage, would deny that objective probabilities exist.
45. Jaynes 1968, 53. See also Jaynes 1994.
46. See Ramsey 1931, 193-96.