

HOW PROBABILITIES REFLECT EVIDENCE

James M. Joyce
The University of Michigan

Many philosophers think of Bayesianism as a theory of practical rationality. This is not at all surprising given that the view's most striking successes have come in decision theory. Ramsey (1931), Savage (1972), and De Finetti (1964) showed how to interpret subjective degrees of belief in terms of betting behavior, and how to derive the central probabilistic requirement of coherence from reflections on the nature of rational choice. This focus on decision-making can obscure the fact that Bayesianism is also an epistemology. Indeed, the great statistician Harold Jeffries (1939), who did more than anyone else to further Bayesian methods, paid rather little heed to the work of Ramsey, de Finetti, and Savage. Jeffries, and those who followed him, saw Bayesianism as a theory of inductive evidence, whose primary role was not to help people make wise choices, but to facilitate sound scientific reasoning.¹ This paper seeks to promote a broadly Bayesian approach to epistemology by showing how certain central questions about the nature of evidence can be addressed using the apparatus of subjective probability theory.

Epistemic Bayesianism, as understood here, is the view that evidential relationships are best represented probabilistically. It has three central components:

Evidential Probability. At any time t , a rational believer's opinions can be faithfully modeled by a family of probability functions c_t , hereafter called her *credal state*,² the members of which accurately reflect her *total evidence* at t .

Learning as Bayesian Updating. Learning experiences can be modeled as shifts from one credal state to another that proceed in accordance with Bayes's Rule.

Confirmational Relativity. A wide range of questions about evidential relationships can be answered on the basis of information about structural features credal states.

The first of these three theses is most fundamental. Much of what Bayesians say about learning and confirmation only makes sense if probabilities in credal

states reflect states of total evidence. It is often said, for instance, that learning one proposition E increases a person's evidence for another X just in case X 's probability conditional on E exceeds X 's unconditional probability. Clearly, this assumes that the unconditional and conditional probabilities in a person's credal state somehow reflect her total evidence.

The aim of this essay is to clarify the thesis of Evidential Probability by explaining how a person's subjective probabilities reflect her total evidence. After a brief discussion of the Bayesian formalism and its epistemological significance, it will be argued that a person's total evidence in favor of any proposition can be decomposed along three dimensions that have rather different probabilistic profiles. The overall *balance* of the evidence is a matter of how decisively the data tells in favor of the proposition. This is what individual probability values reflect. The *weight* of the evidence is a matter of the gross amount of relevant data available. It is reflected in the concentration and stability of probabilities in the face of changing information. The *specificity* of the evidence is a matter of the degree to which the data discriminates the truth of the proposition from that of alternatives. It is reflected in the spread of probability values across a credal state. By appreciating these disparate ways in which probabilities can reflect total evidence we shall come appreciate the richness of the Bayesian formalism, and its importance for epistemology. The central theses of the paper are (i) that any adequate epistemology must be capable of accurately representing the distinctions between the balance, weight and specificity of evidence, and (ii) that only a probabilistic theory is capable of doing this properly.

Many of the points made here have been made by others. Indeed, the distinctions between balance, weight and specificity have all been made before, albeit often in an incomplete or piecemeal way. The novelty here is the integration of these insights into an appealing and coherent probabilistic theory of evidence. We begin with a brief sketch of Bayesianism.

1. Credences as Estimates of Truth-Value.

Any adequate epistemology must recognize that beliefs come in varying *gradations of strength*. Instead of asking whether a person accepts or rejects a proposition outright, we must inquire into her level of confidence in its truth. These confidence levels go by a variety of names—degrees of belief, subjective probabilities, grades of uncertainty—but 'credences' will be the preferred term here. By any name, a person's credence in X is a measure of the extent to which she is disposed to presuppose X in her theoretical and practical reasoning.³

People also have graded *conditional* beliefs that express their degrees of confidence in the truth of some propositions on the supposition that other propositions obtain. It is often said that a believer's credence for X conditional on Y is the credence she would invest in X if she were to *learn* Y . While there is

something right in this idea, it must be handled delicately. A person's credence for X conditional on Y will only coincide with her unconditional credence for X after learning Y when the learning induced belief change is not driven by any arational processes that ignore Y 's content, and when Y encompasses literally everything that the person learns (even the fact that she has learned Y). For current purposes, it is not crucial to get clear about the precise relationship between conditional belief and learning. The essential point is that conditional credences have a clear epistemic interpretation: the epistemic effect of conditioning on Y is to provisionally augment the believer's total evidence by the addition of Y and nothing else.

Graded beliefs help us *estimate* quantities of interest. These can be almost anything: the fair price of a bet, the proportion of balls in an urn, the average velocity of stars in a distant galaxy, the truth-value of a proposition, the frequency of a disease in a population, and so on. Since the values of such quantities often depend on unknown factors, we imagine the believer being uncertain about which member of a given set ω of total contingencies (= possible worlds) actually obtains, and we think of the quantity of interest as a function, or 'random variable', f that assigns each world W in ω a unique real number $f(W)$. The objective in estimation is to come up with an anticipated value f^* for f that is, in some sense, the best possible given the information at hand.

The accuracy of such estimates can be evaluated in a variety of ways. One can employ a *categorical* scale that recognizes only two ways of fitting the facts: getting things exactly right, so that $f^* = f$, or having them wrong. This approach makes no distinctions among different ways of being wrong, so that "a miss is as good as a mile." Alternately, one can use a *gradational*, or "closeness counts," scale that assigns estimates higher *degrees of accuracy* the closer they are to the actual value of quantity being estimated. In (Joyce 1998) it is argued that degrees of belief are principally used to make estimates that are judged on a gradational scale. One can assess the overall quality of a person's credences by considering the accuracy of the estimates they sanction. It is, for example, a flaw in a credence function if it sanctions estimates $f^* > g^*$ when g *dominates* f in the sense that $g(W) \geq f(W)$ for all W in ω .

Different Bayesians construe credences differently depending on the sorts of estimates they tend to consider. The role of credences in estimating utilities of actions is often highlighted. This engenders a Bayesianism that emphasizes the practical virtues of having certain kinds of credences. For example, synchronic "Dutch book" arguments purport to show that one can only avoid choosing strictly dominated acts by having credences that obey the laws of probability. Diachronic Dutch books seek to show, in addition, that one will also be subject to dominance unless one updates by conditioning on the information one receives. In each case the take-home lesson is that defective beliefs spawn defective desires. While it is perfectly legitimate to think in this practical vein, it is crucial to appreciate that (a) credences are used to estimate all sorts of

quantities, (b) they play the same formal role in estimating these quantities as in the estimation of utilities, and (c) for certain purposes it can be illuminating to emphasize their role in estimating quantities that are not so directly tied to actions. With respect to this last point, it is worth noting that (van Fraassen 1983) and (Shimony 1988) have focused on the role of credences in estimating relative frequencies, while (Joyce 1998) emphasizes their role in estimating truth-values (with true = 1 and false = 0). This last approach is best for bringing out epistemologically salient aspects of degrees of belief. We shall, therefore, think of a person's credence in X as being linked to her best estimate of X 's truth-value, where it is understood that such estimates are evaluated on a gradational scale that rewards those who believe truths strongly.⁴

2. Representation of Credences.

Bayesians are often accused of being committed to the existence of sharp numerical degrees of belief. This is not true. The idea that people have sharp degrees of belief is both psychologically implausible and epistemologically calamitous. Sophisticated versions of Bayesianism, as found in, e.g., (Levi 1980, 85–91) and (Kaplan 1996, 27–31), have long recognized that few of our credences are anywhere near definite enough to be precisely quantified. A person's beliefs at a time t are not best represented by any one credence function, but by a *set* of such functions c_t , what we are calling her *credal state*. Each element of c_t is a sharp credence function that assigns a unique real number $0 \leq c(X|Y) \leq 1$ to each proposition X and condition Y .⁵ Each such function defines unconditional credences via the rule $c(X) = c(X|T)$, for T is any logical truth.

Without further ado, we shall assume that all credences in an epistemically rational believer's credal state satisfy the laws of (finitely additive) probability, so that (i) $c(T) = 1$ for T any logical truth, (ii) $c(Y) \geq 0$, (iii) $c(X) + c(Y) = c(X \vee Y) + c(X \& Y)$ for any propositions X and Y , and (iv) $c(X|Y) = c(X \& Y)/c(Y)$ whenever $c(Y) > 0$. Many arguments have been offered for thinking that credences must be probabilistically coherent, but considering them would take us off track. Our question is this: given that credences obey the laws of probability, how do they reflect evidence?

Determinate facts about the person's beliefs correspond to properties that are invariant across all elements of c_t . For example, the person can only be said to determinately believe X to degree x when $c(X) = x$ for every $c \in c_t$, and she is only more confident in X than in Y if $c(X) > c(Y)$ for every $c \in c_t$. Such invariant facts can come in a wide variety of forms. It might be invariant across c_t that a certain quantity has a specific expected value, or that a particular distribution of probabilities has a uniform, binomial, normal, or Poisson form, and so on.

For purposes of epistemology, it is useful to divide the c_t -invariant facts into three classes. Some can be interpreted as *evidential constraints* that are imposed upon the believer by her overall epistemic situation. These will

sometimes be ‘deliverances of experience’ that directly fix facts about credences. To borrow a famous example from Richard Jeffrey (1983, p. 165), looking at a piece of cloth under dim light might lead a believer to assign a credence of 0.7 to the proposition G that it is green, in which case the evidence requires $\mathbf{c}(G) = 0.7$ to be satisfied throughout her credal state. Or, it might be that seeing two pieces of cloth under yellow light leads the person to judge that the first is more likely than the second to be green, so that every \mathbf{c} in \mathcal{c}_r satisfies $\mathbf{c}(G_1) > \mathbf{c}(G_2)$. One can also imagine higher-level evidential constraints. At a given time it might be part of a person’s evidence that a certain test for heroin use has a fifteen percent false positive rate, in which case $\mathbf{c}(+\text{test}|\text{no heroin}) = 0.15$ everywhere in \mathcal{c}_r . In contrast to the constraints that are imposed by the evidence, other \mathcal{c}_r -invariant facts are best interpreted as subjective biases or prejudices. It might, for instance, simply strike the agent as more plausible than not that eight-graders in Cleveland have about eight unmatched socks under their beds. Evidence and prejudice can combine to produce additional \mathcal{c}_r -invariants.

While no attempt will be made here to explain how the invariant features of credal states are divided up into evidential constraints or subjective biases, a few sketchy remarks might allay confusion. Bayesians are often portrayed as radical subjectivists who reject any meaningful epistemic distinction between evidence and biases. On a subjectivist picture, a person’s biases merely reflect her ‘prior’ judgments of credibility about various propositions, while her evidence is the ‘posterior’ information she gains from experience. This suggests a model in which a person starts off with a prior probability \mathbf{c}_0 that reflects her initial judgments of credibility (sophisticated treatments make this a set of priors), and learning proceeds by updating the prior in light of data. In the simplest case where the data specifies that each of the propositions E_1, E_2, \dots, E_n is true, the posterior \mathbf{c}_1 arises from the prior by simple conditioning, so that $\mathbf{c}_1(-) = \mathbf{c}_0(-|E_1 \ \& \ E_2 \ \& \ \dots \ \& \ E_n)$. Priors are required in this process, it is claimed, in order to get inductive reasoning off the ground. So, according to subjectivist Bayesians, a person’s total evidence in favor of a proposition X will encompass both the ‘posterior’ beliefs that she comes to have as the result of learning experiences as well as her ‘prior’ opinions about the intrinsic credibility of various propositions, including X itself.

While this fairly characterizes the views of some Bayesians, the probabilistic approach to epistemology is compatible with the existence of an objective distinction between evidence and bias. Different Bayesians will surely draw the line differently. Some might restrict the class of evidential beliefs to those that reflect observed relative frequencies or known objective chances. Others might go reliabilist and argue that a person’s evidence is found in those invariant features of her credal state that were produced by belief-forming mechanisms that assign high credences to truths and low credences to falsehoods. Others might claim that some constraints are just ‘given’ in experience. There are other options as well: indeed, almost everything epistemologists have had to say about the nature of evidence and be imported into the Bayesian framework.

For present purposes, it does not much matter how one draws the line between evidence and bias, or on which side subjective judgments of credibility lie. What is important is that at any time there should be some set of constraints \mathcal{E}_t that specify those invariant features of a person's credal state that are directly imposed by her evidence. In the examples we consider \mathcal{E}_t will never be anything fancy: it will consist in information about the distribution of objective probabilities over some set of hypotheses about objective chances, and a specification of truth values for data propositions. The goal is to use such simple cases to come to understand how the evidence in \mathcal{E}_t is reflected elsewhere in the believer's credal state. To help us focus on essentials, we shall confine our attention to the ideal case of a person with no biases, so that every invariant feature of \mathcal{C}_t is either an evidential constraint or a consequence of such constraints. This will seem like no restriction to subjectivists, but the more objectively minded will view it as an idealization. Either way, the supposition is needed if we are to isolate those aspects of the believer's credal state that reflect her overall evidential situation.

3. The Distinction Between Balance and Weight.

At an intuitive level, the total evidence for a proposition X is the sum of all those considerations that tell in favor of its truth. Bayesians, and their opponents, have often proceeded as if the total *amount* of evidence for X is directly reflected in X 's credence. When $\mathbf{c}(X) = x$ holds all across \mathcal{C}_t , this amounts to the claim that the number x is a meaningful measure of the total amount of evidence for X . More generally, the view is that (a) the person has more evidence for X than for Y iff $\mathbf{c}(X) > \mathbf{c}(Y)$ over \mathcal{C}_t , (b) she has strong evidence for X iff $\mathbf{c}(X) \approx 1$ over \mathcal{C}_t , (c) E provides the person with incremental evidence for X iff $\mathbf{c}(X|E) > \mathbf{c}(X)$ over \mathcal{C}_t , and so on. This picture of the relationship between credences and evidence is seriously misleading. As we shall see, the total evidence in favor of a hypothesis can be separated into at least three components—balance, weight, and specificity—only one of which is directly reflected in credences.

Let us first distinguish between the *balance* of the evidence, which is a matter of how decisively the data tells for or against the hypothesis, and what J.M. Keynes (1921) called the *weight* of the evidence, which is a matter of the gross amount of data available. Here is Keynes:

As the relevant evidence [for a hypothesis] at our disposal increases, the magnitude of [its] probability may either decrease or increase, according as the new knowledge strengthens the unfavorable or favorable evidence; but *something* seems to have increased in either case—we have a more substantial basis on which to rest our conclusion... New evidence will sometimes decrease the probability of [the hypothesis] but will always increase its 'weight'. (1921, p. 77)

The intuition here is that any body of evidence has both a kind of valence and a size. Its valence is a matter of which way, and how decisively, the relevant data ‘points.’ A body of evidence will often be composed of items of data with different valences that need to be compared. It is this ‘balance of the evidence’ that credences reflect. The size or ‘weight’ of the evidence has to do with how much relevant information the data contains, irrespective of which way it points. As Keynes emphasized, we should not expect the weight of a body of evidence to be reflected in individual credence values. From the fact two hypotheses have the same credence we can infer that the balance of the evidence for each is the same, but we cannot infer anything at all about the relative weights of the evidence in their favor.

To clarify the distinction, it will be useful to consider a simple sampling case.

Four Urns: Jacob and Emily both start out knowing that the urn U was randomly chosen from a set of four urns $\{\text{urn}_0, \text{urn}_1, \text{urn}_2, \text{urn}_3\}$ where urn_i contains three balls, i of which are blue and $3 - i$ of which are green. Since the choice of U was random both subjects assign *equal* credence to the four hypotheses about its contents: $\mathbf{c}(U = \text{urn}_i) = 1/4$. Moreover, both treat these hypotheses as statements about the *objective chance* of drawing a blue ball from U , so that knowledge of $U = \text{urn}_i$ ‘screen offs’ any sampling data in the sense that $\mathbf{c}(B_{\text{next}}|E \ \& \ U = \text{urn}_i) = \mathbf{c}(B_{\text{next}}|U = \text{urn}_i)$, where B_{next} says that the next ball drawn from the urn will be blue and E is a proposition that describes any prior series of random draws with replacement from U . Finally, Jacob and Emily regard random drawing with replacement as an *exchangeable* process, so that any series of draws that produces m blue balls and n green balls is as likely as any other such series, irrespective of order. Use $B^m G^n$ to denote the generic event in which m blue balls and n green balls are drawn at random and with replacement from U . Against this backdrop of shared evidence, suppose Jacob sees five balls drawn at random and with replacement from U and observes that all are blue, so his evidence is $B^5 G^0$. Emily, who sees Jacob’s evidence, looks at fifteen additional draws of which twelve come up blue, so her evidence is $B^{17} G^3$. What should Emily and Jacob think about B_{next} ?

It would be clear what each should think if the true chance hypothesis were known with certainty: since credences should reflect known objective chances,⁶ $\mathbf{c}(B_{\text{next}}|U = \text{urn}_i) = i/3$ should hold throughout Emily and Jacob’s credal states. Unfortunately, neither Emily nor Jacob knows the true chance hypothesis, and so each has to rely on sampling data to form opinions about the various $U = \text{urn}_i$ possibilities.

Intuitively, Jacob’s total evidence points more decisively than Emily’s does toward B_{next} : all the balls he observed were blue, whereas three of the balls she saw were green. On the other hand, Emily has a greater volume of relevant evidence than Jacob does in virtue of having seen more draws. Both these facts are reflected in their credal states. After seeing five blue balls Jacob’s degrees of

belief will have shifted from an even distribution over the chance hypotheses to a distribution in which the urn_3 hypothesis is judged to be very probable, so that every function in his credal state looks like this:

$$\begin{aligned} \text{Jacob: } \mathbf{c}(U = urn_0 | B^5 G^0) &= 0 \\ \mathbf{c}(U = urn_1 | B^5 G^0) &= 0.0036 \\ \mathbf{c}(U = urn_2 | B^5 G^0) &= 0.1159 \\ \mathbf{c}(U = urn_3 | B^5 G^0) &= 0.8804 \end{aligned}$$

The probability of the next ball draw being blue is then $\mathbf{c}(B_{next} | B^5 G^0) = 0.959$. Emily, in contrast, has credences that make her almost certain that there are precisely two blue balls in the urn. All the functions in her credal state look like this:

$$\begin{aligned} \text{Emily: } \mathbf{c}(U = urn_0 | B^{17} G^3) &= 0 \\ \mathbf{c}(U = urn_1 | B^{17} G^3) &= 0.00006 \\ \mathbf{c}(U = urn_2 | B^{17} G^3) &= 0.99994 \\ \mathbf{c}(U = urn_3 | B^{17} G^3) &= 0 \end{aligned}$$

When Emily estimates the probability of B_{next} she comes up with a number indistinguishable from $2/3$ out to the fourth decimal: $\mathbf{c}(B_{next} | B^{17} G^3) = 0.666626$.

This difference in subjective probability reflects a disparity in the respective balances in the total evidence that Jacob and Emily have for B_{next} . The idea of a balance of evidence is fairly clear in cases where subjective credences are mediated by beliefs about objective chances. Chance hypotheses function as evidential funnels: data can only affect a person's beliefs about the proposition of interest by altering his or her opinions about its chance. It is then reasonable to interpret the chances $i/3$ and $(3 - i)/3$ as gauging the strength of the evidence for and against B_{next} when $U = urn_i$ is known for certain.⁷ Thus, someone with enough evidence to justify unreserved confidence in $U = urn_i$ has $i/(3 - i)$ times the evidence for B_{next} as for $\sim B_{next}$. More generally, whatever the value of $\mathbf{c}(U = urn_i)$, the conditional credence $\mathbf{c}(B_{next} | U = urn_i)$ can be interpreted as that proportion of the balance of total evidence for $U = urn_i$ that also contributes toward the balance of total evidence for B_{next} . Given this, a rational believer should use the quantity $\mathbf{c}(B_{next}) = \sum_i \mathbf{c}(U = urn_i) \times \mathbf{c}(B_{next} | U = urn_i) = \sum_i \mathbf{c}(U = urn_i) \times i/3$ as her estimate of the balance of her evidence for B_{next} . That is, she should proportion her beliefs to the evidence by having her credence reflect the expected balance of her evidence for the proposition believed.

This explains why Jacob's credence for B_{next} increases more dramatically than Emily's does. Since both initially know that U was randomly selected from $\{urn_0, urn_1, urn_2, urn_3\}$, each starts out with determinate, perfectly symmetrical evidence for and against the claim that the next ball will be blue. After seeing five blue balls Jacob's evidence requires him to regard the two chance hypotheses that favor B_{next} , $U = urn_3$ and $U = urn_2$, as much more likely than their

symmetrical counterparts, $U = \text{urn}_0$ and $U = \text{urn}_1$. In particular, he assigns $U = \text{urn}_3$, the hypothesis whose truth would conclusively justify B_{next} , a credence of 0.8804, while he assigns $U = \text{urn}_0$, whose truth would conclusively justify $\sim B_{\text{next}}$, a credence of 0. As a result, Jacob's estimated total evidence for B_{next} exceeds his estimated total evidence against B_{next} by a factor of $\mathbf{c}(B_{\text{next}}|B^5G^0)/\mathbf{c}(\sim B_{\text{next}}|B^5G^0) \approx 24$. Emily's evidence, on the other hand, forces her to regard both $U = \text{urn}_3$ and $U = \text{urn}_0$ as certainly false, and to concentrate almost all her credence on $U = \text{urn}_2$. As a result, her estimate of the total evidence for B_{next} exceeds her estimate of the total evidence against B_{next} only by a factor of $\mathbf{c}(B_{\text{next}}|B^{17}G^3)/\mathbf{c}(\sim B_{\text{next}}|B^{17}G^3) \approx 2$. There is thus a clear sense in which Jacob has better evidence than Emily does: on balance, his evidence tells more decisively in favor of B_{next} than hers does.

Emily's evidence is better along another dimension. Since she has seen a greater number of draws, her evidence, though slightly less decisive, provides her with a more settled picture of the situation. Indeed, if both subjects received evidence that tells against B_{next} , then Jacob's beliefs are likely to change more than Emily's will. Suppose that both see five more balls drawn, and all are green. Jacob's credence will fall from near 0.96 to 0.5. Emily's will move hardly at all, dropping from 0.666626 to 0.666016. This illustrates the point, made persuasively by Brian Skyrms (1980), that the weight of the evidence for a proposition X often manifests itself not in X 's unconditional credence, but in the *resilience* of this credence conditional on various potential data sequences. A person's credence for X is resilient with respect to datum E to the extent that her credence for X given E remains close to her unconditional credence for X . Note that resilience is defined relative to a specific item of data: a person's belief about X may be resilient relative to one kind of data, but unstable with respect to another. That said, it is usually the case that the greater volume of data a person has for a hypothesis the more resilient her credence tends to be across a wide range of additional data. Our example illustrates this nicely. Even though Jacob's evidence points more definitively toward a blue ball on the next draw, his credence is less resilient than Emily's with respect to almost every potential data sequence, the sole exceptions being those sequences in which *only* blue balls are drawn. In this regard Emily's evidence is better than Jacob's: even though she is not so sure as he is that a blue ball will be drawn, her level of confidence is better informed than his, and so is less susceptible to change in the face of new data.⁸

This example suggests the following provisional conclusions: (i) As Keynes argued, there is intuitive distinction between the balance of the total evidence in favor of a hypothesis and the weight of this evidence. (ii) Balances of evidence are reflected in credences in two ways: a rational believer's credence for X reflects the balance of her total evidence in favor of X ; her credence for X conditional on Y reflects that proportion of the balance of her total evidence for Y that contributes toward the balance of evidence for X . (iii) Weights of evidence are, at least in some cases, reflected in the resilience of credences in the face of additional data.

These conclusions raise as many questions as they answer. Even if it is plausible to think that probabilities express balances of total evidence in our toy example, why think this is true in general? Indeed, why think that balances of evidence can be expressed in terms of numbers at all? (Keynes himself doubted they could!) Even if they can be quantified, why think balances of evidence are probabilities? Is there any way to make the concept of “weight” formally precise, say by specifying some way of measuring it? Partial answers to these and other questions will be provided in the next two sections.

4. Measuring Balances of Evidence.

Let’s begin by focusing on what it means to say that a body of data provides *some* evidence in favor of a hypothesis. If we recall that a believer’s total evidence is a set \mathcal{E}_t of constraints on credence functions allowed into her credal state, then it is natural to say that \mathcal{E}_t provides *some* evidence for X just in case the joint satisfaction of the constraints in \mathcal{E}_t requires X to have a positive credence throughout \mathcal{C}_t . Consider, for example,

$$\begin{aligned}\mathcal{E}_1 &= \{\mathbf{c}(E) = 0.99, \mathbf{c}(X|E) > \mathbf{c}(X|\sim E)\} \\ \mathcal{E}_2 &= \{\mathbf{c}(E) = 0.99, \mathbf{c}(X|\sim E) > \mathbf{c}(X|E) = 0\} \\ \mathcal{E}_3 &= \{\mathbf{c}(X \vee Y) = 1, \mathbf{c}(Y|E) = 0, \mathbf{c}(E) = 1\}\end{aligned}$$

In all these cases the believer has some evidence in favor of X . The first constraint does not specify any particular credence for X , but it does require it to be positive. The second highlights the fact a person can have some evidence for X even though the data makes X incredible (\mathcal{E}_2 requires $0 < \mathbf{c}(X) < 0.01$). If this seems odd, keep in mind that a single body of data can provide evidence both for and against X . \mathcal{E}_3 provides *conclusive* evidence for X since, in addition to forcing her to recognize that X has some chance of being true, it also requires her to recognize that X has no chance of being false.

Our next step is to understand how the evidence for and against a proposition can be ‘balanced’. It is useful to investigate this matter by considering a more general problem. Given that \mathcal{E}_t provides some evidence in favor of each of X_1, X_2, \dots, X_N , how might this evidence be pooled to yield an overall measure of the amount of total evidence that \mathcal{E}_t provides for the combination of the propositions? More specifically, under what conditions will it make sense to say that \mathcal{E}_t provides a greater overall balance of total evidence for the propositions in $\{X_1, X_2, \dots, X_N\}$ than for those in $\{Y_1, Y_2, \dots, Y_N\}$? One tempting answer invokes a kind of dominance principle:

Combination Principle. If \mathcal{E}_t provides at least as great a balance of total evidence for X_i as for Y_i for each $i \leq N$, then \mathcal{E}_t provides at least as great a balance of total evidence for the combination of the X_i as for the combination of the Y_i .

Unfortunately, if “combination” means “conjunction” or “disjunction,” this is mistaken. For conjunction, note that on a July day we have more evidence for thinking that a fair coin will come up heads on its next toss, X_1 , than for thinking that it will rain in Los Angeles at noon, Y_1 , and we also have more evidence for thinking that the coin will not come up heads, X_2 , than for thinking that Los Angeles will be struck by a major earthquake at noon, Y_2 . Even so, it in no way follows that the balance of the evidence favors X_1 & X_2 , which is impossible, over Y_1 & Y_2 , which is merely improbable. (Readers may figure out the dual argument for disjunction.)

There is, however, another sense of “combination” in which we do have more evidence for the combination of X_1 and X_2 than for the combination Y_1 and Y_2 : we have more evidence *on average* for the X_i s than for the Y_i s. This disparity in average evidence becomes clear when we try to estimate the number of truths in the two sets. Even though we have more evidence for Y_1 & Y_2 than for X_1 & X_2 we also think it unlikely that either Y_1 or Y_2 is true, whereas we know that exactly one of X_1 or X_2 is true. So, if we had to come up with estimates for the number of truths in $\{X_1, X_2\}$ and $\{Y_1, Y_2\}$, we would settle on a value 1 for the first and a value only marginally above 0 for the second. This difference in estimates reflects the different balances of total evidence we have in favor of the X_i and the Y_i . The general principles at work here are these:

- The balances of a person’s total evidence for the propositions in $\{X_1, X_2, \dots, X_N\}$ is reflected in her estimate of the number of truths the set contains.⁹
- If the balance of total evidence in favor of X_i is increased, and if no other X_j experiences a decrease in the balance of total evidence in its favor, then the person’s estimate for the number of truths in $\{X_1, X_2, \dots, X_N\}$ should also increase.

A first stab at the combination principle we seek would require all such estimates to reflect the balances of total evidence in favor of individual hypotheses.

Combination (first approximation). If ϵ_i provides at least as great (a greater) balance of total evidence in favor of X_i as it does in favor of Y_i for each $i \leq N$, then ϵ_i constrains a person’s credal state in such a way that her estimate of the number of truths among the X_i is at least as great as (greater than) her estimate of the number of truths among the Y_i .¹⁰

This is only a first approximation because it applies only to sets of the same cardinality. We often want to compare the evidence for the propositions in one set with the evidence for propositions in another set with more or fewer elements. For example, if the geologists are right, we have significantly more evidence for thinking that California will suffer a major earthquake sometime in the next century, X , than we do for thinking either that a roll of a fair die will

produce a number strictly greater than 2, Y_1 , or that it will produce an odd number, Y_2 . Our estimate of the number of truths in $\{X\}$ is just our credence for X , let's say $9/10$, whereas our estimate of the number of truths in $\{Y_1, Y_2\}$ is $2 \times 1/3 + 1 \times (1/6 + 1/3) = 7/6$. In an aggregate sense, then, we have more evidence for the Y_i s than we have for X , but this is only because there are more Y_i s. To factor out the effect of this difference in cardinality, we can focus on the *average* amount of evidence for X and for the Y_i s by replacing sets of propositions, which, by definition, have no repeated elements, by ordered sequences of propositions, whose elements can repeat. If we compare our estimates of the average number of truths in the sequences $\langle X, X \rangle$ and $\langle Y_1, Y_2 \rangle$ we get $9/10$ and $7/12$. This reflects the fact that, on average, the balance of evidence in favor of X exceeds the balances of evidence in favor of Y_1 and Y_2 .

Generalizing on this idea leads to the correct version of the combination principle,

Combination. Let $\langle X_1, X_2, \dots, X_N \rangle$ and $\langle Y_1, Y_2, \dots, Y_N \rangle$ be ordered sequences of propositions, which may contain repeated elements. If ϵ_i provides at least as great (a greater) balance of total evidence in favor of X_i as in favor of Y_i for each $i < N$, then ϵ_i constrains a person's credal state in such a way that her estimate of the number of truths among the X_i is at least as great as (greater than) her estimate of the number of truths among the Y_i .

Combination supplies the crucial link between balances of evidence and probabilities. The connection is forged with the help of a simple consistency condition that was first explored in (Kraft, et. al, 1959) and was put in a particularly elegant form in (Scott 1964). It says, simply, that if two sequences of propositions have the same number of truths as a matter of logic, then no body of evidence can require a person's estimate of the number of truths in the first sequence to exceed her estimate of the number of truths in the second. More formally, the requirement is this

Isovalence. Suppose two ordered sequences of propositions $\langle X_1, X_2, \dots, X_N \rangle$ and $\langle Y_1, Y_2, \dots, Y_N \rangle$ are *isovalent* in the sense that, as a matter of logic, they contain the same number of truths. If the balance of total evidence in favor of X_i is at least as great as (greater than) the balance of total evidence in favor of Y_i for all $i < N$, then the balance of total evidence in favor of Y_N is at least as great as (greater than) the balance of total evidence in favor of X_N .

Isovalence has the following intuitively appealing consequences:

Normality. The balance total evidence in favor of any proposition never exceeds the balance of total evidence in favor of any logical truth.

Consequence. The balance of total evidence in favor of X never exceeds the balance of total evidence in favor of any proposition that X entails.

Partition. If the balance of balance of total evidence in favor of X & Y exceeds the balance of total evidence in favor of X & Z , then the balance of total evidence in favor of X & $\sim Z$ exceeds the balance of total evidence in favor of X & $\sim Y$.

Each of these brings out a different aspect of the concept of a balance of evidence. The first two principles make it clear that the ‘evidence’ includes not only empirical information, like the colors of balls drawn from urns, but also *logical* information, like the fact that a blue ball is also a blue or green ball. This makes perfect sense as long as we keep in mind that we are interested in balances of *total* evidence. As noted by Carnap (1962), total evidence (which he referred as ‘firmness’) satisfies both Normality and Consequence. While not every notion of evidence has this property (e.g., incremental evidence lacks it), total evidence clearly does. If a body of evidence provides any reason to think that a proposition is true, then it provides at least as much reason to think that any logically weaker proposition is true. The partition condition requires the balance of evidence for a proposition to remain the same no matter how the proposition happens to be partitioned. It says that if the total evidence, on balance, favors one way for X to be true over another, then it must also favor X being true in something other than the second way over its being true in something other than the first way.

As show in (Kraft, et. al., 1959), Isovalence ensures that balances of total evidence can be represented as probabilities. When Isovalence holds there will always be at least one (usually many) finitely additive probability functions \mathbf{c} such that $\mathbf{c}(X) \geq \mathbf{c}(Y)$ whenever \mathbf{e}_t provides at least as great a balance of total evidence in favor of X as in favor of Y . Moreover, for any sequences of hypotheses $\langle X_1, X_2, \dots, X_N \rangle$ and $\langle Y_1, Y_2, \dots, Y_N \rangle$, each such function will satisfy $\mathbf{c}(X_1 + \dots + X_N) \geq \mathbf{c}(Y_1 + \dots + Y_N)$ whenever \mathbf{e}_t provides at least as great a balance of total evidence in favor of X_i as in favor of Y_i for each $i \leq N$. Clearly, this can only happen if balances of total evidence correspond to probabilistically coherent truth-value estimates. Since a rational believer’s truth-value estimates coincide with her credences, it follows that her credences reflect the balances of her total evidence. Though we will not make the case in detail, it also follows that the believer’s credence for X given Y reflects that portion of the balance of her evidence for Y that counts in favor of X .

5. Measuring Weight.

No satisfactory measure of the weight of evidence has yet been devised. Most of the functions that have been suggested for the task—e.g., the log-likelihood ratio $\log(\mathbf{c}(E|X)/\mathbf{c}(E|\sim X))$ of (Good 1984)—are really measures of *evidential relevance* that compare balances of total evidence irrespective of weight. Since the values of these measures can remain fixed even as the volume of data increases, they do not capture the weight of evidence in the sense Keynes

had in mind. The difficulties of formulating a general measure of weight are not to be underestimated. Indeed, Keynes argued that it is impossible to capture the weight of a body of evidence using a single number. This may or may not be so, but it turns out that we can make some headway in the special case where a subject's credence for H depends on her credences for hypotheses about objective chances.

It was suggested above that the weight of evidence manifests itself in the resilience of credences in the face of new data. This is only partly right. While resilience is often a reliable symptom of weight, it is not the heart of the matter. Consider a believer whose credence for X is her estimate of its objective chance, so that $\mathbf{c}(X) = \sum_x \mathbf{c}(\text{Ch}(X) = x) \cdot x$ all across \mathbf{c}_i for some fixed partition of chance hypotheses $\{\text{Ch}(X) = x\}$, and where $\mathbf{c}(X|\text{Ch}(X) = x) = \mathbf{c}(X|E \ \& \ \text{Ch}(X) = x)$ for any potential item of data E . Here, the weight of evidence tends to stabilize X 's credence in a particular way: it stabilizes credences of chance hypotheses, while concentrating most of the credence on a small set of these hypotheses. Since acquiring the datum E will not alter X 's probability conditional on any chance hypothesis, the disparity between $\mathbf{c}(X|E)$ and $\mathbf{c}(X)$ will, other things equal, tend to be small when the disparity between $\mathbf{c}(\text{Ch}(X) = x|E)$ and $\mathbf{c}(\text{Ch}(X) = x)$ is small for most x . That said, data that alters the credence of $\text{Ch}(X) = x$ will also, other things equal, induce a smaller change in X 's credence when x is close to $\mathbf{c}(X)$ than when x is far from $\mathbf{c}(X)$. Consequently, even if E does not alter the probability of a given chance hypothesis at all, so that $\mathbf{c}(\text{Ch}(X) = x|E) = \mathbf{c}(\text{Ch}(X) = x)$, this promotes *instability* in the subject's beliefs if E affects other chance hypotheses in such a way that the distance between x and $\mathbf{c}(X|E)$ is greater than the distance between x and $\mathbf{c}(X)$. The real effect of the weight of evidence is to ensure that such increases in the disparity between chance and credence are compensated by proportional decreases in the probabilities of the offending chance hypotheses. Weight really stabilizes not the probabilities of the chance hypotheses themselves, but their probabilities discounted by the distance between X 's chance and its credence. So, the most basic resilient quantity is not $\mathbf{c}(X)$ or even $\mathbf{c}(\text{Ch}(X) = x)$; it is $\mathbf{c}(\text{Ch}(X) = x) \cdot |x - \mathbf{c}(X)|$ or, what is better for technical reasons, $w(x) = \mathbf{c}(\text{Ch}(X) = x) \cdot (x - \mathbf{c}(X))^2$.

The proposal is this: When a subject's credence for X is mediated by chance hypotheses, the weight of her evidence for X tends to make $w(x)$ resilient, so that the difference between $\mathbf{c}(\text{Ch}(X) = x|E)(x - \mathbf{c}(X|E))^2$ and $\mathbf{c}(\text{Ch}(X) = x)(x - \mathbf{c}(X))^2$ is small for most data propositions E . We can then evaluate the overall weight of the evidence for X relative to E by summing these quantities

$$w(X, E) = \sum_x |\mathbf{c}(\text{Ch}(X) = x|E) \cdot (x - \mathbf{c}(X|E))^2 - \mathbf{c}(\text{Ch}(X) = x) \cdot (x - \mathbf{c}(X))^2|$$

The weightier the evidence for X is, the smaller $w(X, E)$ will tend to be.

w is not a perfect measure of weight. Its applicability is limited since it assumes that X 's credence is mediated by credences for chance hypotheses. Moreover, its value depends on the choice of E (though weighty evidence will

tend to make $w(X, E)$ it small for a wide range of E). Even so, w has some properties that any measure of weight should have. First, it has no evidential valence, i.e., its value is the same for X as for $\sim X$. Second, when $w(X, E)$ is small, X 's credence will tend to be resilient, and when $w(X, E)$ is large X 's credence will tend not to be resilient (except for accidental reasons). Third, w relates the weight of evidence to the stable concentration of credences. It is easy to show that $w(X, E)$ is never less than the absolute difference between X 's variance conditional on E and its unconditional variance: $w(X, E) \geq |\sigma^2(X|E) - \sigma^2(X)|$.¹¹ Consequently, when the evidence for X is weighty $w(X, E)$ places a low upper bound on the amount by which the spread in probabilities for chance hypotheses can change when E is learned. This highlights what is perhaps the most important fact about the weight of evidence. Increasing the gross amount of relevant evidence for X tends to cause credences to concentrate more and more heavily on increasingly smaller subsets of chance hypotheses, and this concentration tends to become more resilient. As a result, the expected chance of X comes to depend more and more heavily on the distribution of credence over a smaller and smaller set of chance hypotheses. This is what weight does, and what w measures.

6. Specificity of Evidence.

Another subtlety in the way credences reflect evidence concerns the handling of *unspecific* data. In the terminology to be used here, data is less than fully specific with respect to X when it is either *incomplete* in the sense that it fails to discriminate X from incompatible alternatives, or when it is *ambiguous* in the sense of being subject to different readings that alter its evidential significance for X .¹² Both incompleteness and ambiguity are defined relative to a given hypothesis, and both are matters of degree. When you are told that Ed is either a professional basketball player or a professional jockey you are given very specific information about the hypothesis that he is an athlete, but somewhat less specific information about the hypothesis that he is a jockey. Likewise, if you draw a ball at random from an urn and examine it under yellow light that makes it hard to distinguish blue from green, then finding that the ball looks blue gives you specific information about how it appears in yellow light, but the data is ambiguous with respect to the hypothesis that the ball is actually blue.

The treatment of unspecific evidence has always posed a challenge for Bayesians. One approach to the problem has been to invoke some version of Laplace's infamous *Principle of Insufficient Reason*. The Principle states that "equipossible" hypotheses, those for which there is symmetrical evidence, should always be assigned equal probabilities. The idea that credences should reflect evidence might seem to require us to endorse Laplace's Principle, and the fact that unspecific evidence tends to be symmetric among possibilities would seem

to show that it is applicable. To see why not, consider the following series of examples:

Four Urns-II: Joshua knows that each of four urns – U_1, U_2, U_3, U_4 – was selected from a (different) population of urns $\{\text{urn}_0, \text{urn}_1, \text{urn}_2, \dots, \text{urn}_{10}\}$ where urn_i contain exactly i blue balls and $10 - i$ green balls. A ball will be randomly chosen from each urn U_i , and B_i is the proposition that it will be blue. Here is Joshua's evidence:

U_1 : Joshua has been allowed to look into the first urn and has seen that it contains exactly five blue balls and five green balls, so he knows that $U_1 = \text{urn}_5$.

U_2 : Joshua knows that the second urn was selected in such a way that each urn_i had an equal probability of being chosen.

U_3 ¹³: Joshua knows that the third urn was selected in such a way that urn_i was chosen with probability $\binom{10}{i}/2^{10}$ where $\binom{10}{i} = 10!/i!(10 - i)!$

U_4 : Joshua has no information whatever concerning the process by which the fourth urn was selected.

What credences should Joshua assign to the various B_i ?

There is perfect symmetry in both the balance and weight of evidence for and against each B_i . So, if rational credences must reflect total evidence, it looks as if Joshua should treat each B_i and its negation equally by assigning each credence $1/2$, just as the Principle of Insufficient Reason suggests. While this seems like a fine idea in the first three cases, it is clearly wrong in the fourth. Let's consider each case in turn.

Because Joshua knows $U_1 = \text{urn}_5$, he has specific and precise evidence that is entirely symmetrical for and against B_1 . This clearly justifies setting $\mathbf{c}(B_1) = 1/2$. Joshua's evidence about U_2 is not quite so definitive, but it is still specific enough to determine a distribution of credences over chance hypotheses: $\mathbf{c}(U_2 = \text{urn}_i) = 1/11$. Since this distribution is uniform, the balance of Joshua's evidence is again captured by $\mathbf{c}(B_2) = 1/2$. The third case is like the second, except that the distribution is binomial rather than uniform. It is still symmetrical about $1/2$, i.e., $\mathbf{c}(U_3 = \text{urn}_i) = \mathbf{c}(U_3 = \text{urn}_{10 - i})$ for each i , so $\mathbf{c}(B_3) = 1/2$ again captures the balance of Joshua's evidence.

Things get dicey in the last case. Since Joshua is in a state of complete ignorance with respect to the fourth urn, his evidence does not pick out any single distribution of credences over the chance hypotheses $U_4 = \text{urn}_i$. This is precisely the point at which some Bayesians tend to overreach by trotting out the Principle of Insufficient Reason and proposing that there is a *single* probability function—an “ignorance prior,” “uninformative prior,” or “objective Bayes” prior—that captures Joshua's evidential state. The siren song sounds like this: Since Joshua has no evidence either for or against any chance hypothesis, he no grounds for thinking that any one of them is more or less likely than any other. Given this perfect symmetry in his reasons, Joshua should not play

favorites. The only way for him to avoid playing favorites is by investing equal credence in each chance hypothesis, so that $\mathbf{c}(U_4 = \text{urn}_i) = 1/11$.

Many philosophers and statisticians object to this sort of reasoning on the grounds that it yields inconsistent results depending upon how the possibilities happen to be partitioned. This objection was first raised by John Venn in the 1800s and has been recapitulated many times since. Here is a version: Suppose you have inherited a large plot of land. You know the parcel is a square that is between one and two kilometers on a side, but this is *all* you know. How many square kilometers of land would you estimate that you own? On one hand, you might partition the possibilities by side-length in kilometers. If so, the Principle of Insufficient Reason seems to require each hypothesis $L_x =$ “Each side has length x km²,” for $1 \leq x \leq 2$, to be accorded the same credence. The expected side length of your parcel is then $3/2$ km and its expected area is $9/4$ km². On the other hand, if you partition the possibilities by area in square kilometers, the Principle tells you to distribute your credence evenly over $A_y =$ “The area of the parcel is y km²,” where $1 \leq y \leq 4$. The expected length and area are then $\sqrt{5}/2$ km and $5/2$ km². So, applying the Principle jointly to length and area produces a contradiction.

While opponents of the Principle of Insufficient Reason often regard this objection as fatal, few of its defenders are troubled by it. It is clear, they argue, that your “prior” should not be uniform over *either* length in meters or area in square kilometers. After all, these distributions are tied to specific units for measuring distance, and it is clear *a priori* that your credences should not depend on the (arbitrary) choice of a distance scale. Specifically, if your credences over the side-lengths have the functional form $\mathbf{c}(L_x) = \mathbf{f}(x)$, and if $\mathbf{t}(x) = u \cdot x$, with $u > 0$, is a transformation that alters the unit of distance ($u = 0.6214$ to switch kilometers to miles), then your credences for the rescaled side-lengths $L^*_{\mathbf{t}(x)} =$ “Each side has length $\mathbf{t}(x)$ in units u ,” should have the form $\mathbf{c}(L^*_{\mathbf{t}(x)}) = \mathbf{f}(\mathbf{t}(x))$. This guarantees, for example, that the probability of finding the side length between 1.2 and 1.3 kilometers is the same as that of finding it between 0.6214 and 0.8078 miles. It is easy to show, (Lee 1997, 101–103), that any \mathbf{f} that has this property for all u produces credences of the form $\mathbf{c}(x) = k_x/x$ where k_x is the normalizing constant $k_x = \int_1^2 dx/x = \ln(2)$. Similar reasoning shows that your credences for area hypotheses must have the form $\mathbf{c}(A_y) = k_y/y$ where $k_y = \int_1^4 dy/y = \ln(4)$. With these priors the contradiction vanishes. Computing expected area by averaging over side-length yields $\int_1^2 x^2 dx/x = 3/2 \ln(2)$, while computing it by averaging over area also yields $\int_1^4 y dy/y = 3/\ln(4) = 3/2 \ln(2)$. As shown in (Jeffries 1939), this sort of maneuver can be used in a wide variety of situations.

Adjudicating this dispute is too large a task to be undertaken here, but it is worth noting that, while the k/x prior has some nice features, it has some odd ones as well. Since it probabilifies values of x in inverse proportion to their size, the function blows up to infinity as x approaches zero. This makes k/x an “improper prior” in the sense that $\int_0^b dx/x = \infty$ for any $b > 0$. Clearly, no

such function can represent the credences of a rational believer over any interval containing zero. Likewise, since $\int_a^\infty 1/x \, dx = \infty$ for any $a > 0$ the k/x prior cannot represent the credences of a believer who is unable to impose an upper bound on the size of x . There are ways to finesse these difficulties. In some situations one can plausibly argue that uncertainty can be confined to an interval $[a, b]$ with $0 < a < b < \infty$. It is also possible to show that many improper priors generate *proper* posterior probabilities when updated using Bayes' Theorem. So, even though they cannot represent coherent beliefs, some improper priors (e.g., k/x) can be used as starting points for learning from experience. Whether or not these gambits actually succeed is a matter of controversy.¹⁴

Whatever the ultimate verdict on the k/x prior and its brethren, there is a deeper problem with “uninformative priors.” The real difficulty is not that the Principle of Insufficient Reason might be incoherent; it is that the Principle, even if it can be made coherent, is defective epistemology. It is wrong-headed to try to capture states of ambiguous or incomplete evidence using a *single* credence function. Those who advocate this approach play on the intuition that someone who lacks evidence that distinguishes among possibilities should not “play favorites,” and so should treat the possibilities equally by investing equal credence in them. The fallacious step is the last one: equal treatment does not require equal credence. When Joshua, who knows nothing about the contents of U_4 , assigns each hypothesis $U_4 = \text{urn}_i$ an equal probability he is pretending to have information he does not possess. His evidence is compatible with *any* distribution of objective probability over the hypotheses, so by distributing his credences uniformly over them Joshua ignores a vast number of possibilities that are consistent with his evidence. Specifically, if we let p range over all probability distributions on $\{U_4 = \text{urn}_i\}$, and if we consider all hypotheses of the form $O_p = “U_4 \text{ was chosen by a probabilistic process governed by distribution } p,”$ then Joshua is ignoring all those O_p in which $p(U_4 = \text{urn}_i) \neq 1/11$.

Proponents of Insufficient Reason might respond that Joshua need not ignore any O_p . With no evidence that favors any one of them over any other, they will say, Joshua should distribute his credences uniformly over the O_p . If he does this he is not ignoring any O_p : in fact, he is treating them equally. Moreover, when he computes expected truth-values relative to the uniform distribution over O_p he arrives at a credence function in which $c(U_4 = \text{urn}_i) = 1/11$. We should not be mollified by this response since it only pushes the problem up a level. Instead of ignoring potential distributions of objective probability over $\{U_4 = \text{urn}_i\}$, Joshua is now ignoring distributions of objective probability over the O_p . He is acting as if he has some reason to rule out those possibilities in which the O_p have different chances of being realized, even though none of his evidence speaks to the issue. One could, of course, move up yet another level, but the same difficulties would rearise. In the end, there is no getting around the fact that the Principle of Insufficient Reason (even if

coherent) is bad epistemology because it requires believers to ignore possibilities that are consistent with their evidence.

As sophisticated Bayesians like Isaac Levi (1980), Richard Jeffrey (1983), Mark Kaplan (1996), have long recognized, the proper response to symmetrically ambiguous or incomplete evidence is not to assign probabilities symmetrically, but to refrain from assigning precise probabilities at all. Indefiniteness in the evidence is reflected not in the values of any single credence function, but in the spread of values across the *family* of all credence functions that the evidence does not exclude. This is why modern Bayesians represent credal states using sets of credence functions. It is not just that sharp degrees of belief are psychologically unrealistic (though they are). Imprecise credences have a clear epistemological motivation: they are the proper response to unspecific evidence.

Joshua's case illustrates this nicely. Given his complete lack of evidence regarding U_4 , Joshua is being epistemically irresponsible unless, for each i and for each x between 0 and 1, his credal state c_i contains at least one credence function such that $c(U_4 = \text{urn}_i) = x$. If his opinions are any more restrictive than this, then he is pretending to have evidence that he does not have. Moreover, if Joshua's credal state is as described, then he treats each chance hypothesis $U_4 = \text{urn}_i$ exactly the same way. For every assignment $c(U_4 = \text{urn}_i) = x_i$ that appears in Joshua's credal state, and for every way s of permuting the eleven indices, the assignment $c_s(U_4 = \text{urn}_i) = x_{s(i)}$ also appears in Joshua's credal state. This is all that good epistemology demands. Symmetrical evidence only mandates equal credences when the data is entirely unequivocal and sufficiently definitive to justify the assignment of sharp numerical probabilities. When the evidence lacks specificity, propositions that are equally well supported by the evidence should receive equal treatment in the probabilistic representation. Proponents of the Principle of Insufficient Reason were right to think that good epistemology requires us to treat hypotheses for which we have symmetrical evidence in the same way; they went wrong in thinking that equal treatment requires equal investments of confidence.

7. The Contrast Between Balance and Specificity.

In general, a body of evidence E_i is specific to the extent that it requires probabilistic facts to hold across all credence functions in a credal state. If E_i is entirely specific with respect to X , then it requires $c(X)$ to have a single value all across c_i . So, perfectly specific evidence produces a determinate balance of evidence for X . Less specific evidence leaves the balance indeterminate. When the evidence for X is unspecific its credence will usually be "interval-valued," i.e., the values of $c(X)$ represented in C_i cover an interval $[x^-, x^+]$.¹⁵ It is then only determinate that the balance of evidence for X is at least x^- and at most x^+ . The difference between the 'upper probability' x^+ and the 'lower probability' x^- provides a rough gauge of the specificity of the evidence with respect to X (where

smaller = more specific). Even if the evidence for X is less than entirely specific, certain facts about balances of evidence can still be determinate. For example, even though one or both of $\mathbf{c}(X)$ and $\mathbf{c}(Y)$ might be interval-valued, the evidence can still dictate that $\mathbf{c}(X) > \mathbf{c}(Y)$ should hold across \mathbf{c}_r , in which case it is determinate that there is a greater balance of evidence for X than for Y .

One thing that might seem to put pressure on the balance/specificity distinction is the fascinating phenomenon of probabilistic *dilation* discussed in (Seidenfeld and Wasserman 1993). Here is a motivating example:

Trick Coins. The Acme Trick Coin Company makes coins in pairs: one silver, one gold. The silver coin in each pair is unremarkable, but always fair. The gold coin is quite remarkable. It contains a tiny device that can detect the result of the most recent toss of the silver coin. The device then determines a bias $\mathbf{c}(G|S) \in [0, 1]$ for a gold head in the event of a silver head, and a bias $\mathbf{c}(\sim G|\sim S) \in [0, 1]$ for a gold tail in the event of a silver tail. The device can be set at the factory so that the first bias is any real number between 0 and 1, but the second bias is always the same as the first. A gold coin set at $\mathbf{c}(G|S) = 2/3$ will come up heads two times in three after the silver coin lands heads, but it will come up heads only one time in three after the silver coin lands tails. You have a pair of Acme coins in front of you, which are about to be tossed in sequence (silver, then gold). How confident should you be that the gold coin will come up heads?

Interestingly, there is a determinate answer to this question even though your evidence is quite unspecific. Since you are completely ignorant about the gold coin's bias, your credal state contains functions whose values for $\mathbf{c}(G|S)$ and $\mathbf{c}(\sim G|\sim S)$ span the whole of $[0,1]$. Despite this, your credence for G is fully determinate. Every function in your credal state will satisfy $\mathbf{c}(G|S) = \mathbf{c}(\sim G|\sim S)$. Since the silver coin is fair, it follows that $\mathbf{c}(G) = \mathbf{c}(S)\mathbf{c}(G|S) + \mathbf{c}(\sim S)\mathbf{c}(G|\sim S) = 1/2$. Thus, your evidence for G and $\sim G$ is determinately and perfectly balanced even though your evidence about the bias of the coin is highly unspecific.

So far there is nothing to worry about: evidence that is unspecific in one respect can be specific in another. Things get hairy when we imagine that the silver coin is tossed. Suppose it comes up heads. Intuitively, it would seem that adding this very specific item of data to your evidence should increase its overall specificity, and the general effect should be a narrowing of ranges of permissible credences. Precisely the opposite happens. Once you learn S your new set of credences for G corresponds to your old set of credences conditioned on S . Since $\mathbf{c}(G|S)$ ranges over the whole of $[0,1]$, $\mathbf{c}_{\text{new}}(G)$ ranges over the whole of $[0,1]$ as well. Obviously, the same will happen if the silver coin comes up tails. This looks like a problem. If the addition of a completely precise and specific item of data can make a body of evidence so much less specific that the balance of evidence in favor of a hypothesis goes from completely determinate to entirely indeterminate, then one wonders whether any cogent distinction between specificity and balance can be maintained.

Things are not as dire as they seem. Even though you start out knowing that $\text{bias}(G|S) = \text{bias}(\sim G|\sim S)$, the evidential relevance of this information for questions about G 's truth-value is rather vexed. Clearly, the information can only be relevant to the extent that it allows you to use evidence about S and $\sim S$ to draw conclusions about G . Oddly, in this case your ability to do this turns on the amount of total evidence you have for S . To see this, suppose for a moment that you do not know the bias of the silver coin. Then $\mathbf{c}(G) = \mathbf{c}(\sim S) + [\mathbf{c}(S) - \mathbf{c}(\sim S)]\mathbf{c}(G|S)$ will hold all across your credal state. You will then be able to draw reasonably determinate conclusions about G if your credences for S and $\sim S$ are close together, but not if they are far apart. For if $\mathbf{c}(S) - \mathbf{c}(\sim S)$ is small then the imprecise term $\mathbf{c}(G|S)$ will not matter much to your views about G , but it will matter if $\mathbf{c}(S) - \mathbf{c}(\sim S)$ is large. The moral is that the relevance of the information $\text{bias}(G|S) = \text{bias}(\sim G|\sim S)$ to your opinions about G declines with the distance between $\mathbf{c}(S)$ to $\mathbf{c}(\sim S)$. It is only when $\mathbf{c}(S) = \mathbf{c}(\sim S)$ that it is completely relevant, and its degree of relevance shrinks to nothing when $\mathbf{c}(S)$ or $\mathbf{c}(\sim S)$ is 1. So, while learning S or $\sim S$ does make your evidence more specific overall, it also decreases the amount of specific evidence that is relevant to G . It does this not by making the evidence any less specific, but by making it less relevant.

8. The Contrast Between Weight and Specificity.

It is particularly difficult to disentangle weight and specificity because increases in one are often accompanied by increases in the other. This is no accident: it follows from the well-known "washing out" theorems,¹⁶ which show that subjective probabilities, as long as they are not too much at odds with one another, tend to converge toward a consensus as more and more data accumulates. Here is a simple case: Imagine that a subject's credal state contains only two functions \mathbf{c} and \mathbf{c}^* , which assign X different probabilities strictly between 0 and 1. Suppose also that there is an infinite sequence of evidence propositions $\{E_1, E_2, E_3, \dots\}$ such that:

- \mathbf{c} and \mathbf{c}^* assign each finite data sequence $D_j = \pm E_1 \ \& \ \dots \ \& \ \pm E_j$ a probability strictly between 0 and 1 (where $\pm E$ is either E or $\sim E$).
- X and $\sim X$ function like a chance hypotheses with respect to the E_j s, so that $\mathbf{c}(E_k|X) = \mathbf{c}(E_k|\pm X \ \& \ E_j) = \mathbf{c}^*(E_k|X) = \mathbf{c}^*(E_k|\pm X \ \& \ E_j)$.
- At each time j the subject acquires (perfectly specific) evidence that makes her certain of either E_j or $\sim E_j$, so that her credal state at j is $\{\mathbf{c}(X/D_j), \mathbf{c}^*(X/D_j)\}$ where D_j is the data she has received up to j .

Under these circumstances the subject's credences for X at successive times form a *martingale sequence* in which each term is the expected value of its successor. The Martingale Convergence Theorem of (Doob 1971) entails that, except for a set of data sequences to which the subject assigns probability zero, $\mathbf{c}_j(X)$ and

$c_j^*(X)$ each converge to a definite limit. Moreover, since X and $\sim X$ function like chance hypotheses these limits coincide.

Results of this sort show that, in a wide variety of circumstances, increasing the amount of specific, relevant data that a subject has for a proposition will tend to shrink the range of its admissible credence values. Of course, increasing the amount of specific, relevant data for X increases the weight of the evidence for X , and this causes the values of $c(X)$ and $c^*(X)$ to be increasingly stable. So, there is a natural convergence of opinion, and an attendant reduction of imprecision, that tends to occur as increasingly weighty evidence is acquired. This makes it difficult to separate the effects of weight from those of specificity, and many who have written on the topic have run the two together. Indeed, in an excellent recent paper Brian Weatherson (2002, 52) has argued that many of Keynes's remarks are best interpreted as supposing that the weight of the evidence for a hypothesis is reflected in the spread of its credence values. Weatherson proposes, on behalf of Keynes, that the weight of a person's evidence for X can be measured as $1 - (x^+ - x^-)$. While this is a very plausible reading of Keynes, it also shows that he conflated weight with specificity. It is a natural conflation to make, given that two quantities typically increase together as data accumulates, but it does run together quite different things. As we have seen, the overall volume of relevant evidence for X is tied not to the spread of values for X 's credence across a credal state, but to the stability and concentration of these values in the face of potential future data. The spread in credence values is a matter of the level of incompleteness or ambiguity in the data.

To illustrate this point we need an example in which the weight evidence for X increases, but its specificity with respect to X does not. It not easy to find such an example that is both simple and uncontrived, so contrived will have to do.

Guess the Weight. You are a contestant on a rather odd game show. The host holds up an opaque bag and tells you that it contains either an iron or an aluminum ingot that was chosen at random from among ingots produced yesterday at the Acme Foundry. Your job is to guess whether the ingot is iron or aluminum. You know that the iron ingots produced at Acme tend to heavier than the aluminum ones, but neither have a uniform weight. The weights of iron ingots are normally distributed about a mean of 500 oz with a variance 200 oz, whereas the aluminum ingots are normally distributed about a mean of 300 oz, again with a variance of 200 oz. You have no specific information about the proportions of iron and aluminum ingots that Acme produces, and so your credence for the proposition I that the ingot in the bag is iron covers all of $[0, 1]$. The host tells you that you may weight the bag ten times on a special scale before guessing. Unfortunately, the scale is not a terribly accurate: its results tend to be normally distributed around the true weight with a variance of 100 oz. Moreover, the scale does not report weights as such. Rather, it contains a detector that determines whether the ingot in the bag is iron or aluminum, and then it reports the difference between the ingot's true weight and the mean weight of ingots of its type. So, if the scale reads 10 oz this might mean that the ingot is iron and the scale determines it to be 510 oz, or that the ingot is

aluminum and the scale determines it to be 310 oz. You place the bag on the scale ten times and get readings of 10, -15, -8, 4, -4, 16, 12, -7, -11, 3. Call this data D , and note that the mean of the data is 0 and that its variance is 100. How does the evidence you have for I before the weighing compare with the evidence you have after you learn D ?

The contrived nature of the example ensures that there is no difference at all in the specificity of your evidence for I before and after the weighing. The information you glean from each individual weighing does nothing to distinguish I from $\sim I$ since each report is symmetrically ambiguous between the two. Moreover, the spread of reported values also provides no grounds for distinguishing I from $\sim I$ since the standard deviation in the weights of iron and aluminum ingots is identical. So, both before and after you learn D your credence for I is spread over all of $[0, 1]$.

Though you gain nothing in specificity, you do gain in weight. After you condition on D your credal state sets probabilities for ingot weights conditional on both I and $\sim I$ that are normally distributed about the same means of 500 oz and 300 oz, but with a common, smaller variance of 9.52. The smaller variance indicates that you are now more certain than you were that the ingot is close to average for its type. Moreover, these conditional means and variances are much more stable in the face of information about the ingot's weight than they were before you learned D . It is easy to see that they will be more stable in the face of more measurements on the host's unspecific scale. What might not be so obvious is that they also tend to be more stable under specific information about the ingot's actual weight (not just its weight relative to the mean in its class). Suppose, for example, that the ingot weighs in at 450 oz on a (real) scale that returns a value that is normally distributed about the true weight with a variance of 50 oz. Then the variance of both conditional distributions will shrink from 9.52 to 9.34 oz, and their means will become, respectively, about 475 oz and 375 oz. Contrast this with what happens if you learned this same fact about the ingot's weight before knowing D . In this case, the variances of both conditional distributions shrink to 40 oz., but their means become 458 oz and 425 oz, respectively. This sort of thing happens across the board. For any value the scale might read (except the midpoint 400 oz), both the changes in variances and the changes in the mean values for weights conditional on I and $\sim I$ will be smaller after D is learned than before. As we saw above, this kind of stable concentration of probability is precisely what one expects when the weight of evidence increases. Again, the weight of a body of evidence and its specificity are different things. While the first is plausibly measured by the spread in credence values over a credal state the second is not.

9. Conclusion.

Subjective probabilities reflect three aspects of a believer's total evidence—balance, weight, and specificity—in significantly different ways. The

unconditional credence of a proposition reflects the balance of total evidence in its favor. The weight of this evidence is reflected in the tendency for credences to stably concentrate on a small set of hypotheses about the proposition's objective chance. The specificity of the evidence is reflected in the spread of credence values for the proposition across the believer's credal state. Any satisfactory epistemology must recognize these three aspects of evidence, and must be attuned to the different ways in which they affect credences. It is a great strength of the Bayesian approach to epistemology that it can characterize the differences between the balance, weight and specificity of evidence in such perspicuous and fruitful ways.¹⁷

Notes

1. For a splendid discussion of Jeffries's work see (Howie 2002).
2. The terminology here is from (Levi 1980).
3. As noted in (Ramsey 1932, p. 169), it is a mistake to equate the strength of an opinion with the intensity of any feeling of conviction the believer might have since the beliefs we hold most strongly are often associated with no feelings at all.
4. It is a matter of some subtlety to say how the accuracy of truth-value estimates should be evaluated. One natural measure is the *Brier score* (Brier 1950), which was developed as a way of judging the accuracy of weather forecasts. For discussion see (Joyce 1998) and (Joyce forthcoming).
5. The proposition X is chosen from some underlying Boolean algebra of propositions, and the condition Y is taken from some distinguished set within this algebra.
6. David Lewis (1980) dubs this the 'Principal Principle'. As Lewis observes, it can come undone in circumstances where 'undermining' evidence is possible, but no such situation will be entertained here.
7. This assumes a measurement scale on which a value of 1 indicates the existence of conclusive evidence for the truth of the proposition, 0 signifies conclusive evidence against its truth, and $1/2$ means that the evidence (determinately) tells for and against the proposition in equal measure. The merits of such a measurement scheme will be discussed below.
8. Failure to keep straight the distinction between balances and weights of evidence can lead to confusion. See, for example, Popper's infamous "paradox of ideal evidence" (1959, 406), and Jeffrey's (1983, 196) decisive refutation of it.
9. Formally, a person's estimate of the number of truths in $\{X_1, X_2, \dots, X_N\}$ is her estimated value of the sum $X_1 + X_2 + \dots + X_N$, where each proposition is an indicator function that has value 1 when true and 0 when false.
10. A person's *estimate* of the number of truths in one set may exceed her estimate of the number of truths in another set even though she invests a *low* credence in the *proposition* that the first set contains more truths than the second. This will happen, for example, when $c(H_1 \& H_2) = 1/3$, $c((H_1 \& \sim H_2) \vee (\sim H_1 \& H_2)) = 0$, $c(\sim H_1 \& \sim H_2) = 2/3$, and $c(G_1 \& G_2) = 0$, $c((G_1 \& \sim G_2) \vee (\sim G_1 \& G_2)) = 1/2$, $c(\sim G_1 \& \sim G_2) = 1/2$.

11. $\sigma^2(X|E) = \sum_x \mathbf{c}(\text{Ch}(X) = x|E) \cdot (x - \mathbf{c}(X|E))^2$ and $\sigma^2(X) = \sum_x \mathbf{c}(\text{Ch}(X) = x) \cdot (x - \mathbf{c}(X))^2$.
12. Readers should be cautioned that unspecificity is different from vagueness. When the evidence for X is vague it is impossible even to assign determinate upper and lower credences to X . The treatment of vague evidence is a difficult problem that goes far beyond the scope of this paper.
13. This distribution is obtained by first randomly choosing from all the 2^{10} possible ordered sequences of blue and green balls, and then selecting the urn $_i$ that has the same number of balls as the chosen sequence.
14. Interested readers are encouraged to consult (Howson 2002, 53–56).
15. This is only “usually” because C_i need not be convex. This can happen, in particular, when E_i only determinately specifies facts about probabilistic dependence and independence.
16. For an especially lucid discussion of the convergence results see (Hawthorne 2005).
17. I am grateful to Aaron Bronfman, Branden Fitelson, Dustin Locke, Louis Loeb, Eric Lormand, Jason Stanley and Michael Woodroffe for their insightful comments on the issues discussed in this paper. This research was supported by a Rackham Interdisciplinary Research Grant from the Rackham School of Graduate Studies at the University of Michigan

References

- Carnap, Rudolf. 1962. *Logical Foundations of Probability*, 2nd. edition. Chicago: University of Chicago Press.
- de Finetti, Bruno. 1964. “Foresight: Its Logical Laws, Its Subjective Sources”, in H. Kyburg and H. Smokler, eds., *Studies in Subjective Probability*. New York: John Wiley and Sons: 93–158; all page references are to the later edition.
- Doob, J. 1971. What is a Martingale,” *American Mathematical Monthly* **78**: 451–462.
- Good, I. J. 1984 “The Best Explicatum for Weight of Evidence,” *Journal of Statistical Computation and Simulation* **19**: 294–299.
- Hawthorne, James. 2005. “Inductive Logic”, *The Stanford Encyclopedia of Philosophy (Summer 2005 Edition)*, Edward N. Zalta (ed.), URL = <<http://www.plato.stanford.edu/archives/sum2005/entries/logic-inductive/>>.
- Howie, David. 2002. *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*. Cambridge: Cambridge University Press.
- Howson, Colin. 2002. “Bayesianism in Statistics,” in R. Swinburne, e.d., *Bayes's Theorem*. Oxford: British Academy/Oxford University Press.
- Jeffries, Harold. 1939. *Theory of Probability*. Oxford: Oxford University Press.
- Joyce, James M. 1998. “A Nonpragmatic Vindication of Probabilism,” *Philosophy of Science* **65**: 575–603.
- 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
- forthcoming. “Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief;” to appear in F. Huber and C. Schmidt-Petri, eds., *Degrees of Belief*. Oxford: Oxford University Press.
- Kaplan, Mark (1996). *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.
- Keynes, John Maynard. 1921. *A Treatise on Probability*. London: Macmillan.

- Kraft, C., Pratt, J., and Seidenberg, A. 1959. "Intuitive Probability on Finite Sets," *Annals of Mathematical Statistics* **30**, pp. 408–419.
- Lee, Peter M. 1997. *Bayesian Statistics: An Introduction*, 2nd edition. London: Arnold.
- Levi, Isaac. 1980. *The Enterprise of Knowledge*. Cambridge, MA: MIT Press.
- Lewis, David. 1980. "A Subjectivist's Guide to Objective Chance," in R. Jeffrey, ed., *Studies in Logic and Inductive Probability*, Vol. 2. Berkeley: University of California Press.
- Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
- Ramsey, Frank. 1931. "Truth and Probability," in R. Braithwaite (ed.), *The Foundations of Mathematics and Other Logical Essays*. London: Kegan Paul: 156–98.
- Savage, Leonard. 1972. *The Foundations of Statistics*, 2nd edition. New York: Dover.
- Scott, Dana. 1964. "Measurement Structures and Linear Inequalities," *Journal of Mathematical Psychology* **1**: 233–47.
- Seidenfeld, T. and L. Wasserman. 1993. "Dilation for convex sets of probabilities," *Annals of Statistics* **21**: 1139–1154.
- Shimony, Abner. 1988. "An Adamite Derivation of the Calculus of Probability," in J. H. Fetzer, ed., *Probability and Causality*. Dordrecht: D. Reidel: 151–161.
- Skyrms, Brian. 1980. *Causal Necessity*. London, Yale University Press
- van Fraassen, Bas. 1983. "Calibration: A Frequency Justification for Personal Probability," in R. Cohen and L. Laudan, eds., *Physics Philosophy and Psychoanalysis*. Dordrecht: D. Reidel: 295–319.
- Weatherston, Brian. 2002. "Keynes, Uncertainty and Interests Rates," *Cambridge Journal of Economics*: 47–62.