Symmetry Is the Very Guide of Life[1]

Alan Hájek

DRAFT ONLY

"Symmetry is the way things have to be"

–Jane Siberry, *Symmetry*

1. Introduction

When Bishop Butler said that "probability is the very guide of life", he implied that there is exactly *one* guide of life, and that probability is it. He needed to say more. For as any visitor to the Taj Mahal knows, there can be all too many "guides", and not all of them take us to our proper destinations. And so it is with probability: some probabilities guide us to the right places, some guide us to the wrong places, and some guide us nowhere at all. What we need, then, is *some guide to the guides:* some principled way of choosing among the many would-be guides until we are left with the one, or ones, that will take us where we want to go.

What more could Butler have said? We have two important choices to make, corresponding to two fundamental problems in the philosophical foundations of probability. First, we must choose some formal theory of probability, some codification of how probabilities are to be represented and how they behave. Second, we must choose an interpretation of probability, an account of what probabilities are, and how they are to be determined. Regarding the first choice, I contend that we should reject the usual Kolmogorov axiomatization of probability, which takes unconditional probabilities as

[1] I thank especially Mark Colyvan, Branden Fitelson, Matthias Hild, Chris Hitchcock, Johan van Bentham, Bas van Fraassen, Peter Vranas, Jim Woodward, and participants at the 2004 Formal Epistemology

basic, and which subsequently takes conditional probabilities to be defined derivatively as certain ratios of them. Instead, we should regard conditional probabilities as primitive, and axiomatize them directly (Hájek 2003b). Regarding the second choice, we should reject any interpretation of probability that leaves us with an arbitrary, unprincipled choice among many 'guides', or with no guidance at all, and that thus leaves us essentially to guide ourselves. Instead, we should restrict ourselves to those interpretations that regard certain appropriate objective features of the world as genuine constraints on our inductive practices. I contend that these interpretations will be based on objective symmetries.

Somewhat surprisingly, the so-called 'reference class problem' unifies both these problems. It is well-known that the reference class problem is a problem for the frequentist interpretations of probability, those that identify probability with some suitably defined relative frequency of outcomes. What is not well-known is that a version of the reference class problem besets almost all of the leading interpretations of probability. But those that manage to evade it are exactly those that give us no guidance at all – 'no-theory theories' that leave us in the dark when we are making inductive inferences. Those interpretations that are genuinely illuminating, then, are as susceptible to a version of the reference class problem as the frequentist interpretations.

But the term 'reference class problem' is misleading, for sensitivity to reference class may not be such a problem after all. Rather, what it reveals is a fundamental fact about the nature of probability: that it is essentially a two-place, relative, conditional notion.

Workshop, Berkeley, for very helpful discussion.

This in turn drives home the point that conditional probabilities should be taken as basic, and that they are the proper objects of axiomatization.

This is a companion piece to my paper "Conditional Probability is the Very Guide of Life" (2003a). Despite possible appearances, the titles – and conclusions – of the two papers are not contradictory, and are in fact complementary. The present paper taxonomizes the chief interpretations of probability on the market, and distinguishes the 'no-theory theories' from those that are genuinely informative. The latter all turn out to be based on symmetries. In the other paper, I generalize the 'reference class problem': it is a problem for any interpretation that is symmetry-based. But rather than try to solve the reference class problem, I propose that we dissolve it: accept the fact that probabilities are essentially reference class-dependent, and honor that fact by taking conditional probabilities as basic.

2. More guidance to 'guides' and 'symmetry'

Butler could also have said more about what 'a guide to life' is, and before I go on, so should I. I assume that we want guidance in two respects: *epistemic* and *practical*. Firstly, we want guidance for our beliefs, or more generally, for our degrees of belief. We want our opinions to cohere with each other, to be appropriately responsive to our evidence, and to 'fit' or correspond to the way the world is. Secondly, we want guidance for our actions. We want our decisions to serve our interests as well as possible. A theory of probability, then, should offer appropriate constraints on our opinions and serve as inputs into our reasoning about what we should do.

I should also say more about what 'symmetry' is, at least insofar as it will feature in this paper. Here I am all too aware that a philosopher offers an analysis of something at his own peril, and it is all too tempting to do a Justice Potter, and simply to say that I know symmetry when I see it—and to hope that you know it when you see it too. Indeed, as long as we can agree that the 'symmetries' that I identify really deserve the name, I will have fulfilled the task that I have set myself. Still, suitably wary, let me try to do better. Said quickly, symmetry is a kind of invariance. In more detail, a symmetry is a relation between an object, or a set of objects, and a set of transformations on this object or set of objects. An object O is *symmetric with respect to* a set of tranformations $\mathcal{T}$ iff for any T $\in \mathcal{T}$, T applied to O yields O, or symbolically, T(O) = O. The entire set $\mathcal{X}$ *is symmetric with respect to* $\mathcal{T}$ iff for any object O $\in \mathcal{X}$, O is symmetric under $\mathcal{T}$. Thus, a given equilateral triangle is symmetric with respect to the set of rotations by 60, 120, or 180 degrees; and the set of all equilateral triangles is symmetric with respect to this set of rotations.

In this paper, the objects will typically be probability functions. We will see various transformations with respect to which these functions are symmetric. We will also find it useful to loosen our definitions, and to consider *near*-symmetries. Thus, we will sometimes allow T(O) to be merely *similar* to O (in specified respects), rather than necessarily identical to O.

There is a worry that for *any* set of objects there is *some* set of transformations under which the objects are symmetric. (If nothing else, there is always the identity transformation.) The symmetries that I detect strike me as being natural, and I hope that they will strike you that way too. In a sense, Goodman's 'grue' paradox involves seeing in unexpected places the symmetries or near-symmetries that underpin induction. And the skeptical lesson that one might take away from the paradox is that such symmetries are in the eye of the beholder. On the other hand, the rest of us insist that some symmetries are better than others, even if our account of the distinction between them is little better than Justice Potter's. It would take me too far afield, and it may be *too* perilous, to try to do much better here. But there is a recurring pattern: the symmetries that I will identify typically involve either *even-handed counting* over countable sets, or *even-handed measuring* over uncountable sets.[2] Theories (as opposed to no-theory theories) of probability are unified by some version of this core idea: the probability of an event is the ratio of the 'size' of the set of outcomes corresponding to the event, to the 'size' of the set of all outcomes deemed possible—'size' being measured in a symmetric way.

3. The interpretations of probability: symmetry-based theories, and no-theory theories

Probability theory begins with an axiomatization of how probabilities behave. Some writers seem to think that it also ends with an axiomatization of how probabilities behave—namely, Kolmogorov's. I will not quarrel here with that axiomatization (I quarrel with it enough elsewhere), taking it as read that it represents the orthodoxy. Instead, let us pause to note how symmetry enters even at this early stage. Consider the additivity axiom:

$$P(A \cup B) = P(A) + P(B) \quad \text{if } A \cap B = \varnothing.$$

---

[2] In a future project, I hope to explore *why* symmetry should play such an important role in probability. Here I am content merely to show that it *does*.

The probabilities of A and of B are weighted equally. Kolmogorov's theory is in this sense even-handed, democratic. It remains so when it is extended to include countable additivity, so that each member in an infinite sequence of probabilities is given the same weight in the sum. It could easily have been otherwise—indeed it is otherwise for various non-additive theories. "Baconian" probability theory, for example, equates $P(A \cup B)$ with $\max\{P(A), P(B)\}$, so that no weight whatsoever is given to the smaller of the two probabilities. The theory is 'dictatorial', insensitive to changes in the smaller probability (unless the change is large enough to switch dictators).

Be that as it may, the symmetries that interest me most come in at the next stage, when we 'interpret' probability, providing analyses of various concepts of probability. Most guidebooks to the main interpretations of probability distinguish the following interpretations of probability: frequentist, classical, logical, propensity, and subjectivist. This is fine as far as it goes, but like Butler, does not go far enough. I will thus refine this taxonomy, dividing each of these species into two sub-species:

1. Frequentist: (i) actual and (ii) hypothetical
2. Classical: (i) finite sample spaces, and (ii) infinite sample spaces
3. Logical: (i) fully constrained and (ii) less constrained.
4. Propensity: (i) symmetry-based and (ii) non-symmetry-based.
5. Subjectivist: (i) radical and (ii) constrained

We will find that most of these accounts at least tacitly, and in some cases explicitly, use symmetries to determine probabilities. Those that do not, I will argue, leave probabilities so *under*determined that they do not function as a guide to life at all. I will call them 'no-theory theories' to convey my dissatisfaction with them.[3]

3.1 Frequentist interpretations

Frequentist interpretations are unified by the view that probabilities are relative frequencies. But which relative frequencies? Here we find a point of divergence.

3.1.(i) Actual frequentism

Actual frequentists such as Venn (1876) and, apparently, various scientists even today[4], identify the probability of an attribute or event A in a reference class B with the relative frequency of actual occurrences of A within B. We thus assign probabilities to events by counting trials in a repeated experiment, taking the ratio of the number of 'favorable' trials in which a given event occurs to the total number of trials. To borrow an example of Venn's, the probability that John Smith, a consumptive Englishman aged fifty, will live to sixty-one, is the frequency of people like him who live to sixty-one, divided by the frequency of all such people. If the reference class is finite, as it typically and perhaps always is, we are done: probability is just a straightforward ratio. If not, then we must imagine the B-occurrences to be suitably ordered in some sequence, and take the probability of A to be the limiting relative frequency of A's in this sequence.

3.1.(ii) Hypothetical frequentism

[3] I borrow this term from Sober (2000), and Hild (unpublished), although I think that my usage of it differs slightly from theirs.

[4] Witness Frieden (1991): "The word 'probability' is but a mathematical abstraction for the intuitively more meaningful term 'frequency of occurrence'" (p. 10).

Hypothetical frequentists such as Reichenbach (1949) and von Mises (1957) are inspired by the dictum that probabilities are *long-run* relative frequencies and are well aware that the actual world may not deliver a long run of trials of the required sort. No matter—we can always consider instead a *hypothetical* sequence of trials that is as long as we want, and the longer, the better. In particular, an infinite sequence is as good as it gets. Thus, we have variants of this proposal:

> *The probability of an attribute A in a reference class B is what the limit of the relative frequency of A's among the B's would be hypothetically if the actual (finite) sequence were extended to an infinite sequence.*

We may want to place further constraints on which infinite sequences are acceptable. For example, there are sequences for which the limiting relative frequency of a given attribute does not exist; Reichenbach thus excludes such sequences. Von Mises gives us a more thoroughgoing restriction to what he calls *collectives,* hypothetical infinite sequences of attributes of specified experiments that meet two axioms:

> the limiting relative frequency of any attribute exists (axiom of convergence),

and

> it is the same in any infinite subsequence whose trial numbers are selected by a recursive rule, a 'place selection' (axiom of randomness)

Commentators on these frequentist interpretations tend to focus on their empiricist/positivist motivations, or their relation to scientific practice. Instead, I want to emphasize frequentism's appeal to a certain sort of *symmetry* – one so obvious that it is easily overlooked. Frequentism gives *equal weight* to each trial, so that trial number or trial outcome plays no role in the determination of probability. We could imagine, instead, various skewed 'schmequentisms' that were not so even-handed. Consider, for example, a perverse variant of actual frequentism that counts the nth trial n times. Suppose a sequence of five tosses of a coin yields:

H T H H T

Our variant treats this sequence the way that normal frequentism would treat the sequence

H TT HHH HHHH TTTTT

giving a 'probability' of heads of 8/15. Worse, another variant might look to the outcomes themselves in its biasing of the trials—for example, giving double weight to each trial in which heads shows up. (Worse, because we know *a priori* that this variant, unlike the previous one, will favor heads.) Needless to say, there seems to be something perverse about these schmequentisms, with their preferential treatment of some trials over others. Good old even-handed frequentism, by contrast, earns our respect (even if not our unanimous approval) by at least playing fair.

Following Fine (1973), we may put this more formally. Suppose we have n independent repetitions E_1, E_2, … E_n of a dichotomous experiment E, whose outcomes x_1, x_2, … x_n are represented as each taking the value 0 or 1. We seek an invariant description of the outcomes of our experiment to form the basis of our definition of probability. Given our assumption of independence of the trials, the description should be independent of the order of performance of the $\{E_i\}$, so that the description is the same for any permutation of the trial numbers. Said another way, the joint probability function of the outcomes $P(x_1, x_2, \ldots x_n)$ is a symmetric function, in that it yields the same result for all permutations of a given set of arguments. So we have all the ingredients for our recipe for symmetry. We have an object: a probability function. We have a set of transformations: all the permutations of the arguments of the function. When we apply the transformations to the object, we get the object back. The probability function is symmetric with respect to the set of permutations of its arguments. (And of course the set of all such functions is symmetric with respect to this set of transformations.) De Finetti calls this symmetry *exchangeability* (though his interest in exchangeability concerns more probability functions thought of as representing degrees of belief of suitable agents). Exchangeability is perhaps the most fundamental probabilistic symmetry, and we will see it several times again before we are done.

$$\frac{1}{n}\sum_{i=1}^{n} x_i$$

It turns out that the only functions of x_1, x_2, … x_n that display such symmetry are functions of n and the number of occurrences of each outcome. For {0, 1}-valued outcomes, the invariant probability for the '1' outcome has the form $f(\sum_{i=1}^{n} x_i, n)$. Another way of saying this is that the frequency of 1's is a sufficient statistic. Since our probability function is to be normalized, this yields $\frac{1}{n}\sum_{i=1}^{n} x_i$ as the invariant description – which is to say, the good old even-handed frequency.

And it is frequencies, not schmequencies, that figure in several of the limit theorems of probability theory. Indeed, a stronger assumption of symmetry than exchangeability, namely that of *independent and identically distributed* (i.i.d.) trials, underlies various forms of the law of large numbers. These assure us (or "almost" assure us!) that relative frequencies converge to the values that they should – the constant probabilities associated with each trial of an infinitely repeated experiment.

As well as having formal justification, frequentism has a certain sort of *salience*. One might well ask in connection with the even-handed treatment of the trials: *if not that, then what?* The alternative schmequentisms, with their various forms of affirmative action, seem arbitrary: why should we favor this sort of preferential treatment of some trials over others, rather than another? Why give double weight to the 'Heads' trials rather than to the 'Tails' trials? And so on. To be sure, salience is another notion that appears to be in the eye of the beholder. Goodman (1983) writes: "Regularities are where you find them, and you can find them anywhere". I would add: Saliences are where you find them, and you can find them anywhere. Still, just as some regularities seem to better than others (otherwise the 'grue' paradox would have no bite), so too some respects of salience seem to be better than others.

Now perhaps I've been too harsh on at least some schmequentisms. Indeed, a certain kind of schmequentism would seem to have a lot going for it. We might weight trials in the frequency count according to how *similar* they are, in some specified sense, to our target. We might, for example, weight trials according to how temporally proximate they are: the more recent they are, the more weight they get. Or recall John Smith's probability of living to sixty-one. We might impose a metric on the space of individuals, measuring their 'distance' from John Smith—that is, their degree of dissimilarity to him. Those individuals who are 'close' to him get high weight, and the 'further' they are from him, the less weight they get. This idea is vague, of course, and I don't pretend that making it precise will be easy—but it is precise enough for us to recognize that it is not frequentism, and if the weighting is done right, it may well be superior to frequentism.

Still, symmetry reappears at a deeper level. All individuals who are equally distant from Smith get *equal* weight. It is not a democracy, but a meritocracy.

Von Mises' notion of a 'collective' embodies a symmetry that goes beyond that common to all versions of frequentism: that imposed by the axiom of randomness. The axiom requires collectives to obey another invariance condition, in this case, a mereological one: certain proper parts of the entire sequence display the same behavior as the sequence itself. Randomness excludes patterned sequences – for example, H T H T H T H T …. A gambling rule (e.g., 'bet on heads on every odd toss') could be formulated to exploit such a pattern in a betting context, matching a predictable *asymmetry* in the sequence with a corresponding asymmetry in its placement of bets. Randomness, then, is another 'fairness' requirement – or, less metaphorically, another symmetry. Let's state it in terms of our formal definition. Call two infinite sequences of heads and tails *equivalent* if they have the same limiting relative frequency of heads. With any given random sequence we can associate an equivalence class generated by this equivalence relation. This equivalence class is our object. Let our transformations correspond to the place selections. Applying a place selection to any member of our equivalence class, we produce a new sequence, which has an associated equivalence class. But this equivalence class is the same one that we started with (since our original sequence obeys the axiom of randomness)—that is, we recover our original object.

But fairness/symmetry has its limits. For there is another sense, again so obvious that it is easily overlooked, in which frequentism treats events or objects *unevenly,* discriminating against some while favoring others. For all versions of frequentism, a first round of discrimination takes place in the selection of a reference class. In determining the probability that John Smith lives to 61, we do not keep a tally of all people (let alone all living things) who live that long. We restrict our attention to a special subset of them, those deemed suitably like Smith, and we give no weight whatsoever to the rest. Symmetry, then, only kicks in when we have chosen such a reference class. (The effect is the same as a 'schemquentism' that weighted individuals according to their distance from Smith according to a dissimilarity metric that assumes just the values 1 and 0.) The notorious reference class problem can be regarded as that of choosing among various competing respects of symmetry, each treating the members of a different set even-handedly. In this sense, frequentism is not democratic: all events are equal, but some are more equal than others.

Hypothetical frequentism involves a second round of discrimination: once the reference class has been fixed, there is still the further problem of ordering its members, so that a relative frequency sequence can be defined. One ordering is privileged over all others. Yet the ordering can make a difference to the limiting relative frequency. Consider an infinite sequence of the results of tossing a coin, as it might be H, T, H, H, H, T, H, T, T, … Suppose for definiteness that the corresponding relative frequency sequence for heads, which begins 1/1, 1/2, 2/3, 3/4, 4/5, 4/6, 5/7, 5/8, 5/9, …, converges to 1/2. By suitably reordering these results, we can make the sequence converge to any value in [0, 1] that we like. (If this is not obvious, consider how the relative frequency of even numbers among positive integers, which intuitively 'should' converge to 1/2, can instead be made to converge to 1/4 by reordering the integers with the even numbers in every fourth place: 1, 3, 5, 2, 7, 9, 11, 4, 13, 15, 17, 6, …) Thus, the relative frequency is only determined once we have fixed not just a reference class, but also an ordering of the members of the class. We might call this the *reference sequence* problem.

To summarize: frequentism is a symmetry-based theory of probability. But the symmetry only arises after an initial choice, or pair of choices, is made.

3.2 The classical interpretation

3.2.(i) Finite sample spaces

The classical interpretation is the symmetry-based account of probability *par excellence*, as everyone recognizes. I want to emphasize how structurally similar the account is to actual frequentism (which makes one wonder why *its* basis in symmetry is not equally recognized). Like actual frequentism, the classical theory assigns probabilities to events by counting cases; like actual frequentism, it takes the ratio of the number of 'favorable' cases in which a given event occurs to the total number of cases. This means that, like actual frequentism, it does not play favorites, treating all the cases as having equal weight. Said another way, both theories reduce the probability of an event to the ratio of the size of the set of cases in which the event occurs to the size of the set of all the cases under consideration—size measured by cardinality.

Classical probabilities, again, are exchangeable: they are invariant over permutations of the labeling of the cases, and all the joint distributions over sequences of outcomes are invariant under permutation of ordering of the outcomes. (We could express this in terms of our formal definition of symmetry, as before.) However, now the cases are not the *actual* outcomes of *repeated* performances of an experiment, but rather the *possible* outcomes of *one* performance of the experiment.

Of course, the classical theory cannot be applied just anywhere, but only when the outcomes of the experiment are, in Laplace's vexing locution, "equipossible". This appears to be a category mistake, for possibility does not come in degrees, but what is really meant is "equiprobable". Now there appears to be a different defect: circularity. However, there is no circularity if "equiprobable" is defined in terms that do not mention probability. And indeed it is: outcomes are equiprobable if there is no evidence to the contrary, no evidence that favors some outcomes over others. This is the infamous 'principle of indifference'. So symmetry enters the theory at a deeper level: classical probabilities are appropriate when the evidence, if there is any, is symmetrically balanced regarding the outcomes.

The classical interpretation, as originally conceived by authors such as Pascal and Laplace, assumes that the classes of possible outcomes are always finite, much as the reference classes that determine actual frequentist probabilities typically and perhaps always are finite. What about classical probabilities in infinite sample spaces? Here the structural analogue ought to be hypothetical frequentism, and the parallel ought to be even more striking, since in both cases we have to 'go modal', conducting our tallies over non-actual possibilities. Yet for some reason the generalization of the classical theory to the infinite case did not go the way that hypothetical frequentism did. We could, I suppose, take our cue from the latter and imagine ordering the possibilities, and then take the limiting relative frequency of favorable cases for a given event. We might even impose axioms of convergence and randomness, à la von Mises, and restrict our classical theory to sequences of equipossibilities that form ‘collectives’. Still, the question would arise, however: *which ordering among the infinitely many possible orderings is the right* *one?* No answer seems privileged. (One may wonder, again, why the parallel question for hypothetical frequentism has not been generally thought to be equally pressing—but let that pass.)

So modern heirs of the classical theory turn to other devices.

3.2.(ii) Infinite sample spaces

When the sample space is countably infinite, one may still appeal to the information-theoretic principle of *maximum entropy*, a generalization of the principle of indifference. (We will encounter new headaches in the uncountably infinite case, as we will see.) Entropy is a measure of the lack of 'informativeness' of a probability distribution. The more concentrated the distribution, the less is its entropy; the more diffuse it is, the greater is its entropy. For a discrete distribution P, the entropy of P is defined as:

$$\mathrm{H(P)} = -\sum_{k} P(X_k)\log P(X_k)$$

The principle of maximum entropy enjoins us to select from the family of all distributions consistent with our background knowledge the distribution that maximizes this quantity. In the special case of choosing the most uninformative prior over a finite set of possible outcomes, this is just the familiar 'flat' classical distribution discussed in the previous section, and symmetry is perfectly respected. Things get more complicated in the infinite case, since there cannot be a flat distribution over denumerably many outcomes, on pain of violating the usual probability calculus.[5] Rather, the best we have are sequences of

[5] Fans of non-standard analysis are quick to point out that infinitesimals permit a flat assignment – but one that violates the probability calculus twice over. After all, that calculus assumes that probabilities are real-valued infinitesimals are not), and that probabilities are countably additivity (a notion that does not even make sense in non-standard analysis).

progressively flatter distributions, none of which is truly flat – or as I prefer to say, progressively more symmetric distributions, none of which is truly symmetric.

This, in turn, raises a problem for the principle of maximum entropy: it presupposes that there *is* a distribution of maximum entropy in the family of distributions. What are we to do if there is no such distribution, but rather an infinite sequence of distributions of greater and greater entropy? The best we can do is impose some *further* constraint that narrows the field to a smaller family in which there *is* a distribution of maximum entropy. For example, we might specify that our family consists of distributions over the positive integers with a given mean, m. Then it turns out that the maximum entropy distribution exists, and is geometric:

$$P(k) = \frac{1}{1+m}\left(\frac{m}{1+m}\right)^k, \ k = 1, 2, \ldots$$

However, not just any further constraint will solve the problem. If instead, our family consists of distributions over the positive integers with finite mean, then once more there is no distribution that achieves maximum entropy. (Intuitively, the larger the mean, the more diffuse we can make the distribution, and there is no bound on the mean.)

The worry is that the further constraint required to solve the maximization problem has to be imposed from outside as background knowledge, but there is no general theory of which external constraint should be applied when. Symmetries themselves can no longer carry the day. At this point, the principle of maximum entropy falls silent, and our generalization of classical probability begins to become a no-theory theory to that extent.

Let us turn now to uncountably infinite spaces. These give rise to problems as notorious as the principle of indifference itself: Bertrand's paradoxes. They all turn on alternative parametrizations of a given problem that are non-linearly related to each other.

The following example (adapted from van Fraassen 1989) nicely illustrates how Bertrand-style paradoxes work. A factory produces cubes with side-length between 0 and 1 foot; what is the probability that a randomly chosen cube has side-length between 0 and 1/2 a foot? The tempting answer is 1/2, as we imagine a process of production that is uniformly distributed over side-length. But the question could have been given an equivalent restatement: A factory produces cubes with face-area between 0 and 1 square-feet; what is the probability that a randomly chosen cube has face-area between 0 and 1/4 square-feet? Now the tempting answer is 1/4, as we imagine a process of production that is uniformly distributed over face-area. And it could have been restated equivalently again: A factory produces cubes with volume between 0 and 1 cubic feet; what is the probability that a randomly chosen cube has volume between 0 and 1/8 cubic-feet? Now the tempting answer is 1/8, as we imagine a process of production that is uniformly distributed over volume. What, then, is *the* probability of the event in question?

It is sometimes said that Bertrand-style paradoxes involve different ways of carving up the space of possibilities, or words to that effect. That is incorrect. In the cube example, (and in general) the space of possibilities is exactly the same in each case. For example, the very same possibility is variously labeled 'the cube's side-length is1/3', 'the cube's faces have area 1/9' and 'the cube's volume is 1/27'. By contrast, a genuine case of carving up the space of possibilities in different ways is this: the possible outcomes of a die toss are {1, 2, 3, 4, 5, 6}, or alternatively {1, not-1}. The classical theory is then precariously poised to deliver conflicting values to the probability of the die landing 1, namely 1/6 and 1/2. What the Bertrand-style paradoxes *do* involve (unlike the die case) are different ways of 'equally weighting' the very same set of possibilities. They arise because the principle of indifference can be used in incompatible ways. We have no evidence that favors the side-length lying in the interval [0, 1/2] over its lying in [1/2, 1], or vice versa, so the principle requires us to give equal probability to each. Unfortunately, we also have no evidence that favors the face-area lying in any of the four intervals [0, 1/4], [1/4, 1/2], [1/2, 3/4], and [3/4, 1] over any of the others, so we much give equal probability to each. Likewise for the eight relevant intervals that the volume could land in. We cannot meet all of these constraints simultaneously. I am a fan of symmetries, but there can be too much of a good thing.

Jaynes attempts to rehabilitate the principle of indifference and to extend the principle of maximum entropy to the continuous case, with his *invariance condition:* in two problems where we have the same knowledge, we should assign the same probabilities. He regards this as a consistency requirement. For any problem, we have a group of *admissible transformations*, those that change the problem into an equivalent form. (This reminds us of our formal definition of symmetry in section 2.) Various details are left unspecified in the problem; equivalent formulations of it fill in the details in different ways. Jaynes' invariance condition bids us to assign equal probabilities to equivalent propositions, reformulations of one another that are arrived at by such admissible transformations of our problem. Any distribution that meets this condition is called an *invariant* distribution. Ideally, our problem will have a unique invariant distribution.

More precisely, our prior knowledge concerning a given problem generates a group of transformations G. From each random experiment (Ω, F, P) we generate its *orbit*, the set of all experiments (Ω, F, P_g), where P_g is given by:

$P_g(X) = P(g^{-1}(X))$ for all $g \in G$ and for all $X \in F$.

The principle of invariance forces us to regard all experiments in an orbit to be equivalent. To take an example from Fine (1973): Suppose we want to find the distribution of a point p, when all we know is that it lies on a circle of unit circumference. Ω is the set of points on the circumference of this circle; F is the Borel field of subsets of Ω; P is the distribution to be determined. Our ignorance is described by the one-parameter rotation group:

$$g(\omega) = e^{i\theta_g \omega}$$

each member of which represents rotation of Ω through an angle of θ_g radians. Applying the principle of invariance, we require all experiments in the orbit of (Ω,F, P) to be equivalent. It can be shown that the unique P for which $P_g = P$ is the uniform density over Ω. This, then, is the invariant distribution.

Rosenkrantz (1981) argues that the invariance condition resolves Bertrand-style paradoxes:

> … an invariant distribution need not be uniform. Indeed, from the present perspective, parameterization is not arbitrary; not every transformation is admissible. Consequently, examples [of the Bertrand paradox variety] put the cart before the horse. Only after we have ascertained which transformations are admissible does it make sense to ask which prior represents our state of knowledge. This procedure will never generate inconsistency not already present. That is, if we allow (mistakenly) that too rich a group of transformations leave our problem unchanged, no invariant distribution will exist for that group." (4.2 – 3).

This is a tantalizing passage. Rozenkrantz does not tell us *why* the transformations employed in Bertrand-style paradoxes (such as the squaring and cubing transformations in this one) are not admissible. The reasoning had better not be that the group of transformations is too rich if it generates such paradoxes, for then it would be the invariance condition solution that puts the cart before the horse. But let us grant for now that there is some independent criterion for determining which transformations are admissible (much as we found an independent criterion for determining when we have 'equiprobable' cases in the Principle of Indifference). We still face the opposite problem that the group of transformations may not be rich enough—a problem of underdetermination rather than overdetermination. When it is not, there will not be a unique invariant distribution. Which, then, is the correct one? Any particular choice appears to be arbitrary. Here our guide apparently deserts us; what remains is a no-theory theory.

3.3 Logical probability

Keynes (1921), Johnson (1932), Jeffreys (1939), and Carnap (1950, 1963) all regard probability theory as a generalization of logic. They think that the degree of support that one sentence, E, bestows on another, H, can be measured, and that logical probability does so. Moreover, Keynes, Johnson, Jeffreys, and the early Carnap contend that this measure is unique: there is exactly one correct measure of such support, one 'confirmation function'. (It is often called the theory of ‘inductive logic’, though this is a misnomer: there is no requirement that E be in any sense ‘inductive’ evidence for H.) The later Carnap gives up on this idea, allowing a family of such measures. We thus distinguish two versions of logical probability.

3.3.(i) Fully constrained logical probability.

Let us concentrate on the early Carnap, since his is the fullest development of a fully constrained logical probability. Once again, he takes symmetries to be essential to the determination of probabilities. This time, the objects to which probabilities are assigned are sentences in a formal language, and it will be symmetries among those that will hold the key. In (1950) he considers a class of very simple languages consisting of a finite number of logically independent monadic predicates (naming properties) applied to countably many individual constants (naming individuals) or variables, and the usual logical connectives. The strongest (consistent) statements that can be made in a given language describe all of the individuals in as much detail as the expressive power of the language allows. They are conjunctions of complete descriptions of each individual, each description itself a conjunction containing exactly one occurrence (negated or unnegated) of each predicate of the language. Call these strongest statements *state descriptions*.

Any probability measure m(–) over the state descriptions automatically extends to a measure over all sentences, since each sentence equivalent to a disjunction of state descriptions; m in turn induces a confirmation function c(–,–):

$$c(h, e) = \frac{m(h \,\&\, e)}{m(e)}$$

There are obviously infinitely many candidates for m, and hence c, even for very simple languages. Carnap argues for his favored measure "m*" by insisting that the only thing that significantly distinguishes individuals from one another is some qualitative difference, not just a difference in labeling. A *structure description* is a maximal set of state descriptions, each of which can be obtained from another by some permutation of the individual names. m* assigns each structure description equal measure, which in turn is divided equally among its constituent state descriptions. Thus, yet again we have a kind of exchangeability: invariance of probability assignments over all permutations of labels (names). Symmetry does all the hard work—again, the symmetry of counting.

Carnap shows that m* allows inductive learning. For example, where E is evidence about individuals in some set, and H makes a claim about at least some individuals outside the set, we can have c*(H, E) > m*(H). So far, so good.

3.3.(ii) Less constrained logical probability

The trouble is that infinitely many confirmation functions also have this property. Which, then, is the correct one? The early Carnap argues that m* stands out as being natural and simple, and perhaps it does. Note, though, that the notions of 'naturalness' and 'simplicity' have no place in deductive logic; this casts doubt on his view that inductive logic is cut from the same cloth.

He later generalizes his confirmation function to a continuum of functions c_λ. Define a *family* of predicates to be a set of predicates such that, for each individual, exactly one member of the set applies, and consider first-order languages containing a finite number of families. Carnap (1963) focuses on the special case of a language containing only one-place predicates. He lays down a host of axioms concerning the confirmation function c, including those induced by the probability calculus itself, various axioms of symmetry (for example, that c(h, e) remains unchanged under permutations of individuals, and of predicates of any family), and axioms that guarantee undogmatic inductive learning, and long-run convergence to relative frequencies. They imply that, for a family $\{P_n\}$, n = 1, ..., k, k > 2:

$$c_\lambda \text{ (individual } s + 1 \text{ is } P_j,\ s_j \text{ of the first } s \text{ individuals are } P_j) = \frac{s_j + \lambda / k}{s + \lambda},$$

where λ is a positive real number.

The higher the value of λ, the less impact evidence has: induction from what is observed becomes progressively more swamped by a classical-style equal assignment to each of the k possibilities regarding individual s + 1.

The problem remains: what is the correct setting of λ, or said another way, how ‘inductive’ should the confirmation function be? Nothing in logic, probability theory, or anything else for that matter seems to dictate an answer. In particular, while symmetries determined much else in Carnap's system, they determine nothing about the value of λ. Arbitrariness threatens—again, in a way that has no analogue in deductive logic. 'Inductive logic' must face becoming a no-theory theory.[6]

Now the problem of arbitrariness of the value of λ would not be so bad if the degrees of confirmation turned out much the same whatever value we picked. If we could confine logical probabilities to relatively small intervals, then they could still be useful. Unfortunately, this is far from the truth. At one extreme, $\lambda=\infty$, our inductive rule is so cautious that the probability above never budges from its a priori value of 1/k. At the other extreme, $\lambda=0$, the rule is so incautious that it becomes the 'straight rule', assigning a probability equal to the observed relative frequency. If the number of predicates is large, the former probability will be small, while the latter probability may be as high as 1. Intermediate values of λ lead to all of the values in between. Which guide, then, should we follow?

.

3.4 Propensity theories

[6] These problems are only exacerbated in the still-later Carnap’s 2-dimensional continuum, in which *two* arbitrary choices of parameters (λ and γ) must be made before logical probabilities can be determined.

Propensity theorists think of probability as a physical propensity, or disposition, or tendency of a given type of physical situation to yield an outcome of a certain kind, or to yield a long run relative frequency of such an outcome. This immediately prompts a distinction between *frequency-based* and *non-frequency-based* propensity theories. Frequency-based theories associate propensities with repeatable experimental arrangements, viewing propensities as tendencies to produce relative frequencies equal to the probabilities. Popper (1959), for example, regards a probability p of an outcome of a certain type to be a propensity of a repeatable experiment to produce outcomes of that type with limiting relative frequency p. Giere (1973), on the other hand, explicitly allows single-case propensities, with no mention of frequencies: probability is just a propensity of a repeatable chance set-up to produce sequences of outcomes. His is thus a non-frequency-based propensity theory.

3.4.(i) Symmetry-based propensities

Frequency-based propensity theories tacitly appeal to a certain sort of symmetry, namely that which is found in frequentism itself, and as such I will call them 'symmetry-based'. We can imagine various skewed, 'schmequency'-based propensity theories. Each one associates propensities with repeatable conditions and identifies them with tendencies to produce 'schmequencies' of outcomes – where 'schmequencies' give weight n to the nth trial, or give double weight to the 'heads' outcomes of a coin-tossing experiment, or what have you. Such theories would in general be absurd, of course.[7] Whatever misgivings we might have about frequency-based propensity theories, we don't blink at the point at

---

7 Except, perhaps, those that piggy-backed on a schmequentism that weighted trials according to their similarity to some target.

which they make a connection to good old even-handed frequencies; whatever the faults of such theories, their tacit appeal to this *symmetric* notion (as opposed to some asymmetric variant of it) was not one of them.

But what about non-frequency-based propensity theories? They cannot enjoy any benefit that frequentism's symmetry might confer. Perhaps they could appeal to other symmetries, and if so, I will still call them 'symmetry-based'. A coin, after all, might be pronounced to be fair in virtue of its *physical* symmetries – its symmetric shape, its symmetric mass distribution, or what have you – with no regard for what the results of tossing it happen to be. Such symmetries must have their limits: the coin cannot be *perfectly* symmetrical with respect to its faces, or else there would be no distinguishing them! Indeed, *too* much symmetry would be an embarrassment for Leibniz's identity of indiscernibles. Still, the symmetry-breaking needed is minimal, and what remains is still symmetry for all practical purposes. More complicated systems, such as dice and roulette wheels, function as well as they do as gambling devices in virtue of further symmetries (or near-symmetries), albeit more complicated.

Indeed, Strevens (2003), building on work by Poincaré, Suppes, Keller, Engel, Diaconis, and others, shows how much high-level simplicity, in the form of stable probabilities, inevitably arises out of low-level complexity, *in virtue of symmetries* of a certain kind. (Think, for example, of how microscopic chaos gives rise to macroscopic order in thermodynamics, or how capricious behaviors of the individual organisms in an ecological system give rise to stable probabilities at the population level.) For Strevens, the key to these seemingly miraculous phenomena is found in the notion of "microconstancy", a certain symmetry property, or near-symmetry property, of the systems concerned. Very roughly, the idea is this. Associated with such a system is an *evolution function*, a function that given as input a micro-state as an initial condition, determines whether a designated macro-state occurs or not. An evolution function is *microconstant* if – here comes the near-symmetry – its behavior is approximately the same over many small contiguous regions of its domain, oscillating in an approximately uniform way between the designated macro-state and other macro-states. In Strevens' terminology: the *strike ratio* of a region of initial condition values is the proportion of initial condition values in the region that produce a designated outcome. An evolution function is *microconstant* just in case its domain can be partitioned into many small, contiguous regions each having approximately the same strike ratio. The probability of the designated outcome is then identified with the strike ratio. When he speaks of 'proportion', I take him to mean 'measure', since often the relevant sets of initial conditions will be uncountable. But which measure? I take him to mean the *uniform* measure over the sets. Thus, probability of an event is reduced to the ratio of the size of the set corresponding to that event, to the size of the set of all possible outcomes, 'size' being measured in a symmetric way.

Or we might leave the actual world, grounding propensities instead in *modal* symmetries. McCall (19xx), for example, envisages a model of the universe as a branched structure, the moving present cutting off all but one branch like an inexorable chain-saw. Propensities are proportions among future branches: this coin has a propensity of 1/2 of landing heads because in half of the futures that lie ahead of us, it *does* land heads, and in half of them it does not. Symmetries rule once again: probability is again reduced to even-handed counting.

3.4.(ii) Non-symmetry-based propensities

What of a propensity theorist who has no truck with symmetries whatsoever? What, for example, of such theorists who simply say that propensities are intrinsic properties of chance set-ups, or inherent dispositions, or unanalyzable tendencies, or graded modalities, or …—and say no more? In particular, they make no mention of symmetries, explicitly or implicitly. I submit that they are offering no-theory theories of probability. They give us no clue as to how such probabilities are to be determined by us, or what determines them. In fact, it is unclear why they should even obey the probability calculus. Do we truly understand probability any better by being told that it is such a propensity? I submit that we have here an 'aleative virtue' theory of probability.[8]

3.5 Subjectivism

Subjectivists regard probabilities as degrees of belief and see the theorems of probability as rationality constraints on degrees of belief. Most add the further constraint that degrees of belief should be updated by successive conditionalization on the evidence as it comes in.

3.5.(i) Radical subjectivism

Radical subjectivists (or "orthodox Bayesians") such as de Finetti regard them as the *only* such constraints. This results in a spectacularly permissive epistemology—thus, precious little guidance. For example, you may without any insult to rationality assign

[8] I owe this quip to Peter Godfrey-Smith.

probability 0.999 to the period at the end of this sentence being the Messiah, provided that you assign 0.001 to this not being the case (and that your other assignments also obey the probability calculus). Similarly, you may see dependences among events in ways that are properly regarded as irrational—for example, we call it the 'gambler's *fallacy*'—without setting off the radical subjectivist alarm. This all stems from the fact that coherent priors can be found that will give rise to such states of opinion, even after repeated conditionalization on your evidence (which, after all, is finite). How are we to rule out such priors? Nothing in the theory guides you; you are left with an arbitrary, unprincipled choice. Radical subjectivism is a no-theory theory.

Subjectivists typically believe that there are analytic connections between an agent's degrees of belief and her behavioral dispositions. For example, de Finetti analyzes subjective probabilities in terms of betting dispositions, and followers of Ramsey and Savage derive such probabilities, along with utilities, from preferences, yielding a representation of rational action as that which maximizes an agent's expected utility. Thus, once you have a probability function (without any outside help), it almost trivially functions as a 'guide to life'.

But is the guide any good? Someone who kneels before the period above and worships it, or who pays an extraordinarily high price for a bet that pays off if it is the Messiah, may well be 'guided' by the aforementioned subjective probability. The high roller at Las Vegas who after a run of 'reds' on the roulette wheel bets heavily on black "because it is *due*" is likewise guided by his subjective probability. But they are, in both senses of the word, *misguided*. The trouble with the 0.999 assignment, or the fallacious gambler's assignment, of course, is that they fly in the face of the facts about our world as we know them to be. Periods at the ends of sentences do not save the world; in fact, they don't do much of anything. Roulette wheels don't have a memory, still less a conscience.

I said earlier that we want our opinions to cohere with each other, to be appropriately responsive to our evidence, and to 'fit' or correspond to the way the world is. Radical subjectivists have had much to say about the first two desiderata, but surprisingly little about the third, which offhand seems to be the most important of them all. That's how our misguided subjects get off the hook so easily. Being saddled with such unwelcome results is the price that the radical subjectivist pays for offering a no-theory theory. Probability functions as a 'guide' in name only, much like the 'guide' to the Taj Mahal who takes you instead to his carpet store. It would trivialize the word to call something a guide just because it leads you *somewhere*. The word is, after all, normatively loaded: when Butler said that probability is the very guide of life, he was praising it. The radical subjectivist's probabilities may not deserve such praise.

3.5.(ii) Constrained subjectivism

Thus, many subjectivists are more demanding of their subjects. Some might impose the constraint that you should regard certain sequences of events as *exchangeable*: your probabilities over them should be invariant over permutation of their order.[9] Paralleling our discussion of frequentism, this is a symmetry condition, one that allows one to prove a limit theorem to the effect that, in the long run, the agent's degrees of belief (which update by conditionalizing on the evidence) will almost surely converge to corresponding

[9] Note that despite his enthusiasm for exchangeability, de Finetti does not think that we *should* regard certain sequences of events as exchangeable; rather, he only thinks that as a matter of fact, we *do*.

relative frequencies. Skyrms (1994) canvases variants of exchangeability – notably *partial,* and *Markov* – that are further symmetry principles in their own right, and that lead to counterpart limit theorems.

There are various proposals for extra constraints on rational opinion. I find it most perspicuous to present them all as instances of a certain canonical form, in which a 'guide' is made explicit. Gaifman (1988) coins the phrases "expert assignment" and "expert probability" for a probability assignment that a given agent strives to track: "The mere knowledge of the [expert] assignment will make the agent adopt it as his subjective probability" (193). The guiding idea is captured by the equation

(*) $P(A|pr(A) = x) = x$

where 'P' is the agent's subjective probability function, and '*pr*(A)' is the assignment that the agent regards as expert. For example, if you regard the local weather forecaster as an expert on matters meteorological, and he assigns probability 0.1 to it raining tomorrow, then you may well follow suit:

$$P(\text{rain}|pr(\text{rain}) = 0.1) = 0.1$$

Perhaps 'probability guide' would be an even better name than 'expert probability', since things can function as guides without themselves being 'expert' about anything. (Think of the guidance that the North Star provides sailors.)

More generally, we might speak of an entire probability function as being such a guide for an agent. Van Fraassen (1989), extending Gaifman's usage, calls *pr* an "expert function" for P if (*) holds *for all x* such that $P(pr(A) = x) > 0$, so that the conditional probability is defined. We should keep in mind the distinction between an expert function and an expert assignment, because an agent may not want to track *all* the assignments of her 'expert'. (If your forecaster gives probability 0 to it raining in Los Angeles tomorrow, you may think that he's gone too far, and may not want to follow him there.)

Various candidates for expert functions for rational agents have been proposed:

The *Principle of Direct Probability* regards *relative frequencies* that way. Let A be an event-type, and let relfreq(A) be the relative frequency of A. Then for any rational agent with probability function P, we have

$$P(A|\text{relfreq}(A) = x) = x, \text{ for all A such that } P(\text{relfreq}(A) = x) > 0.$$

Lewis (1980) posits a similar role for the *objective chance function, ch,* in his Principal Principle:

$$P(A|ch(A) = x) = x, \text{ for all A such that } P(ch(A) = x) > 0.$$[10]

A frequentist who thinks that chances just *are* relative frequencies would presumably think that the Principal Principle just *is* the Principle of Direct Probability; but Lewis' principle may well appeal to those who have a very different view about chances—e.g., propensity theorists.

Van Fraassen (1984), following Goldstein (1980), argues that one's *future probability assignments* play such a role in constraining one's present assignments in his Reflection Principle:

$$P_t(A|P_{t+\Delta}(A) = x) = x.$$

The idea is that a certain sort of epistemic integrity requires you to regard your future self as 'expert' relative to your current self.

One might also give conditionalized versions of these already-conditional principles, capturing the idea that an agent might want to track certain conditional probability

[10] I ignore complications due to Lewis' notion of "admissibility".

assignments of her expert. For example, Lewis (1994) amends the Principal Principle in such a way:

$$P(A|ch(A|B) = x \cap B) = x$$

(We could amend the Principle of Direct Probability and the Reflection Principle similarly.) Finally, if Carnap is to be believed, then *logical probability* plays such a role as expert – perhaps the ultimate one.

We have thus moved beyond no-theory theories of subjective probability. I want to stress the subtle role of symmetries in these principles. Firstly, in a sense that I do not want to press too hard, they *are* symmetry principles in a sense: a second-order conditional probability has the same value as a first-order probability in its scope. The value is preserved when we go 'up' a level. Note that the van Fraassen's coinage, 'Reflection', is deliberately suggestive of symmetry, and he tells me that this is no accident: your current probability function 'mirrors' that of your future self as far as you are concerned. But I do not want to press this sense of ‘symmetry’ too hard because it does not seem to fit my formal definition of section 2. So I have not really earned the right to call it a symmetry, even though I feel that I see it and know it to be such. (I did observe that giving such definitions was perilous.)

Secondly, and more importantly, I submit that to the extent that the principles are reasonable and usable, it is in virtue of their reliance on symmetries, tacit or otherwise. The Principle of Direct Probability would lose all its plausibility if *relfreq* were replaced by *schmelfreq*, some skewed variant of the relative frequency function that played favorites among the outcomes of a sequence of trials.[11] The notion of 'chance' in the

[11] Unless, perhaps, the variant weighted trials according to their similarity to some target.

Principal Principle is, according to some theorists, simply relative frequency, the case just considered. According to some other theorists, it is a propensity. We have seen that some propensity theories inherit symmetries: sometimes they are the symmetries inherent in frequencies, sometimes they are physical symmetries in objects such as coins, sometimes they are symmetries in evolution functions underlying macroscopic processes that are determined by microscopic behavior, and sometimes they are modal symmetries. Still other theorists do not tell us much about chance, thought of as a propensity, beyond its being a tendency, a disposition, …. They are no-theory theorists, and they leave us wondering why the Principal Principle should have any claim on us, and leave it utterly mysterious how to apply the Principle. 'Chance' is resonant name for a ‘guide’ that take us nowhere at all.

The early Carnap's logical probabilities are determined by symmetries. Those of the later Carnap depend partly on a seemingly arbitrary choice of λ. But to the extent that the choice is arbitrary, logical probabilities apparently lose their authority. For since the choice is arbitrary, each choice is as good as any other. But then all the resulting confirmation functions are equally good guides—and as we noted earlier, they can take us to very different destinations.

As for human 'experts' (weather forecasters, your future selves, and so on) and their subjective probabilities, we have a dilemma: either they are constrained by something external to *them* or they are not. In the former case, presumably something else is playing the role of 'expert' for *them*. If they it is frequency information, we will find our symmetries there; if it is information about physical symmetries, or symmetries in evolution functions, or modal symmetries, we will find them there. If the information involves the credences of other human experts, then we repeat the process: either those other experts are constrained by something external to *them* or they are not… Either this recursive process eventually bottoms out with some 'ur' symmetry source, or it does not. If it does, my point is made. If it does not, it is dubious whether they have earned their title as 'experts'; we would be left with a no-theory theory of expertise.

4. Conclusion

The interpretations that are based on symmetries are:

- Actual and hypothetical frequentism. Symmetries: exchangeability, and in the case of hypothetical frequentism, invariance of limiting relative frequencies under place selections.
- Classical probabilities in finite sample spaces. Symmetries: invariance under relabeling of outcomes, and exchangeability.
- Fully constrained logical probability. Symmetries: invariance under permutation of names and predicates in state descriptions.
- Symmetry-based propensities. Symmetries: frequency-based, physical (in the objects themselves), microconstancy (in evolution functions), or modal.
- Constrained subjectivism. Symmetries: exchangeability, and those inherited from various ‘expert assignments’.

These theories reduce probability in one way or another to symmetric measuring of the sizes of sets—taking the ratio of the size of the set of ‘favorable’ outcomes to the size of the set of ‘all possible outcomes’. Determining the set of ‘all possible outcomes’ gives rise to the reference class problem—a topic I take up in the companion piece to this paper, "Conditional Probability is the Very Guide of Life".

Those interpretations that are not based on symmetries, the no-theory theories are:

- Classical probabilities in infinite sample spaces.
- Less constrained logical probability.
- Non-symmetry-based propensities.
- Unconstrained subjectivism.

They do not face a reference class problem. But this is not to commend them, for the way they avoid it is by failing to be guides to life.

Moreover, the reference class problem is not such a problem after all, symptomatic as it is of the fundamental relativity of probability, the fact that probability is essentially conditional. That, however, is a story told in another place, the companion piece. Combining the conclusions of that paper and this one: Bishop Butler was on the right track, but I would go further: *symmetry-based conditional probability is the very guide of life*.

REFERENCES

Bertrand, J. (1889): *Calcul des Probabilités*, 1st edition, Gauthier-Villars; 2nd edition 1907; reprinted 1972 by Chelsea Publishing Co., New York.

Carnap, Rudolf (1950): *Logical Foundations of Probability*, University of Chicago Press.

Carnap, Rudolf (1963): "Replies and Systematic Expositions", in P. A. Schilpp (ed.), The Philosophy of Rudolf Carnap, Open Court, La Salle, Ill, 966-998.

De Finetti, Bruno (1937): “Foresight: Its Logical Laws, Its Subjective Sources”, translated in Kyburg and Smokler (1964).

Fine, Terrence (1973): *Theories of Probability*, Academic Press.

Gaifman, Haim (1988): "A Theory of Higher Order Probabilities", *in Causation, Chance, and Credence*, eds. Brian Skyrms and William L. Harper, Kluwer.

Giere, R. N. (1973): "Objective Single-Case Probabilities and the Foundations of Statistics", in P. Suppes et al. (eds.), *Logic, Methodology and Philosophy of Science* IV, North Holland 467-83.

Gillies, Donald (2000): "Varieties of Propensity", *British Journal of Philosophy of Science* 51, 807-835.

Goldstein, Michael (1983): “The Prevision of a Prevision” *Journal of the American Statistical Association* 77, 822–30.

Goodman, Nelson (1983): *Fact, Fiction, and Forecast (4th ed.)*, Harvard University Press.

Hájek, Alan (2003a): "Conditional Probability is the Very Guide of Life", in *Probability is the Very Guide of Life: The Philosophical Uses of Chance*, eds. Henry Kyburg, Jr. and Mariam Thalos, Open Court, 183-203. Abridged version in *Proceedings of the International Society for Bayesian Analysis 2002*.

Hájek, Alan (2003b): "What Conditional Probability Could Not Be", *Synthese*, Vol. 137, No. 3, December 2003, 273-323.

Hild, Matthias (unpublished): Introduction to *The Concept of Probability: A Reader*.

Jaynes, E. T. (1968): “Prior Probabilities”, *Institute of Electrical and Electronic Engineers Transactions on Systems Science and Cybernetics*, SSC-4, 227-241.

Jeffreys, Harold (1939): *Theory of Probability*; reprinted in Oxford Classics in the Physical Sciences series, Oxford University Press, 1998.

Johnson, W. E. (1932): “Probability: The Deductive and Inductive Problems”, *Mind* 49, 409-423.

Keynes, J. M. (1921): *Treatise on Probability*, Macmillan, London. Reprinted 1962, Harper and Row, New York.

Laplace, Pierre Simon de (1814): “Essai Philosophique sur les Probabilités”, Paris. Translated into English as *A Philosophical Essay on Probabilities*, New York, 1952.

Lewis, David (1980): "A Subjectivist's Guide to Objective Chance", in R.C. Jeffrey ed. *Studies in Inductive Logic and Probability*, Vol II., University of California Press, 263-293; reprinted in *Philosophical Papers Volume II,* Oxford University Press.

Lewis, David (1994): “Humean Supervenience Debugged”, *Mind* 103, 473-490.

McCall, Storrs (19xx): *A Model of the Universe,* Oxford.

Popper, Karl (1959): "The Propensity Interpretation of Probability", *British Journal of Philosophy of Science*10, 25-42.

Ramsey, F. P. (1926): “Truth and Probability”, in *Foundations of Mathematics and other Essays;* reprinted in Kyburg and Smokler (1963), and in D.H. Mellor (ed.) *Philosophical Papers*, Cambridge University Press, Cambridge, 1990.

Reichenbach, Hans (1949): *The Theory of Probability,* University of California Press.

Rosenkrantz, R.D. (1981): *Foundations and Applications of Inductive Probability*, Ridgeview Publishing Company.

Savage, L. J. (1954): *The Foundations of Statistics*, New York, John Wiley.

Skyrms, Brian (1994): “Bayesian Projectibility”, in *Grue!* (D. Stalker ed.), Open Court, 241-62.

Sober, Elliott (2000): *Philosophy of Biology*, Westview Press, 2nd ed.

Strevens, Michael (2003): *Bigger than Chaos*, Harvard University Press.

van Fraassen, Bas (1984): "Belief and the Will", *Journal of Philosophy* 81, 235-256.

van Fraassen, Bas (1989): *Laws and Symmetry*, Clarendon Press, Oxford.

van Fraassen, Bas (1995): "Belief and the Problem of Ulysses and the Sirens", *Philosophical Studies* 77, 7-37.

Venn, John (1876): *The Logic of Chance*, 2nd ed., Macmillan and co.

von Mises, Richard (1957): *Probability, Statistics and Truth*, revised English edition, New York.