BAYESIAN ORGULITY

GORDON BELOT

ABSTRACT. A piece of folklore enjoys some currency among philosophical Bayesians, according to which Bayesian agents that, intuitively speaking, spread their credence over the entire space of available hypotheses are certain to converge to the truth. The goals of the present discussion are to show that kernel of truth in this folklore is in some ways fairly small and to argue that Bayesian convergence-to-the-truth results are a liability for Bayesianism as an account of rationality, since they render a certain sort of arrogance rationally mandatory.

1. INTRODUCTION

Convergence to the truth is sometimes easy for Bayesians, sometimes difficult, sometimes rare, and sometimes virtually impossible. Speaking somewhat loosely (more careful formulations are given below), we have the following examples.

(a) Suppose that the problem is to determine the bias of a coin from knowledge of the outcomes of an infinite sequence of tosses. Typical Bayesian agents are essentially guaranteed to converge to the truth, no matter which hypothesis is true.

(b) Suppose that the problem is to determine the propensities underlying a chance process that has a countable infinity of possible outcomes. Then, for typical hypotheses about the propensities, if any of those hypotheses are true, then typical Bayesian agents are (essentially) guaranteed not to converge to the truth.

(c) Suppose that the problem is to determine whether a given binary sequence, revealed bit by bit, encodes a binary expansion of a rational number. For any Bayesian agent, there is a rich infinite family of sequences the agent could be shown that would prevent

---

Forthcoming in *Philosophy of Science.* Version of June 14, 2013.

For helpful conversations and suggestions, thanks to Bob Geroch, Alistair Isaac, Jim Joyce, Sarah Moss, Laura Ruetsche, and two anonymous referees. This paper owes a great great deal to the approaches of Kelly (1996, Chapter 13) and Earman (1992, Chapter 9). For all those readers out there who are fans of both Stefan Banach and Sir Thomas Malory—fn. 15 is for you.

convergence to the truth. Bayesian agents that are open-minded in a certain appealing sense fail to converge to the truth for typical sequences.

(d) Suppose that the problem is to determine, from knowledge of the outcomes of an infinite sequence of tosses, whether a given coin has a bias of two-thirds in favour of heads or in favour of tails—and also whether it is a nickel, a dime, or a quarter. Because the evidence does not distinguish between each pair of fine-grained hypotheses under consideration, convergence to the truth is possible only in special cases.

Bayesian agents tend towards supreme self-confidence: facing a problem of type (a), (b), or (c), they will assign probability zero the possibility of failing to converge to the truth.

While from the outside it is clear that problems of type (b), (c), and (d) are far more intractable for Bayesians than are problems of type (a), this fact is not reflected in the judgements of Bayesian agents: they rate their chances of success for problems of type (b) or (c) just as highly as for problems of type (a), even while being willing to admit that success is not inevitable for problems of type (d).

In Section 5 below, I will argue that there is a real problem here for Bayesians—their account of rationality renders a certain sort of arrogance rationally mandatory, requiring agents to be certain that they will be successful at certain tasks, even in cases where the task is so-contrived as to make failure the typical outcome.[1]

The first order of business, in Section 2, is to get on the table the notion of typicality that will be in play in the following sections. In Section 3, by way of stage-setting and because I suspect that the results in question are not as widely known among philosophical Bayesians as they might be, I survey some classic results concerning problems of type (a) and (b). In Section 4, I establish the picture sketched above for problems of type (c).

2. Typical!

One often says that a property is typical of a given class of objects if the set of objects exemplifying that property is overwhelmingly large in some appropriate sense—or if the set of objects lacking the property is so small as to be negligible in some appropriate sense.

[1]This can be thought of as a sharpened relative of an objection advanced, somewhat ambivalently, by Dawid—cf. his (1982, §6) and (1985, Note 7.3).

Example: The Unit Interval. Let us begin with a concrete example: the unit interval $I = [0, 1]$ of the real line. We can think of this in a number of ways—notably, as a set, as a topological space (a set together with a notion of convergence of sequences), or as a metric space (a set together with a notion of distance—and hence also a notion of convergence).[2] Each of these brings with it a corresponding notion of smallness and largeness.

If we consider I as a set, then it is natural to think of its countable subsets as being small, and to think of the complements of such subsets as being large.

If we consider I as a equipped with its usual notion of distance, then it is natural to consider a subset X to be small if and only if it is a (Lebesgue) *nullset*—if and only if, that is, for every $\varepsilon > 0$ we can a find a collection of subintervals of I such each point in X is in one of these subintervals while the total length of these subintervals is less than ε. Correlatively, one thinks of complements of nullsets as being large.

In the intermediate case, in which we consider I as a topological space, the relevant notion of smallness is perhaps less familiar. We call a subset $X \subset I$ *meagre in* I if $X = \cup_{k=1}^{\infty} X_k$ where each X_k is a nowhere dense subset of I (in the present case this means: even if we add to X_k those points that arise as limits of sequences of points in X_k, the resulting set still does not contain any subintervals of I as subsets).

To motivate this notion, note that the families of countable subsets of I and the nullsets of I are closed under the operation of taking countable unions. So one way to arrive at a topological notion of small subset of I is to isolate a family of undeniably small sets, then look at the family of sets that arise as countable unions of these. Now, it is natural to consider any open dense $Y \subset I$ large: no subinterval is so small that it avoids such a Y; every point in such a Y lies in some subinterval wholly contained in that Y. The complement of such a Y is a nowhere dense set—so it is natural to regard such sets as small. The meagre sets are those that result from beginning with the nowhere dense sets and closing under the operation taking countable unions. No subinterval of I is meagre—and the complement of a meagre set is uncountable and dense in I.[3]

The relations between these three notions of smallness are complex. Every countable set is both a nullset and a meagre set. But there exist uncountable subsets of I that are both meagre and null. The

[2]All spaces considered here and below are metrizable—so their topologies can be characterized in terms of the convergence of sequences (Dudley, 1964, §2).

[3]See, e.g., Oxtoby (1980, Theorem 1.3).

Cantor set, which will play a role in Section 4 below, is the standard example. Let C_1 be the result of deleting the (open) subinterval that is the middle third of $C_0 = I$. Let C_2 be the result of deleting (open) middle thirds of each of the components of C_1. And so on. Then the Cantor set is $C := \cap C_k$. Equivalently, we can think of C as the set of those numbers in I that can be written in base-three without using the numeral one. C is uncountable.[4] But it is a nullset.[5] And C is itself nowhere dense—and hence is meagre as a subset of I.[6]

Since there are meagre nullsets, there are also sets whose complements are meagre and null. Further, there are nullsets with meagre complements—and hence also meagre sets with complements that are nullsets.[7] In practice, the two notions live side by side, providing complementary, sometimes clashing, notions of smallness.[8]

One of the strengths of the notion of a meagre set is that it is purely topological—and so is invariant under homeomorphisms from the interval to itself (i.e., continuous deformations of the scale on the interval that leave the endpoints fixed). The notion of a nullset, though, depends on the metric structure of the interval (or on some other structure that goes beyond its topology): there are homeomorphisms of I that

---

[4] We define a surjective map $f : C \to I$ as follows: if $.x_1x_2x_3\ldots$ is the ternary expansion of $x \in C$ in which each x_k is 0 or 2, then we define $y_1, y_2, \ldots$ by $y_k = \frac{1}{2}x_k$ and let $f(x)$ be the real number in I whose binary expansion is $.y_1y_2y_3\ldots$. This map is non-decreasing and continuous. But it is not one-to-one: $\frac{1}{2}$ can be written in binary either as as $.0111\ldots$ or as $.100\ldots$—so f maps $\frac{1}{3}$, which admits the ternary expansion $.0222\ldots$, to the same point that it maps $\frac{2}{3}$, which admits the ternary expansion $.200\ldots$. For a characterization of the countable set of pairs of points at which this problem arises, see, e.g., Gelbaum and Olmsted (2003, §8.14).

[5] The lengths of the (disjoint) intervals deleted in constructing C sum to one.

[6] Each interval deleted from C_k in constructing C_{k+1} is open. So each C_k is closed—so C itself is closed. So in order to see that C is nowhere dense, it suffices to note that, since each such subinterval contains numbers that require ones in their ternary expansions, C has empty interior.

[7] See fn. 11 below

[8] There is, however, a sense in which these two notions provide equally good extensions of the notion of a small subset of the real line into the realm of uncountable sets. *Sierpiński–Erdős duality*: given the continuum hypothesis (or something somewhat weaker) it can be shown that there exists a bijection $g : \mathbb{R} \to \mathbb{R}$ that interchanges the nullsets and the meagre sets and which is its own inverse. So from a suitably lofty perspective, the nullsets and the meagre sets are functionally equivalent: any sentence that involves just set-theoretic terminology together with 'meagre' or 'nullset' is equivalent to the sentence that arises by swapping the roles of these two terms. See Oxtoby (1980, §19).

map the Cantor set to non-nullsets.[9] Non-nullsets that arise in this way are known as *fat Cantor sets.*[10] Each fat Cantor set is meagre, being the image of a meagre set under a homeomorphism—and by taking a countable union of well-chosen fat Cantor sets we can construct a meagre set whose complement is a nullset.[11]

Meagre and Residual. We have been focussing on the interval and its subsets. But in any topological space we can define the *meagre* subsets to be those that can be decomposed as countable unions of nowhere dense sets.[12] The complement of a meagre set is called *residual* or *comeagre.* Meagre subsets are also known as *sets of first category*; non-meagre subsets are also known as *sets of second category.*

Intuitively, in well-behaved spaces meagre subsets are negligibly small while residual subsets are overwhelmingly large.[13] Indeed, in the spaces to be considered below, countable sets are always meagre—and meagre sets are always small in a reasonably intuitive sense (namely, meagre subsets of such spaces have dense complements—so in particular, no nonempty open sets are meagre).[14]

---

[9]Recall from fn. 4 above that there is a non-decreasing surjective map $f : C \to I$. Let g be function from I to itself that maps $x \in I$ to $m(f([0, x] \cap C))$, where m is the Lebesgue measure on I. For $x \in I$ let $h(x) = \frac{1}{2}(x + g(x))$. This is a strictly increasing continuous map from I onto itself—and hence a homeomorphism (see, e.g., Dieudonné, 1960, §4.2.2). And since g is constant on the complement of C, the derivative of h on this set is $\frac{1}{2}$—so this set of measure one gets mapped to a set of measure one-half, which means that C must get mapped onto a set with the same measure. This is a special case of a more general result; see Oxtoby (1980, Theorem 13.2).

[10]Here is a recipe for constructing fat Cantor sets. Instead of deleting the middle third of each interval remaining at each stage in our construction of the Cantor set, we choose some $0 < \alpha < 1$, and construct C_1^α by deleting the middle open interval of length $\frac{\alpha}{2}$ from $C_0 = I$; construct C_2^α by deleting the two middle intervals of length $\frac{\alpha}{8}$ from C_1; and so on, deleting at each stage intervals of total length $\frac{\alpha}{2^n}$. The intersection C^α of the C_k^α has Lebesgue measure $1 - \alpha$ and is a fat Cantor set. See Gelbaum and Olmsted (2003, §§8.4 and 8.23).

[11] Let C^* be a countable union of fat Cantor sets C^{α_k} with $\alpha_k \to 1$. See, e.g., Gelbaum and Olmsted (2003, §§8.4 and 8.19).

[12]Recall that a subset A of a topological space X is called *nowhere dense* if the closure of A has empty interior.

[13]But consider the rational numbers and the natural numbers, each equipped with the topology that they inherit from the real numbers. In the case of the rational numbers, each singleton set is a nowhere dense set—so that the space as a whole is meagre. In the case of the natural numbers, singleton sets, being both open and closed, fail to be nowhere dense—so that even finite subsets are not meagre.

[14]Countable sets are meagre in any perfect space (i.e., in any space in which any open neighbourhood of any point includes at least two points); in any space

The so-called *Banach–Mazur game* provides an alternative characterization of meagre subsets that bolsters the intuitive force of their claim to be negligibly small. Let X be a topological space and let $\mathcal{G}$ be a basis for the topology of X: i.e., $\mathcal{G}$ is a collection of open sets of X such that every open set can be written as a union of sets in $\mathcal{G}$. A subset A of X is identified and a two-player game is to be played. Player 1 and Player 2, take turns choosing elements G_1, G_2, G_3, $\ldots$ of $\mathcal{G}$ with Player 1 choosing first and with each G_{i+1} constrained to be contained in G_i (for $i \geq 1$). Player 1 wins if $\cap G_k$ contains a point in A, otherwise Player 2 wins. Intuitively, the smaller A is, the harder it is for Player 1 to win—certainly, if A is all of X, then Player 1 cannot lose, while if A is countable (and X uncountable and well-behaved), it is easy for Player 2 to win, no matter what Player 1 does. It turns out that Player 2 has a winning strategy (a rule for selecting an interval at each stage, given what has gone on before, that is guaranteed to result in victory no matter what Player 1 does) if and only if A is meagre.[15] For applications, see fnn. 44, 48, and 49 below.

Below, our concern will be with the Bayesian framework, in which each agent has in view a space of live hypotheses over which prior and posterior probability measures are defined. In the usual case, spaces of hypotheses and the spaces of probability measures over them have a topological structure (one knows which sequences of points in these spaces converge and which do not), but they are not equipped with natural notions of distance or of measure.[16] So we will count meagre subsets of spaces of hypotheses or of priors as small and residual subsets as large. A property of hypotheses or of priors will count as typical just in case the set of objects exemplifying that property is residual in the relevant space.

For the sort of cases we will be concerned with, the various notions of largeness and smallness mesh in a interesting ways.

---

with the topology of a complete metric space, no nonempty open sets are meagre and every residual set is dense; see, e.g., Kechris (1995, §§8.A f). All of the spaces considered below are perfect and homeomorphic to complete metric spaces.

[15]See Oxtoby (1957, Theorem 1) or Kechris (1995, §8.H). Ulam bought Banach a bottle of wine for proving this for the unit interval—whether it were red or not, the Scottish Book maketh no mention.

[16]For a given space of hypotheses Θ, the corresponding space of priors will be the space of (countably additive) Borel probability measures over Θ, equipped with the topology of weak convergence: to say that the sequence of measures $(P_1, P_2, \ldots)$ converges to the measure P is to say that $\lim_{n\to\infty} \int f dP_n = \int f dP$ for each bounded continuous function f on Θ.

Remark 1. On the one hand: if X is a reasonable topological space with uncountably many points, then every Borel probability measure on X assigns measure zero to some residual subset of X.[17] So there is a sense in which each Bayesian prior embodies some very strong *a priori* bias that is immune to empirical refutation.

Remark 2. On the other hand: If R is a (measurable) meagre subset of a reasonable space of hypotheses, then the probability measures according to which R has measure zero form a residual subset of the space of probability measures over that space of hypotheses.[18]

3. Good News, Bad News

Turn we now to a family of problems from the heartland of Bayesian statistics, by way of getting more clear concerning what it means for a Bayesian agent to be guaranteed to converge to the truth—and to have a quick look at some representative facts about when this is possible.

Suppose that we are interested in a repeatable chance process with a countable number of possible outcomes—tossing a coin, spinning a roulette wheel, drawing a ball from an urn, etc. We assume that each outcome has a stable chance of occurring from repetition to repetition, with the outcomes of various repetitions being independent of one another. The propensities of the various outcomes are unknown. So there is a space Θ of *chance hypotheses* whose points correspond to the various ways these propensities might be: in the case of a coin, Θ will be the set of ordered pairs, (x_1, x_2), of nonnegative numbers that sum to one; in the case of a die, Θ will be the set of sextuples, $(x_1, x_2, \ldots, x_6)$ of nonnegative numbers that sum to one; in the case of a process with a countably infinite number of possible outcomes, Θ will be be the set of sequences, $(x_1, x_2, \ldots)$, of nonnegative numbers that sum to one. In each case, x_i gives the chance of getting outcome i on any given trial.

A Bayesian agent facing this sort of setup will begin with a prior probability measure P_0 over Θ—assigning various degrees of credence to various (sets of) chance hypotheses. After the first trial, P_0 will be updated by conditionalization to yield a posterior probability measure

[17] See Oxtoby (1980, §16). Note, however, that it is possible, e.g., to equip the unit interval with an unorthodox topology relative to which the meagre sets are precisely the sets of Lebesgue measure zero; see Oxtoby (1980, §22).

[18] This follows from Theorem 2.1 of Koumoullis (1996). Illustration: recall that according to the Lebesgue measure on the unit interval, the Cantor set is a set of measure zero whereas the set C^* of fn. 11 above has measure one; our result tells us that typical Borel probability measures consider both to be sets of measure zero.

P_1 on Θ; after the second trial, P_1 will be updated by conditionalization to yield a new posterior P_2 on Θ, and so on.

We are interested in the conditions under which the sequence of posteriors $(P_1, P_2, \ldots)$ generated by P_0 and an infinite sequence of outcomes converges to the truth. The idea is as follows. Suppose that $\theta_0 \in \Theta$ is the true chance hypothesis. Look at a sequence of outcomes generated by the process (given that θ_0 encodes the true propensities), and consider the corresponding sequence of posteriors $(P_1, P_2, \ldots)$ determined by these outcomes and by the given prior P_0. We say that this sequence of probability measures *converges to the truth* if as time goes on its terms becomes more and more tightly peaked around the true hypothesis θ_0. Here and below, to say that a sequence of probability measures $(P_1, P_2, \ldots)$ become more and more tightly peaked about a given hypothesis θ_0 is to say that the P_k converges (in the weak topology—see fn. 16 above) to the probability measure δ_{θ_0} that puts unit weight on θ_0.[19]

Clearly, there can be no general guarantee of success, even for Bayesian agents with eminently reasonable priors. Faced with an infinite sequence of tosses in which the limiting frequency of heads is one-third, any such agent will, as time goes on, become more and more certain that the bias of the coin is very close to one-third. But of course such a sequence of outcomes is possible even if the coin is in truth fair. So there can be no prior that is *absolutely* guaranteed to lead us to the right answer for this sort of problem.

But we can have something almost as good (although we have to put up with the unfortunate name that statisticians have given this good thing).

> We call a prior P_0 on Θ *consistent at* $\theta \in \Theta$ if according to the chance hypothesis θ, the chance of a sequence of outcomes arising that, together with P_0, would generate a sequence $(P_1, P_2, \ldots)$ of posteriors that did *not* become more and more tightly peaked around θ is zero. And we call a prior P_0 *consistent* if it is consistent at every $\theta \in \Theta$.

[19]In the present context, this boils down to the following (Diaconis and Freedman, 1986, §1). Let $N \in \mathbb{N} \cup \infty$ be the number of possible outcomes. For $\varepsilon > 0$ and m a natural number less than or equal to N, let $M_{m\varepsilon}$ be the set of $\phi \in \Theta$ that such that $|x_i - y_i| < \varepsilon$, for $i = 1, \ldots, m$, where x_i and y_i are the weights assigned to the ith possible outcome by θ_0 and by ϕ, respectively. $P_n \to \delta_{\theta_0}$ weakly if and only if for all m and ε, $P_n(M_{m\varepsilon}) \to 1$.

In the chance setting, a consistent prior is the best one can ask for: such a prior is essentially guaranteed to lead to the truth, in the sense that no matter which chance hypothesis is true, any non-pathological stream of data generated by that hypothesis would lead an agent with that prior to pile up more and more credence on smaller and smaller neighbourhoods of the true hypothesis.

For problems of the kind under consideration we have the following sort of result about subjective certainty:

> For any prior P, the set of hypotheses at which P is not consistent is a set of P-measure zero (Freedman, 1963, §1).

Now, the fact that a given prior assigns measure zero to the set of hypotheses at which it is inconsistent is compatible with that set being very large indeed. If I start out life certain that the coin has a one-third chance of coming up heads on any toss, then my prior is consistent for just one point in Θ.

A natural hope is that this sort of blatant *a priori* bias is the only obstruction to consistency—that any prior that was in some suitable sense spread over the entire space of hypotheses would turn out to be consistent.

We say that a hypothesis θ is in the *support* of prior P if P assigns positive measure to every open set containing θ. We say that a prior has *full support* if it assigns positive measure to every nonempty open subset of Θ.

> When the number of possible outcomes of our chance process is finite, a prior P is consistent at a hypothesis θ if and only if θ is in the support of P.[20] (Freedman, 1963, §3)

A prior is consistent at a hypothesis if it assigns that hypothesis a positive probability. But a prior can also be consistent at a hypothesis that it assigns zero probability, so long as it puts a finite amount of probability on each open set containing that hypothesis. The consistent priors are those that are nondogmatic in an intuitive sense: in this setting the consistent priors are just the priors of full support—those

[20]For $\varepsilon > 0$ and $\theta = (x_1, \ldots, x_n)$, let $B_\varepsilon(\theta)$ be the set of $(y_1, \ldots, y_n) \in \Theta$ such that $\sum(x_k - y_k)^2 < \varepsilon$. In the present context, a prior has θ in its support if and only if it assigns positive measure to each $B_\varepsilon(\theta)$.

that put positive measure on every open neighbourhood of every hypothesis. And the priors of full support form a residual family in the space of priors on Θ.[21]

The situation changes markedly when we consider chance processes with a countable infinity of possible outcomes.

> When the number of possible outcomes is infinite, for a prior P to be consistent at a hypothesis θ it is necessary (but not sufficient) that θ be in the support of P.[22] (Freedman, 1963, §5)

In this setting, for any two hypotheses $\theta_0, \theta_1 \in \Theta$, it is possible to find a prior P_0 with θ_0 in its support, such that conditionalizing P_0 on any data sequence that is nonpathological according to θ_0 leads to a sequence of posteriors $(P_1, P_2, \ldots)$ that become become more and more tightly peaked around θ_1 as more and more data is revealed (Freedman, 1963, Theorem 5). So an agent with prior P_0 can be expected to be misled into thinking that the true chance propensities are given by θ_1 when they are in fact given by θ_0.

Further, there are priors with full support that are consistent only at a meagre subset of Θ (Freedman, 1963, §5, Remark 6). Indeed, the priors of full support that are consistent only at a meagre set of hypotheses form a residual set in the space of priors defined on Θ.[23] So there is a sense in which the typical prior is of full support—but fails to converge to the truth in typical situations. But this is not because

[21]Since we will need it repeatedly, let us wheel out some heavy artillery. If a space is compact and metrizable, so is the space of Borel probability measures on that space when equipped with the weak topology (Kechris, 1995, Theorem 17.22). And in any compact metrizable space, any set that is dense and arises as the intersection of a countable number of open sets is residual (Oxtoby, 1980, Theorems 9.1 and 9.2). And within the set of probability measures on a compact metrizable space, the set of measures of full support is dense and arises as a countable intersection of open sets (Dubins and Freedman, 1964, §3.13). And, finally, the Θ presently under consideration are compact and metrizable (Freedman, 1963, 1387).

[22]In this context, each $\theta \in \Theta$ corresponds to an infinite sequence $(x_1, x_2, \ldots)$ of weights assigned to each possible outcome (relative to some fixed enumeration of the outcomes). For each $\theta \in \Theta$ and $\varepsilon > 0$, let $B_\varepsilon(\theta)$ be the set of $\phi = \{y_k\}$ in Θ such that $\sum_k 2^{-k-1}(x_k - y_k)(1 + x_k - y_k)^{-1} < \varepsilon$. A prior on Θ includes $\theta \in \Theta$ in its support if and only if it assigns positive measure to $B_\varepsilon(\theta)$ for each $\varepsilon > 0$.

[23]See Freedman (1965), the Corollary of which tells us that priors consistent only at a meagre set of hypotheses form a residual subset of the space of priors, while Remark 2 tells us that the set of priors of full support is residual in the space of priors. The intersection of two residual sets is residual.

the problem is so-contrived as to put convergence to the truth beyond the pale of possibility: there exist consistent priors for this problem.[24]

Where does this leave us? In ideally well-behaved settings the posteriors of Bayesian agents whose priors have full support are essentially guaranteed to converge to the truth—and such priors are typical. But in the last case considered typical priors fail to converge to the truth in typical situations—although there still exist special priors guaranteed to converge to the truth in every situation under consideration. For some problems, convergence to the truth is all but automatic for Bayesians, for others, it is very difficult (but not impossible) to come by.

Remark 3. We have been concerned with the most straightforward sort of statistical inference problem, in which the hypotheses under consideration are very closely connected with the chance distributions of observations. In many problems, this connection is more subtle. In some problems, convergence to the truth is in general impossible—and is not expected by typical Bayesian agents—because multiple hypotheses induce the same chance distribution on the space of sequences of measurement outcomes, so that observation of outcomes provides no way to discriminate between hypotheses. Statisticians call this phenomenon *nonidentifiability.* It is easy to construct silly examples of nonidentifiability. Example: suppose that one is told the outcomes of a sequence of tosses of a coin and asked to determine whether the coin is biased two-thirds in favour of heads or of tails and also whether it is a nickel, a dime, or a quarter. One expects that this evidence will allow one to correctly identify the bias of the coin but not its denomination.[25]

Remark 4. Nonidentifiable models aside, statisticians have a rule of thumb that being of full support should suffice for a prior to be consistent when the space of hypotheses is finite-dimensional (*parametric models*) but not when the space of hypotheses is infinite-dimensional (*nonparametric models*).[26] Even in the nonparametric setting, sufficient conditions are known for a prior to be consistent—often they involve requiring that the problem be identifiable, that the prior avoid

[24]See the discussion of tail-free priors in Freedman (1963, §6) Close relatives of these priors are natural generalizations of the priors associated with Carnap's continuum of inductive methods; see Skyrms (1993).

[25]For more interesting examples, see, e.g., Schervish (1995, 430), Skyrms (1991, §4), Sober and Steel (2002), or Steel *et al.* (1994).

[26]On the importance of this divide, and for parametric exceptions to the general rule of thumb, see Ghosh and Ramamoorthi (2003, Chapter 1).

certain sorts of jaggedness, and that it spread credence over the space of hypotheses in some sense more demanding than having full support.[27]

4. Good News, Rotten News

Let us now take as our space of hypotheses the space $\mathcal{C}$ of infinite binary sequences. We will need a few facts about this space, known as the *Cantor space.* $\mathcal{C}$ is the result of taking the product of $\{0, 1\}$ with itself infinitely many times and as such carries the natural product space topology—the topology of pointwise convergence of sequences.[28] We can characterize the open sets in $\mathcal{C}$ as follows: let $w = x_1 x_2 \dots x_n$ be a string of bits; then the set B_w consisting of those sequences whose first n bits are given by w is an open set and every open set arises as a union of sets of this form. It is straightforward to set up a correspondence between the points in the Cantor space and the points of the standard Cantor set: the mapping $\boldsymbol{x} = (x_1, x_2, \dots) \in \mathcal{C} \mapsto \sum \frac{2x_k}{3^k} \in C$ is a homeomorphism.

Think of each sequence in $\mathcal{C}$ as corresponding to a possible sequence of records of outcomes of some family of observations. We might be looking at the outcomes of sequence of coin tosses (perhaps with each toss involving the same coin, perhaps with different tosses involving coins of different biases). Or we might be looking at successive bits in a binary expansion of a constant of nature, or determining whether there is more gold in China or in India, minute by minute.

We will be interested in Bayesian agents who begin life with priors over $\mathcal{C}$, then update by conditionalization as the bits of data making up some particular sequence are revealed one by one.

In the present context, conditionalization works as follows. Consider the first bit learned. Using the notation introduced above, B_0 and B_1 (the set of sequences that begin with 0 and the set of sequences that begin with 1) are open sets. When you learn the first bit of data, you set your prior to zero on one of these sets and, up to normalization, leave it as it is on the other (i.e., from now on, you assign measure zero to any set contained in one of B_0 and B_1, but leave unchanged the ratio between the measures of sets contained in the other). And similarly, for each new bit you learn—if the first bit was 1, then after seeing that bit your posterior is concentrated on B_1. After seeing the

[27] See Ghosh and Ramamoorthi (2003, Chapter 4) and Wasserman (1998).

[28] Let $\boldsymbol{x} = (x_1, x_2, \dots)$ be a point in $\mathcal{C}$. And for each $j = 1, 2, \dots$ let $\boldsymbol{x}^j = (x_1^j, x_2^j, \dots)$ be a point in $\mathcal{C}$. Then $\boldsymbol{x}^j \to \boldsymbol{x}$ in this topology if and only if: for each k, there is a J_k such that $x_k^j = x_k$ for all $j > J_k$.

second bit, you have a new posterior that vanishes either on B_{10} or on B_{11} (depending on what the second bit was). And so on.[29]

It is easy to construct priors of full support in this setting. For example, the *fair coin* measure, that assigns measure $\frac{1}{2^n}$ to any open set of the form $B_{x_1 \ldots x_n}$ is a measure of full support. So is any measure that assigns non-zero weight to each point in a countable dense subset of $\mathcal{C}$ (since then any open set includes points assigned positive measure).

Consistency in the sense of §3 is easy to come by for this problem: every prior of full support is consistent. Given the way that conditionalization works for this problem, whichever the true sequence $\boldsymbol{x}_0$ is, the result of conditionalizing a prior P_0 of full support on longer and longer initial segments of $\boldsymbol{x}_0$ is to deliver a sequence of posterior probability measures $(P_1, P_2, \ldots)$ whose support is limited to a sequence of smaller and smaller open sets that shrink towards $\boldsymbol{x}_0$.[30] So as time goes on, these posteriors become as tightly peaked as you like around the true hypothesis.[31] So convergence to the truth is guaranteed for priors of full support—and priors of full support are typical in the space of priors.[32]

Nonetheless, there is a sense in which convergence to the truth can be formidably difficult in this setting: as we will see, there are questions we could ask our Bayesian agents about the sequences they are seeing such that then any prior will fail to converge to the truth for infinitely many sequences in $\mathcal{C}$—and such that priors that display a certain sort of attractive open-mindedness fail to converge to the truth for typical hypotheses in $\mathcal{C}$.

[29]As always, it will not be obvious how to define the posterior if the the prior is concentrated on a region ruled out by the data. Of course, this problem cannot arise for priors of full support.

[30]Full support is required in order to assure the the posteriors are well-defined. See fn. 29 above.

[31]That is, the P_n converge weakly to $\delta_{\boldsymbol{x}_0}$. This follows from Billingsley (1999, Theorem 1.2.2): suppose that the class of subsets $\mathcal{A}$ of some topological space X is closed under finite intersections and that every open subset of X is a countable union of sets in $\mathcal{A}$; then if Q, Q_1, Q_2, ... are measures on X such that if $Q_k(A) \to Q(A)$ for every $A \in \mathcal{A}$, then the Q_k converge weakly to Q. Now, in our case the class of B_w satisfy these hypotheses. So we need only show that for any binary string w, $\lim_{k\to\infty} P_k(B_w) = \delta_{\boldsymbol{x}_0}(B_w)$. To see this, note that if $\boldsymbol{x}_0$ is in B_w then $\delta_{\boldsymbol{x}_0}(B_w) = 1$—and for for sufficiently large k, the support of P_k is contained in B_w so that $\lim_{k\to\infty} P_k(B_w) = 1$; while if $\boldsymbol{x}_0 \notin B_w$, then $\delta_{\boldsymbol{x}_0}(B_w) = 0$—and for sufficiently large k, B_w is outside the support of P_k so we have $\lim_{k\to\infty} P_k(B_w) = 0$.

[32]$\mathcal{C}$ has the topology of the Cantor set—which as a closed subset of a compact metrizable space is compact and metrizable. So the results quoted in fn. 21 above apply to probability measures on $\mathcal{C}$.

Let R be any countable dense subset of $\mathcal{C}$.[33] Now, to say that R is a dense subset $\mathcal{C}$ is to say that every point in $\mathcal{C}$ is a limit of a sequence of points in R—or, equivalently, that every for any finite binary string, w, the set B_w contains an element of R. Further, each $B_w \subset \mathcal{C}$ is an uncountable set—and so must also have non-empty intersection with the complement of R. So the complement of our countable dense R is also dense in $\mathcal{C}$. Examples of countable dense subsets of $\mathcal{C}$: the set of binary sequences that correspond to binary expansions of rational numbers; the set of binary sequences that are eventually constant; the set of binary sequences that are periodic.[34]

For the remainder of this section we will set our Bayesian agents the following problem: determine whether or not the true sequence $\boldsymbol{x}$ lies in R.[35] Since the prior P_0 and the sequence $(P_1, P_2, \ldots)$ of posteriors sum up the Bayesian's opinion at each stage, the natural method for a Bayesian facing this problem will be to be guided at each stage by $P_n(R)$.

What would count as getting at the truth here? Well, a rather undemanding criterion would be that *either* $x \in R$ and for sufficiently large n, $P_n(R)$ is always greater than or equal to one-half; *or* $x \notin R$ and for sufficiently large n, $P_n(R)$ is always less than one-half. So let's say that for the problem currently under consideration, the *failure set* for a prior is the set of hypotheses for which this condition does not hold. As one would expect, there is the usual sort of guarantee of subjective certainty of convergence to the truth:

> Let P_0 be any prior on $\mathcal{C}$. Then the corresponding failure set has P_0-measure zero.[36]

The question, of course, is what the failure sets of various priors look like. Some priors correspond to agents who begin life certain

---

[33] R is of course (Borel) measurable, being a countable union of singleton sets (each of which is closed).

[34] A sequence of either of the latter two kinds is specified by specifying a finite binary string—so these sets are countable. To see that they are dense, note that an arbitrary binary sequence $\boldsymbol{x}$ arises as a limit of a sequence $\boldsymbol{x}^1$, $\boldsymbol{x}^2$, … drawn from either set—for each k, let $\boldsymbol{x}^k$ be an eventually constant or periodic binary sequence that begins with the same first k bits as $\boldsymbol{x}$.

[35] More generally, we might set our Bayesians the problem of determining whether the true sequence belongs to a dense, measurable set S with dense complement. Everything below would still go through, except for the argument of fn. 40.

[36] This is a consequence of some very general results of Schervish and Seidenfeld (1990). Applied to our special case, their Theorem 2 guarantees that the set of hypotheses x such that *neither* $x \in R$ and $P_n(R) \to 1$ *nor* $x \notin R$ and $P_n(R) \to 0$ is a set of measure zero. Their Corollary 5 justifies us in treating every subset of this set as a set of measure zero—and the failure set is itself such a subset.

that the true sequence lies in R, and who remain thus no matter what data they see—for these priors, the failure set is the uncountable dense complement of R in $\mathcal{C}$.[37] Other priors correspond to agents who begin life certain that the true sequence x lies outside of R, and who maintain this certainty no matter what data they see—for such priors, the failure set is the countable dense set R.[38]

Either sort of prior can have full support.[39] So in the present context, being of full support does not suffice to guarantee convergence to the truth—indeed it is compatible with being entirely closed-minded about whether or not the true sequence belongs to R.

What would it take to truly leave the door open concerning the question whether the true sequence belongs to R? Since this problem has the feature that any finite amount of data is consistent both with the true sequence being in R and with it not being in R, it is natural to consider a condition along the following lines.

> Let P_0 be a measure on $\mathcal{C}$ and R be a (measurable) subset of $\mathcal{C}$. P_0 is *open-minded for* R if, for any data set an agent with P_0 as prior sees, there are extensions of this data set that would lead the agent to assign a credence of at least one-half to the proposition that the true sequence lies in R, and other extensions of the data that would lead the agent to assign this proposition credence less than one-half.

Agents with open-minded priors never make up their minds irredeemably on the basis of a finite amount of data. It is not hard to see that open-minded priors exist for countable and dense R of the sort we are concerned with.[40]

---

[37]To construct such a prior, enumerate the hypotheses in R, then put $\frac{9}{10}$ unit of credence on the first, $\frac{9}{100}$ on the second, etc.

[38]The fair coin measure is an example—it assigns measure zero to each singleton, and so assigns measure zero to R as a whole.

[39]The fair coin measure and the measures of fn. 37 have full support.

[40] Any prior P_0 that is a non-trivial mixture of the fair coin measure with a measure of the sort discussed in fn. 37 is open-minded. As usual, we write P_n for the posterior generated from P_0 and n bits of data. Let us also write a_n for $P_n(R)$ and b_n for $1 - a_n$. In order to show that our P_0 is open-minded, it suffices to show that after seeing k bits: (i) if $a_k \geq b_k$, then there is a way of continuing that data set to include m bits such that $a_m < b_m$; and (ii) if $a_k < b_k$, then there is a way of continuing that data set to include m bits such that $a_m \geq b_m$. Since all that matters is the ratio of a_m to b_m, let us ignore normalization—update P_n by just killing off the part of it that lives on the region inconsistent with the bit revealed at stage $(n + 1)$. In order to see (i), note that at each stage, no matter what bit turns up next, the measure of the complement of R will go down by a factor of

We need one more definition before getting down to business.

> We say that hypothesis $\boldsymbol{x}$ *flummoxes* prior P_0 if there are infinitely many m for which $P_m(R) \geq \frac{1}{2}$ and infinitely many m for which $P_m(R) < \frac{1}{2}$.

Agents faced with flummoxers flip-flop *ad infinitum* on the question whether it is at least as likely as not that the true sequence lies in R.

(1) *Any prior whose failure set was empty would have to be open-minded.* No matter what data an agent has been shown, that data is consistent with both $\boldsymbol{x} \in R$ and $\boldsymbol{x} \notin R$. If the prior has an empty failure set, it has to prepared for either eventuality—so there must be ways of continuing that would lead to credence above one-half and others that would lead to credence below one-half.

(2) *Any open-minded prior is flummoxed by some hypothesis.* Given an open-minded prior, we can construct a flummoxing hypothesis as follows.[41] Start off by showing some data that leads to a credence greater than one-half that the true sequence is in R. There are sequences that begin this way but which are not in R (since the complement of R is dense in $\mathcal{C}$). Pick one of them, and continue the data stream by showing a block of bits from this sequence long enough to force the agent's credence that the true sequence is in R below one-half—there must be such a continuation, since the agent is open-minded. Now—whatever string of bits has been shown to the agent, there is some sequence in R that begins this way (since R is dense in $\mathcal{C}$). So, since the agent is open-minded, we can choose a continuation of the data stream that will eventually force the agent's credence that the sequence being revealed lies in R above one-half. Continue this bait and switch *ad infinitum.* In this way we construct a binary sequence that flummoxes our open-minded agent.

---

two. Meanwhile, the measure of R decreases by a factor of nine whenever a new bit is shown that rules out the member of R of maximum probability among those consistent with the data thus far. So by choosing at each stage to show a bit of this kind, we can ensure that after a finite number of new bits, $a_m < b_m$. In order to see (ii), let $\boldsymbol{x}$ be a point in R consistent with the data through stage k. Then for some j, $P_k(\boldsymbol{x}) = 9 \times 10^{-j}$. So if we extend the data set by showing bits from $\boldsymbol{x}$, we can ensure that $a_m \geq 9 \times 10^{-j}$, even while b_n gets halved every time a new bit is revealed. So after a finite number of bits are revealed on this scheme, $a_m \geq b_m$, as desired.

[41]The argument of this paragraph is essentially the proof of Proposition 12 in Kelly and Glymour (1989).

(3) *No prior has an empty failure set.* This follows immediately from the preceding two points and the fact that every flummoxer of a prior is in its failure set.

(4) *Each prior has an infinite failure set that is dense in the space of hypotheses.* Let P be any prior and let $\boldsymbol{x} = (x_1, x_2, \ldots)$ be any binary sequence in our space of hypotheses. We need to show that there is a sequence $\boldsymbol{x}^1 = (x_1^1, x_2^1, \ldots)$, $\boldsymbol{x}^2 = (x_1^2, x_2^2, \ldots)$, $\boldsymbol{x}^3 = (x_1^3, x_2^3, \ldots)$, ... of hypotheses in the failure set of P that converges to $\boldsymbol{x}$ (this will establish that the failure set of P is dense in $\mathcal{C}$—and hence infinite). It suffices to show that that for each $k = 1, 2, 3, \ldots$ there is an $\boldsymbol{x}^k$ in the failure set of P that has the same first k bits as $\boldsymbol{x}$.[42] So for k arbitrary, let w_k be the initial k-bit string of $\boldsymbol{x}$. There are two cases: there is some flummoxer of P that begins with w_k, or there is not. If there is, then we can take this flummoxer as $\boldsymbol{x}^k$. If not, then for every $\boldsymbol{y} \in B_{w_k}$, there is some point in the binary sequence $\boldsymbol{y}$ after which P has made up its mind irredeemably, and can no longer be caused to flip-flop on the question whether the sequence being revealed lies in R. In particular, this is the case for $\boldsymbol{x}$ itself. Suppose that P makes up its mind irredeemably after being shown the first m bits of $\boldsymbol{x}$. So no matter how we continue the data sequence after w_m (the first m bits of $\boldsymbol{x}$), we always find that P gives us the same answer about whether it is more likely than not that the binary sequence being revealed lies in R. But this means that for some binary sequences in B_{w_m}, P settles on the *wrong* answer—since some sequences in B_{w_m} lie in R and others do not. So in this case we can let $\boldsymbol{x}^k$ be one of these binary sequences that P is wrong about.

(5) *The set of flummoxers of an open-minded prior is residual in* $\mathcal{C}$. We can think of a prior as a machine that, when fed a binary sequence bit by bit, outputs a sequence of Y's and N's (telling us, at each stage, whether or not $P_n(R) \geq \frac{1}{2}$). For a given prior P and a given data sequence $\boldsymbol{x}$, we say that the prior *never switches* for that data sequence if the prior outputs a sequence that is all Y's or all N's. We say that it *switches at least once* if in the output sequence begins with one or more Y's followed by one or more N's (or vice versa). We say that it *switches at least twice* if the output begins with some Y's followed by some N's followed by some Y's (or vice versa). And so on. Consider an open-minded prior P. It is not hard to see that for each $k = 0, 1, 2, \ldots$, the set S_k of $\boldsymbol{x} \in \mathcal{C}$

[42] For in this case, we know that for each j, $x_j^n = x_j$ for $n > j$—so we have $\boldsymbol{x}^k \to \boldsymbol{x}$ pointwise.

for which the prior switches at least k times is open and dense in $\mathcal{C}$.[43] Now, the set of flummoxers of P is just the set of points in $\mathcal{C}$ that get P to switch at least k times for each $k = 1, 2, \ldots$—i.e., the set of flummoxers of P is just the intersection of all the S_k. But a countable intersection of open dense sets is residual. So the set of flummoxers of P is residual in $\mathcal{C}$.[44]

(6) *Open-minded priors fail to converge to the truth for typical hypotheses when faced with the problem of deciding whether or not the sequence they are being shown lies in R.* The set of flummoxers of a given prior is a subset of its failure set. So the failure set of an open-minded prior is residual in the space of hypotheses. So, in particular, the failure set is an uncountable set, dense in the space of hypotheses.[45]

5. Orgulity

As we have just seen, if a Bayesian agent is shown a binary sequence bit by bit, and asked after each bit is revealed for the probability that the sequence as a whole has a property of a certain type—such as being a binary expansion of a rational number, or being periodic, or being eventually constant—then there will be a rich family of sequences that the agent could be shown that would frustrate convergence to the truth on this question. Agents that are open-minded, in the sense that they

[43]For suppose $\boldsymbol{x}_0$ causes P to switch at least k times. Then there must be some n such that the first n bits of $\boldsymbol{x}_0$ already suffice to make P switch k times. So P must switch at least k times for every $\boldsymbol{x} \in B_w$ (where w is the n-bit initial segment of $\boldsymbol{x}_0$). So S_k is open. And if $\boldsymbol{y}$ is any point in $\mathcal{C}$, then we can construct a sequence $(\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots)$ of points in S_k that converge to $\boldsymbol{y}$: each $\boldsymbol{y}^m$ starts with the first m bits of $\boldsymbol{y}$, then continues in some way that makes P change its mind k times (possible since P is open-minded). The $\boldsymbol{y}^n$ converge pointwise to $\boldsymbol{y}$—for each j, we know that $y^n_j = y_j$ for $n > j$. So S_k is dense in $\mathcal{C}$.

[44] Here is an alternative argument via the Banach–Mazur game. Take $\mathcal{G}$ to be class of (non-empty) sets of form B_w, so that we can think of Player 1 and Player 2 as taking turns specifying finite binary strings which are then concatenated to yield a binary sequence, with Player 1 winning if and only if the resulting sequence lies in the target set A. Let A be the set of non-flummoxers of prior P. Player 2 has a winning strategy: no matter what Player 1 does, Player 2 plays a string of bits that causes P to output Y followed by a string that causes it to output N. This always results in an infinite sequence that flummoxes P. Since Player 2 has a winning strategy, the set of non-flummoxers must be meagre in $\mathcal{C}$.

[45]Further, the failure set of an open-minded prior is a set of measure one according to typical probability measures (Remark 2 above)—and not because typical measures hide their mass in some corner of the space of hypotheses (fn. 32 above).

never make up their minds irredeemably based on a finite amount of evidence, are frustrated in this way by typical binary sequences.

This in itself does not represent a shortcoming of Bayesianism: define a *method* to be a function from finite binary strings to guesses about whether the sequence being revealed has the target property; and define an *open-minded* method to be one that never makes up its mind irredeemably based on a finite amount of evidence, and so on; then analogues of (1)–(6) of Section 4 carry over to this more general setting.

What sets Bayesian agents apart is that, despite the patent intractability of the problem at hand, they assign probability zero to the set of sequences that they cannot handle successfully.[46]

Suppose that I face a problem of type under discussion. I am inclined to be a Bayesian and have extraordinary powers of voluntary belief. So I sit down to look at some priors to decide which to adopt. One factor among many that I will weigh up in evaluating a prior is the nature of its failure set—the set of binary sequences with the feature that if I am shown them, I will fail to latch on to the truth about whether the sequence has the property of interest. The failure set of each prior I consider is nontrivial—always infinite and dense, in some cases even residual. And each prior considers its own failure set to be negligible, assigning it measure zero. I may find a prior that I like on other grounds, but worry that its failure set is rather large—but tinkering with it will never get me very far. For example, if a prior P has a countable failure set, I can construct a new prior P^* arbitrarily similar to P that converges to the truth for every sequence in the failure set of P—but now P^* will fail to converge to the truth for infinitely many sequences that P was able to handle.[47]

Reflecting on this I may well, even while endorsing my favourite method as well-suited to this problem, also think that there is some chance that nature will be unkind and frustrate my desire to reach the truth. According to Bayesianism, this combination of thoughts is incoherent: to commit to a method is to commit to a prior; and (for

---

[46]In light of the Remark 1 above, it was a foregone conclusion that each Bayesian agent has to assign probability zero to some residual set—but one might have hoped that the sets in question would be recondite and of no obvious epistemological interest.

[47]You can construct P^* by rescaling P to make available an arbitrarily small amount of credence, then parcelling it out among the sequences in the failure set of P. P^* will have an infinite failure set, disjoint from that of P. The failure set of P^* may even be residual whereas the failure set of P was meagre—e.g., P may be the fair coin measure and P^* may be a measure of the type considered in fn. 40 above.

this problem) no prior licenses the belief that there is some chance that it will fail to lead to the truth in the long run.

Bayesian convergence to the truth theorems tell us that Bayesian agents are forbidden to think that there is any chance that they will be fooled in the long run—even when they know that their credence function is defined on a space that includes many hypotheses that would frustrate their desire to reach the truth. This is a bizarre limitation—and one that ought to be unwelcome to those who are attracted to Bayesianism because of its apparent flexibility.

Let me wind up by heading off two lines of response that may seem appealing.

(i) Some Bayesians may be attracted to a view on which an agent with a given prior is entitled to consider all hypotheses in its failure set to be mere skeptical hypotheses. Collectively, the set of skeptical scenarios *ought* to be assigned measure zero. So there is no embarrassment here, no matter how large the failure set should turn out to be.

But that would be a strange way to think of the present problematic. We are interested in the behaviour of a Bayesian agent's posteriors in the infinite long run, in a setting in which every false hypothesis is eventually ruled out definitively by observation and in which these posteriors converge to the delta function measure concentrated on the true hypothesis. By ordinary standards, the hypotheses in the failure set of a prior are not skeptical scenarios.

(ii) Bayesians who countenance substantive restrictions on rational priors my be be tempted to forbid open-minded priors for problems of the type under consideration.

This is in itself a strange idea. Suppose, for instance, that I am interested in some physical system whose dynamics may well be chaotic and that the binary sequence I am to be shown encodes some information about the history of the system (imperfect example: it encodes which of two points on the surface of Hyperion is closer to the centre of Saturn, second by second). Suppose, further, that by examining this sequence bit by bit, I aim to determine whether or not the sequence as a whole is periodic. I begin by thinking that the system's dynamics is likely chaotic—but that there is some chance that it is instead periodic. If it is chaotic, then I expect that the sequence I am shown will not be periodic—but that it will mimic periodic sequences for long stretches of time.[48] If the system is periodic, the sequence will be periodic—but whatever reasons I have to think that the system is likely chaotic are also reasons to think that if the sequence is periodic, it will be very complicated. Is it really plausible that rationality requires me to think that there is some possible finite evidence set that would decide the question one way or the other, no matter what evidence should later c0me in?

In any case, there are close relatives of our problem that afflict all priors, open-minded or not. Suppose that a Bayesian agent is shown a binary sequence $(a_1, a_2, \ldots)$ encoding, for each of a sequence of days, whether or not it snows at a given location. On day k, our agent is required to announce the probability p_{k+1} of snow on the next day, given the pattern of snowy and snowless days to date. The *discrepancy* on day k is $p_k - a_k$. A Bayesian agent is *high-low calibrated* relative to a given binary sequence if the mean discrepancy goes to zero when we specialize to days on which $p_i \geq .5$ and also to days on which $p_i < .5$. If a forecaster fails to be high-low calibrated, then a rival can complain that there is a mismatch, for days on which the forecaster predicted (or did not predict) snow, between the actual relative frequency of snowy days and the forecast probabilities of snow. All Bayesian forecasters assign probability zero to the set of sequences for which they are not high-low calibrated (Dawid, 1982). And all Bayesian forecasters fail to be calibrated for typical data sequences.[49]

Some have seen in the tendency of Bayesian agents to converge to the truth—and in related results concerning the eventual merger of opinion between Bayesian agents whose initial credences share a certain amount of common ground—the materials for acquitting personalist

[48]Let $A \subset \mathcal{C}$ consist of those sequences that do *not* start with some string w_k repeated k times, for each $k = 1, 2, 3, \ldots$. Consider a Banach–Mazur game of the type described in fn. 48 above with A as the designated set. Player 2 has a winning strategy: on the nth turn, play n repetitions of whatever has been played so far. So A is meagre. This means that even though typical sequences in $\mathcal{C}$ are not periodic (the set of periodic sequences is countable—and so meagre), typical sequences begin by repeating some string twice, some string three times, some string four times, etc.

[49]Consider a Banach–Mazur game of the type described in fn. 48 above, with the designated set A consisting of sequences for which a given forecaster is high-low calibrated. Player 2 has a winning strategy: each bit played is designed to make the forecaster look bad (so a 1 is played as the kth bit if and only of $p_k < .5$); and enough bits are played on each turn to make the mean discrepancy on days for which snow is forecast go above .25 of the mean discrepancy for days on which no snow is forecast go below $-.25$. For further discussion, see Belot (unpublished).

Bayesianism of the charge of excessive subjectivity.[50] But recent philosophical commentators (some Bayesians among them) have tended to downplay the significance of these results, pointing out that they what they guarantee is that Bayesian agents *think* that there is no chance that their own future opinions will fail to converge to the truth, which is not the same thing as saying that the opinions of each Bayesian agent are *in fact* certain to converge to the truth.[51] Thehe truth concerning Bayesian convergence to the truth results is significantly worse than has been generally allowed—they constitute a real liability for Bayesianism by forbidding a reasonable epistemological modesty.

References

Belot, Gordon (unpublished) "Failure of Calibration is Typical."

Billingsley, Patrick (1999) *Convergence of Probability Measures*. New York: Wiley, second edition.

Dawid, Philip (1982) "The Well-Calibrated Bayesian." *Journal of the American Statistical Association* 77: 605–610.

Dawid, Philip (1985) "Calibration-Based Empirical Probability." *The Annals of Statistics* 13: 1251–1273.

Diaconis, Persi and David Freedman (1986) "On the Consistency of Bayes Estimates." *The Annals of Statistics* 14: 1–67.

Dieudonné, Jean (1960) *Foundations of Modern Analysis*. New York: Academic Press.

Dubins, Lester and David Freedman (1964) "Measurable Sets of Measures." *Pacific Journal of Mathematics* 14: 1211–1222.

Dudley, R.M. (1964) "On Sequential Convergence." *Transations of the American Mathematical Society* 112: 483–507.

Earman, John (1992) *Bayses or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Edwards, Ward, Harold Lindman, and Leonard J. Savage (1963) "Bayesian Statistical Inference for Psychological Research." *Psychological Review* 70: 193–242.

Freedman, David (1963) "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case." *Annals of Mathematical Statistics* 34: 1386–1403.

[50]For examples, see: Edwards *et al.* (1963); Howson and Urbach (1989, §§10.a and 11.j) (but cf. their 2006, §§2.d and 8.a); Savage (1954, §§3.6 and 4.6); and Schervish and Seidenfeld (1990). For further references and discussion, see Earman (1992, §6.3).

[51]For discussions emphasizing this point, see: Earman (1992, §6.6), Glymour (1980, 73), Howson (2000, 210), and Kelly *et al.* (1997, §2).

Freedman, David (1965) "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case. II." *Annals of Mathematical Statistics* 36: 454–456.

Gelbaum, Bernard and John Olmsted (2003) *Counterexamples in Analysis*. New York: Dover.

Ghosh, J.K. and R.V. Ramamoorthi (2003) *Bayesian Nonparametrics*. New York: Springer–Verlag.

Glymour, Clark (1980) *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Howson, Colin (2000) *Hume's Problem: Induction and the Justification of Belief*. Oxford: Oxford University Press.

Howson, Colin and Peter Urbach (1989) *Scientific Reasoning: The Bayesian Approach*. La Salle, Il: Open Court, first edition.

Howson, Colin and Peter Urbach (2006) *Scientific Reasoning: The Bayesian Approach*. La Salle, Il: Open Court, third edition.

Kechris, Alexander (1995) *Classical Descriptive Set Theory*. New York: Springer–Verlag.

Kelly, Kevin (1996) *The Logic of Reliable Inquiry*. New York: Oxford University Press.

Kelly, Kevin and Clark Glymour (1989) "Convergence to the Truth and Nothing but the Truth." *Philosophy of Science* 56: 185–220.

Kelly, Kevin, Oliver Schulte, and Cory Juhl (1997) "Learning Theory and the Philosophy of Science." *Philosophy of Science* 64: 245–267.

Koumoullis, George (1996) "Baire Category in Spaces of Measures." *Advances in Mathematics* 124: 1–24.

Oxtoby, John C. (1957) "The Banach–Mazur Game and the Banach Category Theorem." In Melvin Dresher, Albert Tucker, and Philip Wolfe (eds.), *Contributions to the Theory of Games*, Princeton, NJ: Princeton University Press, volume III, 159–163.

Oxtoby, John C. (1980) *Measure and Category: A Survey of the Analogies between Topological and Measure Spaces*. Springer–Verlag, second edition.

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

Schervish, Mark J. (1995) *Theory of Statistics*. New York: Springer–Verlag.

Schervish, Mark J. and Teddy Seidenfeld (1990) "An Approach to Consensus and Certainty with Increasing Evidence." *Journal of Statistical Planning and Inference* 25: 401 –414.

Skyrms, Brian (1991) "Carnapian Inductive Logic for Markov Chains." *Erkenntnis* 35: 469–460.

Skyrms, Brian (1993) "Carnapian Inductive Logic for a Value Continuum." *Midwest Studies in Philosophy* 18: 78–89.

Sober, Elliott and Michael Steel (2002) "Testing the Hypothesis of Common Ancestry." *Journal of Theoretical Biology* 218: 395–408.

Steel, Michael, László Székely, and Michael Hendy (1994) "Reconstructing Trees When Sequence Sites Evolve at Variable Rates." *Journal of Computational Biology* 1: 153–163.

Wasserman, Larry (1998) "Asymptotic Properties of Nonparametric Bayesian Procedures." In Dipak Dey, Peter Müller, and Debajyoti Sinha (eds.), *Practical Nonparametric and Semiparametric Bayesian Statistics*, New York: Springer–Verlag, 293–304.

Department of Philosophy, University of Michigan
E-mail address: belot@umich.edu