August 7, 11, 13, 2009

Cosmology and Inductive Inference: A Bayesian Failure

John D. Norton[1]
Department of History and Philosophy of Science
Center for Philosophy of Science
University of Pittsburgh

For updates, see www.pitt.edu/~jdnorton

Prepared for "Philosophy of Cosmology: Characterising Science and Beyond"
St. Anne's College, Oxford
September 20-22, 2009

It is natural for scientists to employ a familiar formal tool, the probability calculus, to give quantitative expression to relations of partial evidential support. However this probabilistic representation is unable to separate cleanly neutral support from disfavoring evidence (or ignorance from disbelief). Since this separation is essential for the analysis of evidential relations in cosmology in the context of multiverse and anthropic reasoning, the use of probabilistic representations may introduce spurious results stemming from its expressive inadequacy. That such spurious results arise in the Bayesian "doomsday argument" is shown by a reanalysis that employs fragments of inductive logic able to represent evidential neutrality. Similarly, the mere supposition of a multiverse is not yet enough to warrant the introduction of probabilities without some analog of a randomizer over the multiverses. The improper introduction of such probabilities is illustrated with the "self-sampling assumption." A concluding heretical thought: perhaps the values of some cosmic parameters are

[1] I am grateful to Jeremy Butterfield and Eric Hatleback for helpful discussion.

unexplained by current theory simply because no explanation is possible; they just are what they are.

1. Introduction[2]

A corner of philosophy of science that may prove useful to cosmologists is the study of inductive inference. Recent work in cosmology has identified certain observed features of the cosmos as surprising and uses this identification to motivate new theories or forms of explanation. The judgment that these features are surprising makes essential use of inductive notions as does the injunction that the appropriate response is a new theory or new forms of explanations.

My purpose in this note is address a specific way in which these inductive notions are expressed. It is common both within philosophy of science and outside it to take inductive notions just to be probabilistic notions, even though probabilistic notions are, I believe, only a special case of inductive notions and cannot always be applied. Of course, these probabilistic notions can often be very useful. However, sometimes they are not. They can be quite misleading. Their use in connection with surprises in cosmology is such a case.

The difficulty is that probabilistic notions cannot separate two cases. There is the case in which evidential support is completely neutral over the possibilities; and the case in which evidence disfavors particular ones. Probabilistic notions fare poorly at capturing neutrality, but are good at capturing disfavor. If one insists that probabilistic notions should be used in cases of

---

[2] My thanks go the organizers of this conference for giving me the opportunity to participate. It is an appealing thought that philosophers may have something useful to say to cosmologists and may aid them in the profound conceptual and theoretical challenges they face. It is an appealing thought in the abstract, but it is a daunting one when this particular philosopher is invited to be useful. The word "philosopher" has its origins in the Greek, where its root is a "*lover* of wisdom." There is no assurance that a lover of wisdom has any, just as an anglophile is not assured to have an Englishman locked in the basement. For a broader picture of how a philosopher of science approaches the philosophical problems of cosmology, see John Earman, "Cosmology: A Special Case?" in this conference.

evidential neutrality, both cases end up being represented by the same low probability values and this paucity of expressive power leads to a conflation of neutrality and disfavor. Evidential neutrality warrants fewer definite conclusions than does evidential disfavor; so mistakening the second for the first can lead to spurious conclusions that are merely an artifact of a poor choice of inductive logic.

In the following note, I will describe how this confusion has arisen in cosmology. It arises in connection with what I call the "Surprising Analysis," outlined in Section 2 below, along with the Bayesian attempt to use probability theory to give a more precise account of it. In Section 3, I will argue that the Bayesian reformulation fails since it cannot distinguish neutrality of support from disfavor; and I will show how the extreme case of completely neutral support must be represented in an inherently a non-probabilistic way. In Section 4, I will sketch how this state could be incorporated into alternative inductive logics. Section 5 will illustrate how such alternative logics can be used to good effect. There we shall see how the implausible results of the Bayesian "doomsday argument" arise as an artifact of the inability of the Bayesian system to represent neutral evidential support. A reanalysis in an inductive logic that can express evidential neutrality no longer returns the implausible results. Sections 6 will review how probabilistic representations can properly be introduced into cosmology. An ensemble provided by a multiverse, it is argued, is not enough. Some analog of a randomizer is also needed. The "self-sampling assumption" illustrates what happens when there is no analog to the randomizer. Finally, Section 7 will entertain the idea that the values of cosmic parameters that initiate the Surprising Analysis may admit no further explanation after all.

2. The Surprising Analysis

In recent work in cosmology, the following analysis has become common. We find some aspect of our cosmos, often a parameter, whose value is left indeterminate by our present science. Its value must be found by observation. The observed value is then reported and judged to be surprising. It is surprising that our cosmic spatial geometry is so close to flat or that values of certain fundamental constants lie within the very narrow boundaries that permit just our form of life to develop. We then pass to a demand for an explanation of why the cosmic feature has this value. In the cases cited, explanations are located in cosmic inflation and anthropic reasoning respectively.

The analysis has three phases:

Surprising Analysis (informal version)

1. Establishment that prior theory is neutral with regard to the particular cosmic feature.
2. The claim that the specific value observed for the feature is surprising and in need of explanation.
3. The provision of the explanation to which we should infer.

This analysis is, at its heart, an instance of inductive reasoning.[3] The cosmic features identified in 1. and 2. are taken to provide inductive support for the explanation offered in 3. It is also an imprecise analysis in so far as certain central notions—"surprising" and "explanation"—are not well understood. As a result, it has proven attractive to recast the analysis in a form that uses the probability calculus as a means of representing the otherwise vaguely indicated evidential relations. For then the imprecision of the informal version is replaced by a precise computation. Sometimes this analysis is given as a full-blown Bayesian analysis. At other times it is merely alluded to by allowing fragments of probability talk to slip into the discussion.[4] In either case, it is clearly comfortable for physicists to replace vaguer inductive notions with a precise mathematical instrument, probabilities, that are the standard tool for dealing with uncertainties in physics.

A generic recasting of the surprising analysis in Bayesian terms can proceed as follows.

Surprising Analysis (Bayesian version)

i. There is some cosmic or physical parameter k whose value is left indeterminate by our prior theories, all collected into our background knowledge "B". We represent that

---

[3] Here I construe the notion of inductive inference very broadly. It extends well beyond its older meaning of generalization from particulars; it includes all cases in which propositions are judged to favor or disfavor other propositions while not deductively establishing or refuting them.

[4] Reviewing the articles collected in Carr (2007), for example, one finds probabilities appearing in full-blown Bayesian analyses, in casual mentions and much in between.

indeterminateness by a probability distribution p(k|B) which is widely spread over the admissible values of k. In particular, the observed value k_{obs} has low probability.[5]

ii. Were some theory T to be the case, then the probability of k_{obs} would be much higher; that is the likelihood $p(k_{obs}|T\&B)$ is large.

iii. Bayes' theorem in the form

$$\frac{p(T \mid k_{obs} \,\&\, B)}{p(T \mid B)} = \frac{p(k_{obs} \mid T \,\&\, B)}{p(k_{obs} \mid B)}$$

now assures us that the numerical ratio $p(T|k_{obs}\&B)/p(T|B)$ is large so that the evidence k_{obs} lends strong support to the theory T. For the posterior probability $p(T|k_{obs}\&B)$ of the theory T given the observation is much greater than its prior probability p(T|B), conditioned only on our background knowledge B.

3. What is Wrong with the Bayesian Version

3.1 Evidential Neutrality versus Disfavoring

While we should try to replace of a vaguer analysis with a more precise one, the replacement will only be successful if the more precise instrument used is the appropriate one. In many cases, inductive relations are well explicated by probabilistic analysis. These are cases in which the systems inherently contain probabilities, such as stochastic systems or random sampling processes. However there are other cases in which evidential relations are, contrary to Bayesian dogma, poorly represented by a probabilistic analysis.[6] The Surprising Analysis above is such a case. That is already suggested by the discomfort many feel at the talk of the probability of the observed parameter k_{obs} given only background knowledge, that is, when there is no presumption of whether the theory T or some other theory--who knows which--is true. Equally mysterious are other quantities that appear in a full Bayesian analysis such as the likelihood of the observed parameter k_{obs} given that the theory T does *not* obtain, $p(k_{obs}|\text{not-}T\&B)$. For how

---

[5] More exactly, observation will pick out an small interval surrounding k_{obs} to which the probability density p(k|B) will assign a small value.

[6] For a survey of these boundaries of Bayesian applicability, see Norton (manuscript).

can we reasonably comprehend the expanse of possibilities arising when our favored theory is not the case? The idea that assigning these probabilities is an arbitrary and even risky project appears often in the multiverse literature.[7] However that concern does not seem to stop the use of these probabilities or to trigger consideration of alternative ways of representing evidential relations.

The most serious failing of the Bayesian reformulation of the Surprising Analysis lies in the first step (i). In the original analysis, our background theory is neutral in its support of any particular value of k. That neutrality is represented by a probability distribution. The difficulty is that probability distributions are unable to represent evidential neutrality. The numerical values of probability measures span a range from zero to one. High values near one represent strong support; low values near zero represent strong disfavoring. This complementary relationship between support and disfavoring results directly from the additivity of probability measures.[8] For a proposition A, we have

$$P(A|B) + P(\text{not-}A|B) = 1$$

So, if we assign a low probability near zero to P(A|B), then we must assign a probability near unity to P(not-A|B). That high probability means that the evidence strongly supports not-A, which means it must strongly disfavor its contradiction, A, to which we assigned the low probability. More generally, if there are n mutually exclusive and exhaustive outcomes A_1, …, A_n, additivity requires

$$P(A_1|B) + P(A_2|B) + \ldots + P(A_n|B) = 1$$

or, in other words, that the measure is normalized to unity. This normalization condition means that evidence can only favor one outcome or set of outcomes if it disfavors others.

The additivity of probabilities is the mathematical expression of the complementary relationship of support and disfavoring. It leaves no place in the representation for neutrality. We shall see in several cases below of neutral support that additivity, in the form of a requirement

[7] See for example Aguirre (2007), Page (2007, p. 422), Tegmark (2007, pp. 121-22).

[8] For further discussion, see Norton (2007, Section 4.1).

that probability measures normalize to unity, is directly responsible for the failure of the probabilistic analysis.[9]

3.2 Representing Evidential Neutrality

The difficult is that the full spectrum of evidential support cannot simply be represented by the degrees of a one-dimensional continuum, such as the reals in [0,1]. The full spectrum forms a multi-dimensional space with, loosely speaking, disfavoring and neutrality proceeding in different directions. I know of no adequate theoretical representation of this space. However we can discern what a small portion of it looks like. Let us take the case of *complete* evidential neutrality. The background B, in this case, simply has no evidential bearing in any direction on the contingent[10] propositions $A_1, \ldots, A_n$, of some finite Boolean algebra of propositions. Two instruments independently give us the same characterization of this state of evidential neutrality.

3.2.1 Invariance under Negation

Assume we have a finite Boolean algebra of propositions such that our background knowledge B is completely neutral with respect to its contingent propositions, $A_1, \ldots, A_n$. Let the degree of support in this state of complete neutrality be written as "$[A_i|B]$". Its value will be unchanged in this special, extreme case if we replace A_i by its negation, not-A_i.

$$[A_i|B] = [\text{not-}A_i|B]$$

The point is, I think, quite straightforward. If you are unsure, try this thought experiment. The proposition A_i is written on a card in a sealed envelope. You have formed your neutral degree

---

[9] Why doesn't a probability of 1/n represent evidential neutrality when there are n options? The deeper problems with all probabilistic representations of neutrality emerge in the following section. For a foretaste, note that the "neutral value" of 1/n is now context dependent. If there are 5 options, 1/5 represents neutrality. But if there are ten, 1/5 represent strong favoring. Compare this with strong support. A probability of 0.999 is strong in every context.

[10] Contingent propositions are just all those that are not the always-false contradiction F or the always-true tautology T, whose truth values are same no matter how the world may be. Contingent proposition may be either true or false according to the state of the world.

$[A_i|B]$. You then learn that, in error, the negation not-A_i was written on the card so you must now adjust your degree. No change would be needed.

Thus, the degrees of support representing evidential neutrality are unchanged if we replace contingent propositions by their negations. This invariance condition is very strong. If we couple it with a weak condition of "monotonicity" it is easy to show that the only admissible set of degrees assigns the same value "I" (for Indifferent or Ignorance) to all contingent propositions:[11]

Completely neutral support

$[A_i|B] = I$ for all contingent propositions

For n>2, no probability measure can provide this uniform set of values; for it always assigns the same value to the disjunction $A_i vA_k$ and each of its disjunctive parts A_i and A_k.

The condition of monotonicity invoked above requires that we have a comparative relation on degrees of support such that, when A deductively entails C, the degree [A|B] is not greater than the degree [C|A]. This merely requires that the deductive consequences of a proposition are at least as well supported as the proposition. (For a full elaboration of the derivation see Norton, 2008, Section 6.2-6.3.)

3.2.2 Invariance under Disjunctive Refinement

A second way of arriving at this distribution of neutral support is afforded by a fixture in many accounts of probability theory, the so-called "paradoxes" of the principle of indifference. Here's a rendering of them. Assume that there are two coins lying on a table. Our background has no information about how they came to be there; it does *not* tell us they were tossed. So our background in quite neutral to how many heads are showing: zero, one or two. That is,

$$[0H \mid B] = [1H \mid B] = [2H \mid B]$$

---

[11] What of the standard Bayesian device of seeking to represent complete neutrality as a convex set of probability measures? It has problems, such as a failure to assign any single degree of support locally. The deepest problem is that the convex set is still not invariant under negation. Rather, under negation, the set becomes a set of what I have elsewhere called "dual additive measures." See Norton (2007a).

The evidence is also neutral over the four possibilities arising if we distinguish the coins, so that the outcome 1H has two mutually exclusive disjunctive parts HT, the first coin shows the head, or TH, the second coin shows the head.

$$[HH \mid B] = [HT \mid B] = [TH \mid B] = [TT \mid B]$$

We can combine these last two sets of equalities with the relations 0H = HH and 1H = HT v TH to recover

$$[HT \vee TH \mid B] = [HT \mid B] = [TH \mid B]$$

That is, the proposition HT v TH has the same degree of support as each of its mutually exclusive, disjunctive parts, HT and TH.

In the literature on probability theory, this outcome is taken to be paradoxical since it can be satisfied non-trivially by no probability measure. It should now be clear that there is no paradox. The paradox is generated by the presumption that a probability measure can represent evidential neutrality in the first place. The better response is that this analysis returns a general property of a state of completely neutral support: For any two mutually exclusion contingent proposition A_i and A_k (such that their disjunction $A_i \vee A_k$ is still contingent), we have

$$[A_i \vee A_k \mid B] = [A_i \mid B] = [A_k \mid B]$$

It is easy to see that applying this condition to all contingent propositions in the algebra rapidly returns the above distribution of completely neutral support. (For details and an entry into the massive literature on the principle of indifference and its paradoxes, see Norton, 2008.)

3.3 Neutrality and Disfavor versus ignorance and disbelief.

Those familiar with the literature will find the last discussion non-standard. It is everywhere expressed in terms of evidential support. The now dominant subjective Bayesians have replaced all such talk with talk of "degrees of belief." Neutrality becomes ignorance; disfavoring becomes disbelief.

This transformation has merged two notions that should be kept distinct. One is the degree to which this body of propositions inductively supports that proposition. These degrees are objective matters, independent of our thoughts and opinions. The second is the degrees of belief that you or I may decide to assign to various bodies of propositions. Once we add our thoughts and opinions, these degrees will likely vary from person to person according to our individual prejudices.

For those of us interested only in inductive inference, the transformation has been retrograde. The evidential relations that interest us are obscured by a fog of personal opinion. This concern has led to a revival of so-called "objective Bayesianism," which seeks to limit the analysis to objective relations. (For discussion, see Williamson, 2009.) A persistent problem facing this objective approach is that a probability measure cannot supply an initial neutral state of support, for the reasons just elaborated above. As result, objective Bayesians cannot realize the goal of a full account of learning from evidence that takes us by Bayesian conditionalization from an initial neutral state to our final state. Since any initial state must be some probability distribution, it always expresses more relations of support and disfavoring that we are entitled to in an initial state.

Subjective Bayesians seek to escape the problem by declaring these relations in the initial state as mere, ungrounded opinion that may vary from person to person. The hope is that, in the long run, continued conditionalization will wash away this unfounded opinion from the mix, leaving behind the nuggets of evidential warrant. While limit theorems purport to illustrate the process, it has long been recognized that the mix remains in the short term of real practice. It will be helpful for the further discussion to illustrate the problem.

3.4 Pure Opinion Masquerading as Knowledge

Let us assume that some cosmic parameter can take a countably infinite set of values $k=k_1$, $k=k_3$, $k=k_3$,… We have no idea of which is the correct value, so we assign a prior probability arbitrarily. Its variations encode no knowledge, but just the arbitrary choices made in ignorance. Since there are infinitely many possibilities, our probability assignments must eventually decrease without limit, else the total probability will not sum to unity. Let us say that, with the decrease needed, we assign the following two prior probabilities

$$P(k_{135}|B) = 0.00095 \qquad P(k_{136}|B) = 0.00005$$

We now learn that one of these two values of k is the correct one. That is, we acquire evidence $E = k_{135} \vee k_{136}$. Bayes theorem in the ratio form assures us that

$$\frac{P(k_{135} \mid E \,\&\, B)}{P(k_{136} \mid E \,\&\, B)} = \frac{P(k_{135} \mid B)}{P(k_{136} \mid B)} = \frac{0.00095}{0.00005}$$

Since the two posterior probabilities must sum to unity, it now follows that

$$P(k_{135}|E\&B) = 0.95 \qquad P(k_{136}|E\&B) = 0.05$$

We have become close to certain of k_{135} and strongly doubt k_{136}. Yet it is clear that our strong preference for k_{135} is entirely an artifact of the pure opinion encoded in our priors.

In sum, the Bayesian reformulation of the Surprising Analysis fails in its first step when it attempts to use a probability distribution to represent neutrality of support. If this Bayesian reformulation fails, how can we respond? In the following, I will pursue three distinct approaches. In the first (Sections 4-5), I will investigate the use of an alternative inductive logic that can tolerate the neutrality of support just outlines. In the second (Section 6), I will investigate what would be needed to bring back a probabilistic analysis. Finally, in the third (Section 7), I will ask if the problem is inductive at all.

4. Inductive Logics that Tolerate Neutrality of Support

We have no good characterization of the multidimensional space of degrees that accommodates both disfavoring and neutrality of evidence. As a result we have no complete inductive logic that accommodates both.[12] However it is possible to discern how such a logic might deal with conditionalization that proceeds from the initial state of completely neutral support.

To see how such a logic might proceed, it is helpful to note that the Bayesian system can be decomposed into parts, as I have done in detail in Norton (2007). It turns out that the additivity of the system is independent of the components that give the system its characteristic dynamics under conditionalization. Those dynamics rests on a procedure that I have called "refute and rescale." To conditionalize a theory T on evidence E, we first preserve just that part of T logically compatible with E, that is, T&E ("refute"). We then rescale the probabilities of everything that survives in proportion to its prior probability. This two-part process is captured by the simple theorem

$$\frac{P(T_1 \mid E \,\&\, B)}{P(T_2 \mid E \,\&\, B)} = \frac{P(T_1 \,\&\, E \mid B)}{P(T_2 \,\&\, E \mid B)} = \frac{P(T_1 \mid B)}{P(T_2 \mid B)}$$

where the second equality holds only in the special case in which each of T_1 and T_2 entail the evidence E.

[12] I have tried to survey the terrain of possible logics in Norton (manuscript a). It includes a sample "partial ignorance logic."

It follows immediately for this special case that if the priors $P(T_1|B) = P(T_2|B)$, then the equality persists for the posteriors $P(T_1|E\&B) = P(T_2|E\&B)$. This last result can be imported as a postulate into an inductive logic that tolerates complete neutrality of support:

Conditionalizing from complete neutrality of support

If our background knowledge B is completely neutral with respect to two theories T_1 and T_2, so that $[T_1|B] = [T_2|B] = I$, and both theories entail the evidence E, then $[T_1|E\&B] = [T_2|E\&B]$.

The virtue of this rule is that it immediately solves the subjective Bayesian problem of "Pure Opinion Masquerading at Knowledge." All we do is to replace the probabilistic prior by a neutral prior and conditionalize according to the last rule. We then have for our priors

$$[k_{135}|B] = [k_{136}|B] = I$$

Since each of k_{135} and k_{136} entails the evidence $E = k_{135} \vee k_{136}$, we can apply the above rule of conditionalization to get the result we should have arrived at before

$$[k_{135}|E\&B] = [k_{136}|E\&B]$$

Our background support did not treat k_{135} and k_{136} differently; the evidence E did not treat them differently; so the combined support of background and evidence should not treat them differently.

There are other techniques that implement notions of neutrality of support. They include invariances under parameter transformation and will be illustrated in the next section. Unfortunately, I am not confident that any of them can be used to supply an alternative, more precise version of the Surprizing Analysis.

5. The Doomsday Argument

A further illustration of these alternative logics can be found in a re-analysis of the doomsday argument. This argument has been connected with the notion of multiverses by Bostrom (2007). See also Bostrom (2002, Ch. 6-7) for an introduction to the literature on the argument.

In its Bayesian form, the argument purports to give remarkable results on a foundation that seems too slender. Re-analysis that employs a more careful representation of neutrality of support can no longer reproduce these results, revealing that the strong results are merely an artifact of the defective probabilistic representation of neutral support.

5.1 The Bayesian Analysis

Consider a process, such as our universe, that may have a life of T years, where T can have any value. What do we learn about T from our learning that the process has already persisted for t years? We begin by assigning a prior probability distribution p(T|B) to T. We assign a likelihood to our learning that the process has persisted t years:

$$p(t|T\&B) = 1/T$$

The rationale is that we, the observer, have no reason to expect that we will be realized in one portion of the T year span than in any other. We now apply Bayes' theorem in the ratio form for two different values of T, both greater than t:

$$\frac{p(T_1 \mid t \,\&\, B)}{p(T_2 \mid t \,\&\, B)} = \frac{p(t \mid T_1 \,\&\, B)}{p(t \mid T_2 \,\&\, B)} \cdot \frac{p(T_1 \mid B)}{p(T_2 \mid B)} = \frac{T_2}{T_1} \cdot \frac{p(T_1 \mid B)}{p(T_2 \mid B)}$$

If $T_1 < T_2$, it now follows that conditionalizing on our evidence shifts support to T_1 by increasing the ratio of probabilities in T_1's favor by a factor of T_2/T_1. That is, the evidence of t shifts support differentially to all times T closer t. More compactly, we have

$$p(T \mid t \,\&\, B) \propto 1/T$$

If T is the time of the end of the world, we are to believe it is coming sooner rather than later.

There is, of course, some room to tinker. A notable candidate is the likelihood p(t|T&B) = 1/T. It amounts to saying we are equally likely to be realized in any year in the process. Since there are more people alive later in the universe's history, a better analysis might scale the likelihood according to how many people are alive. This merely amounts to using a different clock. Instead of the familiar clock time of physics, we rescale to a people clock

$$T' = n(T) \qquad t' = n(t)$$

where the function n(.) gives the number of people alive at the time indicated. The new analysis uses a likelihood

$$p(t'|T'\&B) = 1/T'$$

It will proceed exactly as before and arrive at the same conclusion. Support shifts to a sooner end.

One surely cannot help but feel a sense that this is something for nothing. We have supplied essentially no information to the analysis. We know there is a process; we have no idea how long it will last; we know it has lasted t years. On this meager basis, we somehow are supposed to believe that it will end sooner.

It is also clear that the favoring of earlier times is an artifact of the additivity of the probability measures used. For the analysis depends essentially on the likelihood p(t|T&B)=1/T, which varies according to T. The idea that likelihood was trying to express was merely that, even with a T chosen, no value of t in the admissible range t=0 to t=T is preferred. That uniformity could be expressed by merely setting p(t|T&B) to a constant. The additivity of probabilities, however, requires that all probabilities sum to unity. As a result that constant must vary with different values of T as 1/T so that

$$\int_{t=0}^{T} p(t \mid T \ \& \ B)dt = \int_{t=0}^{T} \text{constant } dt = \text{constant}.T = 1.$$

So it is additivity that forces the result. Yet this additivity is just the formal property of probability measures that precludes them properly representing the evidential neutrality appropriate to this case.

5.2 The Barest Re-analysis

This illusion that we get something for nothing starts to evaporate once we re-analyse the problem in a way that eschews the troublesome additivity of the probability measures and more adequately incorporates neutrality of support. Here is a very bare version. We will start will completely neutral support

$$[T_1|B] = [T_2|B] = I$$

Let us take the evidence of t merely to reside in the logically weaker assertion that we know T>t. Call this E. It now follows that the hypothesis of any T greater than t entails the evidence. Hence we can use the rule of conditionalization of Section 4 and infer that

$$[T_1|E\&B] = [T_2|E\&B] = I$$

That is, knowing that the end, T, must come after t, gives us no basis for discriminating among different end times T_1 and T_2.

What should we do if we do want to incorporate the further information that some fixed t is more or less likely if different T are the case? A return to the Bayesian analysis will show us a way to proceed.

5.3 The Bayesian Analysis Again

The Bayesian analysis of Section 5.1 is only a fragment of a fuller Bayesian analysis. When we explore that fuller analysis, we find the Bayesian analysis fails. Where it founders is on a requirement that the analysis should be insensitive to the units used to measure time.

To see how this comes about, consider the posterior probability, as delivered by Bayes' theorem:

$$p(T \mid t \,\&\, B) = p(t \mid T \,\&\, B) \cdot \frac{p(T \mid B)}{p(t \mid B)} = \frac{1}{T} \cdot \frac{p(T \mid B)}{p(t \mid B)}$$

for T>t. What seems unknowable is the ratio of priors p(T|B)/p(t|B). It turns out, however, that that the ratio must be a constant, independent of T (but not necessarily independent of t). This follows from the requirement that the analysis proceeds the same way no matter what system of units we use—whether we measure time in days or years. To assume otherwise would not be unreasonable. If, for example, the process is the life span of an oak tree, we know that its average life span is 400-500 years. With this time scale information in hand, we should expect a very different analysis of the time to death if our datum is that the oak is 100 days old or 100 years old. However that is a different problem; the doomsday problem as posed provides no information on the time scale and no grounds to analyze differently according to the unit used to measure time.

To proceed, we assume that there is a single probability density p(.|.) appropriate to the analysis, so that the problem is soluble at all; and, to capture the condition of independence from units of time, we assume that the same probability density p(.|.) is used whichever unit is used to measure time. This entails that the probability density p(.|.) is invariant under a linear rescaling of the times t and T (that, for example, corresponds to changing measurements in years to measurements in days):

$$t' = At \qquad T' = AT$$

This is a familiar condition applied standardly to prior probabilities that are functions of some dimensioned quantity T. Such a probability distribution, it turns out, must be the "Jeffreys prior," which is:[13]

---

[13] See, for example, Jaynes (2003, p. 382). The probability assigned to the small interval dT must be unchanged when we change units. That is: p(T|t&B)dT = p(T'|t'&B)dT'. Since T'=AT,

$$p(T|t\&B) = C(t)/T \qquad \text{for } T>t$$

where C(t) is a constant, independent of T.

The difficulty with this probability density in T is that it cannot be normalized to unity. The summed probability over all time T diverges:[14]

$$\int_{T=t}^{\infty} p(T \mid t) dT = \int_{T=t}^{\infty} (C(t)/T) dT = \infty$$

The Bayesian literature has learned to accommodate such improper behavior in prior probability distributions. The key requirement is that, on conditionalization, the improper prior probability distribution must return a normalizable posterior probability distribution. Here, however, the improper distribution is already the posterior distribution. So the failure is not merely a familiar failure of the Bayesian analysis to provide a suitable prior probability; it is its failure to be able to express a distribution of support over different times independent of units of measure.

The failure of normalization of probability is not easily accommodated. It immediately breaks connections with frequencies. While we may posit that ratios of the finite-valued probabilities are approximated by ratios of frequencies of the corresponding outcomes in the usual way, there is no comparable accommodation for outcomes with infinite probability. Their ratios are ill-defined.

We may wish to proceed nonetheless, interpreting the unnormalized probabilities just as degrees of support in some variant inductive logic. The result is curious. Consider the degree of support assigned to the set of end times T in any finite interval T_1 to T_2:

$$P(T_1 < T < T_2) = \int_{T_1}^{T_2} p(T \mid t \,\&\, B) dT = \int_{T_1}^{T_2} C(t)/T \cdot dT = \text{finite}$$

The degree assigned to the set of end times greater than some nominated T_2

$$P(T > T_2) = \int_{T_2}^{\infty} p(T \mid t \,\&\, B) dT = \int_{T_2}^{\infty} C(t)/T \cdot dT = \infty$$

---

we have dT'/dT = A = T'/T, so that p(T|t&B).T = p(T'|t'&B).T', from which the Jeffreys prior follows immediately.

[14] We could restore normalizability by positing an upper cut-off to the time T. That would be ad hoc and would contradict the formulation of the problem by introducing extra information: doom must come before the cutoff. Is such an artifice really preferable to admitting that the probabilistic representation is just poorly matched to the problem?

As a result, finite degree is assigned to any finite interval of times; and, no matter how big a finite interval we take, an infinite degree is always assigned to the set of times that comes after. Since support must follow the infinite degree, all support is accrued by arbitrarily late times. No matter how large we take T_2 to be, all support must be located on the proposition that the end time T comes after it. The standard doomsday argument assures us that, on a pairwise comparison, more support is accrued by the earlier time for doom. This extended analysis agrees with that. It adds, however, that, when we consider the support accrued by intervals of times, maximum possible support shifts to the latest possible times.

5.4 A Richer Analysis

The analysis of the last section shows two things: the unsustainability of the Bayesian analysis and the power of invariance requirements. Here is a way that invariance requirements can be brought to bear on the problem. We seek the degree of support $[T_1, T_2|t]$ for an end time in the interval T_1 to T_2 given by the observation that the process has progressed to time t. We assume both T_1 to T_2 are greater than t.

The Bayesian analysis of Section 5.1 required that we know which of all possible clocks is the correct one in the sense that the likelihood of our observation is uniformly distributed over its time scale. Of course it is virtually impossible to know which is the right one. We somehow need to judge how the cosmos is distributing our moments of consciousnesses as observers. Are they distributed uniformly in time? Are they distributed uniformly over the volumes of expanding space? Are they distributed uniformly over all people; or weighted according to how long each person lives? Are they distributed uniformly over all people or all people and primates with advanced cognitive functions? Or is the distribution weighted to favor beings according to the degree of advancement of their cognitive functions?

Let us presume that there is such a preferred clock in this analysis as well. In addition, we assume that we have no idea from out background knowledge which is the correct clock. As a result, we must treat all clocks the same. This condition is an invariance condition. The degrees of support assigned to various intervals of time must be unchanged as we rescale the clocks used to label the times. A consequence of this invariance is that the degrees of support assigned to all finite intervals must be the same; that is, for any $T_2>T_1>t$ and any other $T_4>T_3>t$, we will have[15]

$$[T_1, T_2|t] = [T_3, T_4|t] = I$$

This will still be the case if either interval in a proper subinterval of the other. In this regard, after conditionalization on t, we have a distribution with the properties of completely neutral support. For this reason, I give the single universal value the symbol "I", as before.

That is, contrary to Bayesian analysis, learning that t has passed does not invest us with miraculous, oracular powers of prognostication. On that evidence, we have no reason to prefer any finite time interval in the future over any other.[16]

6. Bringing Back Probabilities

6.1 A Mere Ensemble is Not Enough

When one recognizes the difficulty of using inductive logics other than the Bayesian system, one might yearn to bring back the probabilistic analysis. In some cases that is appropriate. How are we to know when a particular inductive logic is applicable? The "material theory of induction" (Norton, 2003, 2005) was developed to answer precisely this question. Its essential idea is that the warrant for an inductive inference is not a universal formal template; it is a locally obtaining matter of fact. Imagine, for example, that we are given a Uranium atom randomly sampled from

---

[15] To see this, consider any monotonic rescaling f of the clock with the properties: $t'=f(t)=t$; $T_1' = f(T_1) = T_3$; and $T_2' = f(T_2) = T_4$. Since we have only relabeled the times, the degrees of support must be unchanged so that $[T_1, T_2|t] = [T_1', T_2'|t']' = [T_3, T_4|t]'$. The prime on $[.,.|.]'$ indicates that we are using the rule for computing degrees of support pertinent to the rescaled clock. The invariance, however tells us that both original and rescaled systems use the same rule, so that the two function $[.,.|.]$and $[.,.|.]'$ are the same. Hence $[T_1, T_2|t] = [T_3, T_4|t]$ as claimed.

[16] This result does not automatically extend to intervals open to infinity. However it is clear that a minor alteration of the analysis will return $[T_1, \infty|t] = [T_2, \infty|t] = I^*$ for any $T_1>t$ and $T_2>t$. It is plausible that some further condition will give us the stronger $[T_1, \infty|t] = [T_1, T_2|t]$, so that $I^*=I$. However I do not think invariance conditions are able to force it.

natural Uranium ore and asked to form beliefs over whether it will decay radioactively over the next year. We will draw on the facts of the distribution of different isotopes of Uranium, that the sampling was random and on the probabilistic law of radioactive decay to compute a probability of decay over the next year. That probability is a serviceable degree of support accorded by the information at hand to decay over the next year.[17]

There are many such circumstances in which a probabilistic logic is warranted. A case relevant to cosmology arises when we have an ensemble and, in addition, some scheme that lends evidential support to its members in the complementary fashion discussed in Section 3.1 above. That is, a set A of ensemble members is favored by the scheme just to the extent that the complementary set is disfavored. The simplest example of such a scheme is a randomizer. It assigns a high or low physical chance to the set A just to the extent that it assigns a reversed low or high chance to the complement. The point is a familiar one. We do not have a probability of 1/52 of the ace of hearts when we merely have a deck of cards. We must in addition shuffle it well and deal a card before we have the probability. Without this randomizer, merely having neutral evidential support for all cards is insufficient to induce the probabilities.

A prosaic case of such an ensemble and randomizer arose for young residents of English port towns in centuries past. Then press gangs roamed the streets, kidnapping drunken victims, who would then awaken below decks on the high seas. That awakening would give the unfortunates good (anthropic!) reason to believe that they were one of the unlucky victims of the press gangs and decry their bad fortune in realizing an outcome of lower probability. The facts needed to warrant the probabilistic analysis are in place. There is an ensemble, the community of young, tavern-frequenting men in a port town; and a randomizer, the process of roaming press gangs selecting their victims.[18]

---

[17] The facts that warrant a probabilistic analysis need not be facts about physical probabilities. Imagine that one is at a racetrack placing bets with a "Dutch bookie" and that the constellation of assumptions surrounding the Dutch book arguments obtain. (See Howson and Urbach, 2006, Ch. 3.) These facts warrant one conforming one's inductive reasoning with the probability calculus—but only as long as these facts obtain.

[18] The reasoning is anthropic. However the illustration also reflects the puzzling and even dubious nature of anthropic explanation. The awakening on the ship does not explain how the

In the cosmology literature, there are efforts to use the physical facts of the cosmology to ground the assigning of probabilities to the components of a multiverse.[19] This is the right way to proceed, although there is always scope for the facts invoked to fall short of what is needed. One such way arises when the proposal supplies an ensemble but no analog of the randomizer. The proposal developed in Gibbons et al. (1987), Hawking and Page (1988) and Gibbons and Turok (2008) employs a Hamiltonian formulation of the cosmological theories and derives its probabilities from the naturally occurring canonical measures in them.

At first this seems promising since it is reminiscent of the natural measure of the Hamiltonian formulation of ordinary statistical physics. In this latter case, the association of a probability measure with the canonical phase space volume is underwritten by some expectation of a dynamics that is, in some sense, ergodic.[20] That means we expect the dynamical evolution to be such that the system will spend roughly equal times in equal volumes of phase space, as it explores the full extent of the phase space. This behavior functions as a randomizer. It allows us to connect frequencies of occupation of a portion of the phase space with its phase volume, so that the familiar connection between frequencies and probabilities is recoverable. In the Gibbons et al. proposal, however, it is not clear that such ergodic-like behavior is expected. It is not clear that we should expect that, over time, a single model will explore a fuller part of the model space. Rather, the proposal is justified by the remark (p. 736):

> Giving the models equal weight corresponds to adopting Laplace's 'principle of indifference', which claims that in the absence of any further information, all outcomes are equally likely.

---

victim came to be there in any of the usual modes of explanation. It is not a causal explanation: the cause was the kidnapping by the press gang. It is not the factor that raises the probability of going to sea, as is required in statistical relevance explanation. That probability raiser is the victim's visiting and drinking in portside taverns. The awakening at sea does not explain *why* the victim was kidnapped. Rather it provides evidence *that* the victim was kidnapped, which is an assurance that an otherwise a lower probability occurrence has happened.

[19] For other examples of such efforts, see Tegmark et al. (2006) and Weinberg (2000).

[20] It is merely an expectation but not an assurance, since a formal demonstration of the sort of behavior expected remains elusive.

If that truly is the basis of the proposal, then its basis does not warrant the assigning of probabilities. We have seen in Section 3.2.2 above that application of the principle of indifference leads to the non-probabilistic representation of completely neutral support.

Where this proposal would be falling short is that its multiverses are providing an ensemble but there is no analog to the randomizer needed to induce probabilities over the ensemble.

6.2 The Self-Sampling Assumption

A similar failing arises in connection with the "self-sampling assumption" of Bostrom (2002, Ch. 4, 5 and 9; 2002a; 2007). The analysis calls on the "Level I" of the multiverse hierarchy (Tegmark, 2007), in which a very large or spatially infinite universe realizes, to arbitrary closeness, a duplicate of any experiment we may undertake on earth; and does it very often and even infinitely often. When we perform an experiment, we might ask which of these many duplications is ours. Our background knowledge is quite neutral on the matter. So the appropriate representation is that of completely neutral evidence, as described in Section 3.2 above. That representation provides no basis for a probabilistic analysis.

One can force a probabilistic analysis onto the problem by stipulation. That is the effect of the self-sampling assumption. It enjoins us as follows (2007, p. 433):

> One should reason as if one were a random sample from the set of all observers in one's reference class.

Since the analysis is probabilistic, we assign equal probability to the outcome that each of the many replicas of the experiment is ours, a conclusion already suggested by the talk of a "random sample." Bostrom (2002a, p. 618) regards the assumption as "as kind of restricted indifference principle," which, I have already argued provides no basis for assigning probabilities. The assumption imposes a stronger probabilistic representation onto the problem than the weaker one warranted by the neutrality of the evidence, thereby risking again that conclusions are artifacts of a poorly chosen logic. One such artifact arises through the great difficulties probability measures face in spreading support uniformly over a countable infinity of outcomes.

The severity of the problem becomes clear if we work through an example that also shows that the analysis resulting from the self-sampling assumption is at best unnecessary and at worst unsuccessful. Take the proposition:

E: Penzias and Wilson measure a 3^oK cosmic background radiation.

We are interested in its probability conditioned on the unfavorable theory that:

T: The cosmic background radiation has a true temperature of 100^oK.

This particular probability is troubled by none of the above concerns with evidential neutrality; it is a physical chance that can be computed within a physical theory. If the cosmic background were 100^oK, then there would be some unimaginably small but non-zero probability "q" that, through a theoretically computable random fluctuation, Penzias and Wilson could measure 3^oK. That physical chance is expressed in the likelihood

(L) $P(3^oK \text{ measured in specific experiment} \mid \text{true background } 100^0K) = q$

The direct approach is simply to apply L to E, since E reports a specific experiment of the relevant type. That is, we recover directly

$$P(E|T) = q$$

Bostrom's analysis proceeds by assuming that this direct analysis cannot be applied. The worry of Bostrom (2002, Section 1) seems to be that an unlikely occurrence, such as measuring a 3^oK background in a universe with a true background of 100^oK, is virtually certain to happen somewhere if there are enough Penzias and Wilson clones performing the experiment. That is, the probability of an E-like event somewhere given the truth of T is close to one. The obvious point is that this reasoning conflates an E-like event with E. Nonetheless a probabilistic analysis is possible that allows for the possibility of many trials each able to produce an E-like event and then factors in that each is unlikely to be our trial. It is an indirect analysis conducted on a proposition logically equivalent to E:

F: There is a measurement of 3^oK cosmic background radiation somewhere in the universe AND it is our Penzias and Wilson's.

While the E and F are logically equivalent, a probabilistic analysis of them need not be equivalent. In the best case, the analysis of F yields the same conditional probability, but only by introducing dubious sampling probabilities. In the worst case, these extra probabilities incur sufficient problems to defeat the analysis entirely. In either case, there is no clear gain in replacing the analysis of E by that of F, for the second analysis still requires use of the likelihood (L) that can be applied directly to E to get the result without any further fuss.

We recover the same result analyzing E and F in the simpler case in which there are only a finite number n of replicas of the experiment in the multiverse.[21] We now apply the self-sampling assumption, on the basis that our Penzias and Wilson experiment is done in our world. We conclude that the probability that our Penzias and Wilson experiment is any one of the n available replicas is just 1/n. We compute the probability that our Penzias and Wilson do measure 3^{o}K by summing the probabilities of these n, mutually exclusive cases. That is, we compute

$$P(F|L) = q.(1/n) + q.(1/n) + \ldots + q.(1/n) = q$$

This is the same result as given by the direct analysis of P(E|L). However its generation is much less satisfactory. For the analysis of P(E|L) employed only the physical chances supplied by a physical theory. The analysis of P(F|L) arrived at the same result, but mixed the physical chance q with the dubious self-sampling probabilities 1/n. Indeed this last sum shows that something less than the self-sampling assumption is all that is needed. We could instead assume any set of probabilities $p_1, p_2, \ldots, p_n$ for the n trials. Since these n probabilities sum to unity, the above sum is replaced by

$$P(F|L) = q.\ p_1 + q.\ p_2 + \ldots + q.\ p_n = q$$

Thus, the indirect approach asks us to introduce an unnecessary assumption to generate dubious probabilities to conduct an unnecessary calculation to get the same result as was recovered directly without computation.

Or at least we have the same result in the case of a finite universe. If the universe is spatially infinite, so that there are infinitely many replicas of Penzias and Wilson, then the analysis fails. The self-sampling assumption entails that the probability that each particular replica is theirs is zero. It must be zero since each outcome of random sampling must have the same probability and any non-zero probability would lead to an infinite total probability. This

[21] While the analysis of the finite case does not require it, Bostrom assumes in cases like this that n is very large so that the probability that measurement of 3^{o}K occurs in at least one of the replicas is $1-(1-q)^n$, which can be brought arbitrarily close to one by sufficiently large n.

zero is a disaster for Bayesian analysis since all the probabilities produced by summing of the individual cases will themselves be zero. For example, P(F|L) is now given by[22]

$$P(F|L) = q.\,0 + q.\,0 + \ldots 0 = 0$$

which is the wrong result. Indeed the probability that any of the measurements is Penzias and Wilson's is a sum of infinitely many zeroes, which is just zero.

This last failure is simply the manifestation all over again of the inability of a probabilistic analysis to capture neutral support. That failure depends essentially on the additivity of probability measures. In the case of a countable infinity of mutually exclusive outcomes all of which are equally supported by the background, additivity requires that we assign a zero probability to each, on pain of violating normalization to unity.

At best, the indirect analysis that incorporates the self-sampling assumption must presume everything in the easy, direct analysis, and merely gives the same result. At worst, the indirect analysis fails and it does so because of the inability of a probability measure to represent neutral support.

7. When the Problem is Not Inductive

The material theory of induction asserts that the warrant for an inductive inference resides in locally obtaining facts. There is an inverse assertion. If there are no warranting facts, then there is no inductive inference warranted, even though the case may look eminently suited to inductive inference. Such a case is supplied by the problem of extendability in spacetime. Take any classical spacetime. One can always excise a piece as a purely mathematical construction to arrive at a new spacetime, observationally identical to the first. While a piece has been cut out,

---

[22] Should we compute P(F|L) by first computing P(F|L) for the finite case to recover q and then "take the limit as n goes to infinity." That procedure presumes that summing over all cases and taking the limit to infinity are operations that commute. In finite cases they do. In this infinite case they do not. A more promising approach is to explore the use of "improper" prior probabilities in which each of the countable infinity of outcomes is assigned the same, small non-zero probability. It is not clear that these improper probability distributions can be made to work in this case. If they can be, however, it will be at the expense of the self-sampling assumption. For no regime of random sampling allows probabilities whose sum is infinite.

there will be no observational trace of the excision. Everything in the surrounding spacetime will remain just as if the piece were still there. Even if our world tubes as observers were to pass through the hole, we would not notice anything unusual. We just cease to exist at the hole's past edge and we would be reconstituted at the future edge, complete with spurious memories of the non-existent, excised part, in exactly the state of someone who experienced everything in the excised hole.

It is natural to expect that some inductive inference grounded in our observations could somehow still force the conclusion that our spacetime is hole free. However (Norton, manuscript b) I have been unable to find any facts that warrant the relevant induction. We might call upon a common postulate in spacetime theories that the spacetime is inextendible. However that postulate is anomalous in being independent from a physical perspective from the other postulates. It is includes as much for mathematical convenience. The other justifications I have seen tend to the esthetic or oddly metaphysical. Somehow we must assume a Leibnizian principle of plenitude, so that the universe contains as much as is possible and no less. We are now descending into a fruitless battle of metaphysics. For this Leibnizian principle contradicts the metaphysics of parsimony encouraged by Ockham's razor. That tells us that we should presume the least possible. This degeneration surely tells us that we are no longer dealing with a problem that is properly understood as inductive in the first place.

Now return to the Surprising Analysis. I have challenged the Bayesian attempt to make it precise but not yet queried whether there is a good inductive inference in it at the informal level. It is now time to make that query. We judge the disposition of cosmic parameters to be surprising; we judge this or that theory not just to accommodate them but to explain them; so we infer to the theory. Yet we are unable to give a clear explication of why the values are surprising and just what it takes to be a good explanation.

Might we have to take these difficulties as an indication that no deeper explication is to be had? In expecting otherwise, are we being misled by other cases? We cannot fail to be impressed when we learn that a deeper theory forces some apparently arbitrary parameter to the value it has. Light, we learned from direct measurements, travels at 300,000 km/sec in a vacuum. We then found that, given the values of the basic constants in the theory of electricity and magnetism, it has to have that speed. Are we concluding too hastily that every parameter in our theories must have some similar, deeper accounting? While we may dream, as did Einstein, of a theory with no arbitrary parameters, that would not exhaust the arbitrariness of our theories. Why, for example, would our spacetime have as many dimensions as it has? Why might its spacetime metric be Lorentz signature and not (Newtonian) degenerate? Why should the structures in it be scalars or spinors or vectors or second rank tensors? Why should our spacetime or the inner spaces the quantum states of matter host just the symmetries they do? Why should quantum states evolve according to the Schroedinger equation? Why should gravity and spacetime curvature be connected?

If we demand a deeper explanation of everything, we trigger an unsatisfiable infinite regress. Perhaps some things in the cosmos are just as they are and no further accounting is possible. Perhaps the cosmic parameters of the Surprising Analysis just are what they are and no further explanation should be sought. Perhaps they are not the inductive springboards to multiverses and other exotic structures from which we imagine our world to be sampled. Perhaps they just are the way they are.

8. Conclusion

The attempt to use Bayesian analysis to supply a more precise reformulation of the Surprising Analysis fails. It founders on the inability of a probability measure to represent adequately the evidential neutrality with which the analysis begins. Once one grounds an analysis in a representation that conflates neutrality and disfavor, we should not readily accept its judgments of what is favored. The case of the doomsday argument shows how readily a faulty representation can return incorrect results. It assures us, on impossibly slender evidence, that we should believe the end of the universe is coming sooner. When we excise the probabilistic representation and replace it with a better representation of completely neutral support, the strong results evaporate. We have no grounds to expect the end sooner or later on its slender evidence.

These concerns do not always preclude a probabilistic analysis. It can be employed if there is a sufficient factual basis to warrant the probabilities. The mere supposition of a multiverse, however, is not sufficient. It does not provide probabilities until we add an analog of a randomizer, just as a deck of cards or die provide no probabilities until we introduce the randomizer of shuffling or throwing. The mere fact of evidential neutrality or ignorance over the members of the multiverse is insufficient grounds for the introduction of probabilities.

Inductive inference can give us new knowledge that extends beyond our evidence. But it can only do so by taking an inductive risk. The risk taken and the fragility of the results increase the more slender the evidence. In the extreme case of completely neutral evidence, we should no longer expect an inductive logic to wrestle new knowledge from nothing. This miraculous feat is what probabilistic analysis purports to do. Inductive logics that incorporate neutrality of support, when analyzing the same problems, return nothing. That is what we should expect.

Nihil ex nihil fit.[23]

Finally, the Surprizing Analysis is grounded in the idea that certain cosmic parameter values require a deeper explanation. While a call for explanation may initially seem natural, are we risking an unsatisfiable infinite regress if we continue call for further explanations of the contingent features of each explanation? The explanatory regress must terminate in some brute facts. Might we already have arrived at this terminus in certain aspects of our present cosmology? Might these cosmic parameters just be what they are with no further explanation possible?

References

Aguirre, Anthony (2007), "Making Predictions in a Multiverse: Conundrums, Dangers, Coincidences," pp. 367-86 in Carr (2007).

Bostrom, Nick (2002) *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. New York: Routledge.

Bostrom, Nick (2002a) "Self-Locating Belief in Big Worlds: Cosmology's Missing Link to Observation," *Journal of Philosophy*, **99**, pp. 607-623.

Bostrom, Nick (2007) "Observation Selection Theory and Cosmological Fine-Tuning" pp. 431-43 in Carr (2007).

Carr, Bernard, ed., (2007) *Universe or Multiverse*. Cambridge: Cambridge University Press.

Gibbons, G. W.; Hawkings, S. W. and Stewart, J. M. (1987) "A Natural Measure on the Set of All Universes," *Nuclear Physics*, B281, pp. 736-51.

Gibbons, G. W. and Turok, Neil (2008) "Measure Problem in Cosmology," *Physical Review,* **D77**, pp. 063516-1-12.

[23] "Nothing comes of nothing."

Hawking, S. W. and Page, Don N. (1988) “How Probable is Inflation?” *Nuclear Physics*, **B298**, pp. 789-809.

Howson, Colin and Urbach, Peter (2006) *Scientific Reasoning: The Bayesian Approach*. 3rd ed. Chicago and La Salle, Il: Open Court.

Jaynes, E. T. (2003), *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Norton, John D. (2003) "A Material Theory of Induction" *Philosophy of Science*, **70**(October 2003), pp. 647-70.

Norton, John D. (2005) "A Little Survey of Induction," in P. Achinstein, ed., *Scientific Evidence: Philosophical Theories and Applications*. Johns Hopkins University Press, 1905. pp. 9-34.

Norton, John D. (2007) "Probability Disassembled" *British Journal for the Philosophy of Science*, **58** (2007), pp. 141-171.

Norton, John D. (2007a) "Disbelief as the Dual of Belief." *International Studies in the Philosophy of Science*, **21**(2007), pp. 231-252.

Norton, John D. (2008) "Ignorance and Indifference." *Philosophy of Science*, **75** (2008), pp. 45-68.

Norton, John D. (manuscript) “Challenges to Bayesian Confirmation Theory,” Prepared for Prasanta S. Bandyopadhyay and Malcolm Forster (eds.), *Philosophy of Statistics: Vol. 7 Handbook of the Philosophy of Science*. Elsevier. http://www.pitt.edu/~jdnorton/

Norton, John D. (manuscript a) "Deductively Definable Logics of Induction." http://www.pitt.edu/~jdnorton/

Norton, John D. (manuscript b) "Observationally Indistinguishable Spacetimes: A Challenge for Any Inductivist." http://www.pitt.edu/~jdnorton/

Tegmark Max (2007), “The Multiverse Hierarchy,” pp. 99-125 in Carr (2007).

Tegmark, Max; Aguirre, Anthony; Rees, Martin J. and Wilczek, Frank (2006) “Dimensionless Constants, Cosmology, and Other Dark Matters,” *Physical Review*, **D73**, pp. 023505-1-28

Page, Don N. (2007) “Prediction and Tests of Multiverse Theories,” pp. 411-29 in Carr (2007).

Weinberg, Steven (2000) “*A Priori* Probabilities of the Cosmological Constant,” *Physical Review*. **D61**, pp. 103505-1-4.

Williamson, Jon (2009) “Philosophies of Probability” pp. 493-534 in, A. Irvine, ed., *Philosophy of Mathematics*.Amsterdam: North-Holland.