Cosmic Confusions: Not Supporting versus Supporting Not
Author(s): John D. Norton
Source: *Philosophy of Science*, Vol. 77, No. 4 (October 2010), pp. 501-523
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/10.1086/661504
Accessed: 10/06/2011 09:43

Cosmic Confusions: Not Supporting versus Supporting Not*

John D. Norton†‡

Bayesian probabilistic explication of inductive inference conflates neutrality of supporting evidence for some hypothesis H ("not supporting H") with disfavoring evidence ("supporting not-H"). This expressive inadequacy leads to spurious results that are artifacts of a poor choice of inductive logic. I illustrate how such artifacts have arisen in simple inductive inferences in cosmology. In the inductive disjunctive fallacy, neutral support for many possibilities is spuriously converted into strong support for their disjunction. The Bayesian "doomsday argument" is shown to rely entirely on a similar artifact.

1. Introduction. One cannot have any doubt of the many successes of the Bayesian project of explicating inductive inferences. Its successes have been widely and justly celebrated. What has received less attention are the limits of these successes.[1] The purpose of this article is to describe one circumstance in which Bayesian analysis fails. This is the extreme case of complete neutrality of evidential support. The Bayesian system is unable to distinguish it cleanly from strongly disfavoring evidence. The system tries to represent this complete neutrality with a broadly spread probability measure that ends up assigning a very low probability to each possibility. The trouble is that this same very low value of probability is correctly used when that same possibility is strongly disfavored by the evidence or, equivalently, its negation is strongly favored. In short, for a

*Received January 2010; revised March 2010.

†To contact the author, please write to: Departments of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260; e-mail: jdnorton@pitt.edu.

‡For helpful discussion, I thank Jeremy Butterfield, Eric Hatleback, Wayne Myrvold, and participants at the conference "Philosophy of Cosmology: Characterising Science and Beyond" at St. Anne's College, Oxford, September 20–22, 2009.

1. For other limits, see Norton ("Challenges to Bayesian Confirmation Theory," forthcoming).

Philosophy of Science, 77 (October 2010) pp. 501–523. 0031-8248/2010/7704-0005$10.00

hypothesis H, a Bayesian analysis conflates the cases of evidence *not supporting* H with evidence *supporting not*-H.

If one insists that probabilistic notions should be used in cases of evidential neutrality, one ends up assigning neutral support to the formal properties of evidential disfavoring. Since evidential neutrality warrants fewer definite conclusions than does evidential disfavor, this conflation leads to spurious conclusions that are merely artifacts of a poor choice of inductive logic.

My contention in this article is that this conflation of neutral and disfavoring evidence has occurred repeatedly in philosophical and physical analyses in cosmology. Since cosmology often deals with problems of universal scope for which evidence is meager, it is rich in cases of neutral support and thus especially prone to the confusion. My purpose in this article is to elaborate the difference between neutral and disfavoring evidence, to show how nonprobabilistic formal tools may be used to represent completely neutral evidential support and to give examples of the conflation of neutral and disfavoring evidence in cosmology.

Section 2 will develop a simple example of neutral evidential support in cosmology in order to fix the notion more clearly. Section 3 will investigate how this neutrality can be represented formally. It will be argued that a probability measure represents degrees of favoring and disfavoring but does not capture neutrality. Rather, an inherently nonadditive representation must be used for completely neutral support. Section 4 will show that misdescription of neutrality of support by a probability measure leads to the "inductive disjunctive fallacy" in which disjunctions of neutrally supported possibilities are mistakenly judged as strongly supported. Illustrations in the literature include van Inwagen's argument for why there is very probably something rather than nothing. Section 5 will sketch how the nonadditive representation of completely neutral support can be incorporated into an alternative inductive logic.

Section 6 will show that 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. Section 7 will review how probabilistic representations can properly be introduced into cosmology. An ensemble provided by a multiverse is not enough. What is needed are some facts that specifically warrant probabilities. The difficulties of the "self-sampling assumption" arise because there are no such facts. Finally, the concluding section 8 will suggest that mainstream cosmological theorizing is at risk of committing the same fallacies as sketched in earlier sections.

2. A Cosmological Case of Neutral Support. A clear example of complete neutrality of support in cosmology arises in a more extreme version of multiverse theory. There we may postulate other universes, disconnected from ours, in which the same fundamental laws of physics obtain. In these other universes, the fundamental constants like h, c, G, and the parameters of the standard model of particle physics have different values, but our supposition is that we have no indication at all of what those values might be. Even so, we can still know a lot about these other universes. Except in degenerate cases, they will emit wavelike propagations of electromagnetic radiation. If the various fundamental forces are appropriately balanced, they will harbor chemical elements like our own, with characteristic quantized atomic spectra. But what can we say of the values of fundamental constants themselves? Our evidence tells us nothing. We have no reason at all to favor one set of values of Planck's constant h over any other. The evidence is neutral.[2]

This case is to be distinguished from another multiverse theory in which we have disfavoring evidence for the same parameter. In this other multiverse theory, new universes are born from singularities through stochastic processes whose governing law, we shall suppose, provides a broadly spread probability distribution over the possible values of h. In this case, that h lies in any small interval of values is very improbable; our background evidence disfavors that small interval. Correspondingly, we have strong evidence that the actualized value of h lies outside this interval.

In the first multiverse theory, we simply have no support for the value of h to be in or not to be in some particular small interval of values. In the second, it is improbable that h lies in some small interval and probable that it lies outside it.

We should not conflate the two cases. Should we try to represent the neutrality of the first theory by assigning a low probability to h lying in the interval, we have contradicted that neutrality, for that assignment forces a high probability on h lying outside the interval, an outcome for which we must now assign strong support. That high probability and the resulting near certainty is a spurious artifact of the use of the wrong inductive logic. It is the support that would arise in the second theory in which the evidence disfavors strongly the small interval and thus strongly favors values outside that interval.

2. Comparing fundamental constants across universes requires that we also compare the units of measurement used. Readers who wish to avoid these complications should replicate the arguments of this article using dimensionless quantities, such as the fine structure constant.

3. Representing Neutral Evidential Support.

3.1. The Failure of a Probabilistic Representation. If a probability measure is able to represent degrees of evidential support at all, then a probability $P(\mathrm{H}|\mathrm{E})$ near unity must represent the case of evidence E providing strong support for the hypothesis H. It immediately follows from the additivity of probability measures

$$P(\mathrm{H}|\mathrm{E}) + P(\text{not-H}|\mathrm{E}) = 1$$

that $P(\text{not-H}|\mathrm{E})$ is close to zero. Since E favors H just to the extent that it disfavors the negation not-H, we must now conclude that, when $P(\text{not-H}|\mathrm{E})$ is close to zero, evidence E strongly disfavors not-H. Reversing H and not-H, we can now conclude that, when $P(\mathrm{H}|\mathrm{E})$ is close to zero, evidence E strongly disfavors H. More generally, if there are n mutually exclusive and exhaustive outcomes $\mathrm{A}_1, \ldots, \mathrm{A}_n$, additivity requires

$$P(\mathrm{A}_1|\mathrm{B}) + P(\mathrm{A}_2|\mathrm{B}) + \ldots + P(\mathrm{A}_n|\mathrm{B}) = 1$$

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

The additivity of probabilities is the mathematical expression of the complementary relationship of support and disfavoring.[3] It leaves no place in the representation for neutrality. The standard device of representing neutrality with a broadly spread probability distribution merely assigns a very low probability to each possible outcome, that is, the case of evidential disfavor, not neutrality.[4]

3.2. Representing Evidential Neutrality. How are we to represent evidential neutrality? The difficulty for the most general case 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 multidimensional 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. Write $[\mathrm{A}|\mathrm{B}]$ as the inductive support proposition A is accorded by proposition

3. Conversely, it has been argued (Norton 2007b, sec. 4.1) that the presumption that the range of values of the degrees of support span favoring to disfavoring leads us directly to an additive measure.

4. What of the popular device of representing neutrality by sets of probability measures? It has been argued in Norton (2007a, sec. 6; 2007b, sec. 4.2) that this device fails for several reasons. The most serious is that it is an attempt to simulate an inherently nonadditive logic with an additive measure, rather than to seek the logic directly.

B. The use of a new notation reminds us that these degrees of support need not be probabilities. Let us take the case of *complete* evidential neutrality. This extreme case can be captured by an essentially nonadditive representation. The support accorded any contingent proposition A by the background B is just one fixed value that we write "I" (for indifference or ignorance) that figures in the following distribution:

(CNS) Completely neutral support
$[T|B] = 1$, for all propositions T deductively entailed by B;
$[A|B] = I$, for all contingent propositions A;
$[F|B] = 0$, for all propositions F that logically contradict B.

The 1 and 0 of the two extreme cases are less interesting; this is merely the assigning of extreme values to propositions we know deductively to be true or false given B. The interesting part is that all contingent propositions, whose truth values are left undecided by B, are accorded the same neutral value I.

The quantity $[A|B]$ of CNS should not be confused with other quantities that arise in Bayesian analyses.[5] This quantity expresses the total support accorded to an outcome A by the background B. It is a function of two propositions, A and B, only. It is distinct from a relation of differential or incremental support: the support accorded A specifically by evidence E in the context of background B. This is a tertiary function of three propositions, A, E, and B. In a Bayesian analysis, it is measured by comparing the posteriors and priors, $P(A|E\&B)$ and $P(A|B)$, such as through a difference measure $P(A|E\&B) - P(A|B)$, and the analysis may seek to express neutrality through the probabilistic independence of E and A when they are conditioned on the background B. My concern is not the differential evidential import of E but that a probabilistic prior such as $P(A|B)$ must fail to capture total neutrality of support.

That CNS is the appropriate representation of completely neutral support has been argued at length by Norton (2008).[6] I refer readers to it for a formally precise development. In the discussion below, I shall indicate informally how the result comes about. It comes from two independent invariance conditions, each of which yields the same outcome.

5. I am grateful to Jonah Schupbach for raising this issue.

6. In Norton (2008), I describe neutral support as an "ignorance" distribution. In using that description, regrettably I succumbed to the subjective Bayesian's insistence that inductive logic is really about degrees of belief, whereas I now think we must insist that it is about objective degrees of support, as do objective Bayesians. The intrusion of opinion must be resisted since it corrupts evidential relations of support and obscures the limits of applicability of Bayesianism.

3.2.1. *Invariance under Redescription* (*and the Principle of Indifference*). The principle of indifference asserts that, if the evidence bears equally on two outcomes, then the support accorded each by the evidence should be the same. This principle is so weak as to border on truism. It does have some strong consequences, however, if we allow that indifference and the resulting equality of evidential import persists when the outcome space is redescribed. The invariance of this indifference leads directly to CNS.

Take the multiverse example of section 2. The evidence is completely neutral over different values of h. As a result, it supports equally that h is in each of the intervals $0 < h \leq 1$, $1 < h \leq 2$, $2 < h \leq 3$, $3 < h \leq 4$, and so on. The neutrality of support persists if we consider the quantity h^2, and, by the same reasoning, the evidence supports equally that h^2 is in each of the intervals $0 < h^2 \leq 1$, $1 < h^2 \leq 2$, and so on. But this is equivalent to asserting equal support for the intervals $0 < h \leq 1$ and $1 < h \leq 4$. Combining the two cases, we have equal support for the intervals $1 < h \leq 2$ and $1 < h \leq 4$, even though the first is a proper part of the second. We can write

$$[(1 < h \leq 2)\,|B] = [(2 < h \leq 3)\,|B] = [(3 < h \leq 4)\,|B] = [(1 < h \leq 4)\,|B].$$

By continuing this sort of argumentation with different rescalings of h, we arrive at equal support for all nonempty, proper subintervals of $0 < h < \infty$. We can infer equality of support for finite and infinite intervals by rescaling from h to 1/h.[7] Thus, all nonempty, proper subintervals of $0 < h < \infty$ must be assigned the same support I,

$$[\text{any subinterval}|B] = I,$$

which is the expression of CNS for a continuous parameter.

3.2.2. *Invariance under Negation.* The same result is recovered from a different invariance to which completely neutral support conforms. That is invariance under negation. What motivates it is that completely neutral evidence cannot offer differential support to a contingent proposition and to its negation, so that

$$[A|B] = [\text{not-}A|B] \qquad \text{for all contingent propositions A.}$$

This invariance condition is very strong. If we couple it with a condition of monotonicity, it is easy to show that the only admissible set of degrees assigns the same value to all contingent propositions.

As an illustration, again take the case of completely neutral support described in section 2 above. If the proposition A locates h in the interval

7. Then we have equal support for $0 < 1/h \leq 1$ and $1 < 1/h \leq 2$. But that corresponds to equal support for $\infty > h \geq 1$ and $1 > h \geq 1/2$.

$0 < h \leq 1$, then its negation, not-A, locates h in the complementary interval $1 < h \leq \infty$. Applying negation invariance to this and similar cases we have

$$[(0 < h \leq 1)\,|B] = [(1 < h \leq \infty)\,|B],$$

$$[(0 < h \leq 2)\,|B] = [(2 < h \leq \infty)\,|B].$$

The requirement of monotonicity asserts that no interval, such as $0 < h \leq 2$, can accrue less support than one of its proper parts, such as $0 < h \leq 1$.[8] Thus, we have

$$[(0 < h \leq 1)\,|B] \leq [(0 < h \leq 2)\,|B],$$

$$[(1 < h \leq \infty)\,|B] \geq [(2 < h \leq \infty)\,|B].$$

It follows that all four strengths are equal. Combining and extending this analysis to other intervals, we arrive as before at the distribution CNS for a continuous parameter,

$$[\text{any interval}|B] = I.$$

3.3. Neutrality and Disfavor versus Ignorance and Disbelief. Those familiar with the literature will find the last discussion nonstandard. 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 a proposition inductively supports another. 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

8. More generally, monotonicity requires that, when A deductively entails C, $[A|B]$ is not greater than $[C|B]$. This merely requires that the deductive consequences of a proposition are at least as well supported as the proposition (see Norton 2008, secs. 6.2–6.3).

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 completely neutral 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 $k = k_1$, $k = k_2$, $k = k_3$, and so on. We have no idea which is the correct value, so, as a subjective Bayesians, we would 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.[9] Let us say that, with the decrease needed, we assign the following two prior probabilities

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

Now we begin collecting evidence. We learn, say, that k_n has $n < 1{,}000$ and then that k_n has $n < 500$ and then that k_n has $100 < n < 200$. All the while, we conditionalize on this new evidence and the probabilities of the remaining k_n mount. Finally, we acquire evidence $E = k_{135} \vee k_{136}$. This is the most specific evidence possible. The next stage of evidence collection would merely declare which of k_{135} or k_{136} is the correct one. At this last stage of conditionalization, Bayes's theorem in the ratio form assures us that

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

9. A countable infinity of outcomes can be accorded equal prior probabilities if the prior is "improper," that is, it does not sum to unity. However, this uniformity will not be preserved under redescription of the outcomes. The posterior might remain improper if the evidence merely reduces the possibilities to a smaller infinite set, such as all k_n with even n. The need to abandon one of the most important axioms of the probability calculus is a direct admission that probabilities are the wrong representation for the problem.

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 from this last computation that our strong preference for k_{135} is entirely an artifact of the pure opinion encoded in the ratio of the priors. Far from being washed out, the priors have risen to dominate the outcome entirely.

4. Neutrality of Support and the Inductive Disjunction Fallacy. The mistake of conflating evidential disfavor with evidential neutrality leads us directly to an inductive fallacy that I will call the "inductive disjunctive fallacy." To see how it arises, recall that completely neutral support accords the same neutral degree of support to all contingent propositions. Take the case of an outcome space with contingent, mutually exclusive outcomes a_1, a_2, a_3, and so on. Taking disjunctions—"or-ing" these outcomes together—does not lead us to propositions with any greater support:

$$I = [a_1|B] = [a_1 \vee a_2|B] = [a_1 \vee a_2 \vee a_3|B]$$
$$= [a_1 \vee a_2 \vee a_3 \vee a_4|B] = \ldots .$$

If, however, we seek to represent the neutrality of support with a broadly spread probability distribution so that

$$P(a_n|B) = \text{some small value} \qquad \text{for all } n,$$

then taking disjunctions will generate propositions with increasing probability:

$$P(a_1|B) < P(a_1 \vee a_2|B) < P(a_1 \vee a_2 \vee a_3|B)$$
$$< P(a_1 \vee a_2 \vee a_3 \vee a_4|B) < \ldots .$$

Eventually, if we accumulate enough disjuncts, we can bring our probability close to unity, which we must interpret as strong support.

If we mistakenly believe that our probabilistic representation expresses neutrality of support for each proposition a_n, then we can commit the inductive disjunctive fallacy: simple arithmetic will lead us to infer incorrectly that taking a big enough disjunction of neutrally supported propositions gives us a proposition that is strongly supported.

The fallacy can be illustrated with the example of completely neutral support of section 2. The different values of the parameter h have completely neutral support. So the propositions that h lies in $0 < h \leq 1$ or $1 < h \leq 2$ or $2 < h \leq 3$, and so on, are each equally supported. Setting aside difficulties of normalization, we might then represent these as equal probability cases. It then follows that virtually all the probability mass must be assigned to the complement of the interval $0 < h \leq 1$; that is, $1 < h \leq \infty$. So we infer spuriously that a value of h not in $0 < h \leq 1$ is strongly supported.

One might imagine that this fallacy is too transparent to be committed in serious work. It turns out, however, that it is committed routinely in cosmology. Here are a few examples. Van Inwagen, while not himself a cosmologist, addresses a cosmological question. He proposes to answer the question that is "supposed to be the most profound and difficult of all questions" (1996, 95): "Why is there anything at all?" The argument is elaborate, so I shall jump to the essential step. Van Inwagen presents the premises that there is only one possible world in which there are no beings but there are infinitely many possible worlds in which there are beings. The latter is arrived at by arguing that there are many ways for beings to be but only one way for them not to be. He then urges that the probability of being actual for each possible universe is the same. (I set aside the problem that this instantly conflicts with the requirement that probability measures normalize to unity.) It now follows that the probability "of there being nothing is 0." It is "as improbable as anything can be" (99). Hence, no doubt, we are to infer that there being anything at all is as probable as anything can be.

Van Inwagen prudently admits that he is "unhappy about the argument. . . . No doubt there is something wrong with it . . . but I should like to be told what it is" (1996, 99). What is wrong is that it is an instance of the inductive disjunctive fallacy. Our background assumptions are near vacuous and provide completely neutral support for the actuality of each possible world; therefore, they provide completely neutral support for any disjunction of these possibilities. What van Inwagen has done is to represent this neutrality incorrectly by a widely spread probability measure, thereby committing himself fallaciously to the conclusion that a disjunction of all but one of them is strongly supported.

The same fallacy is committed by Olum (2004) as part of a challenge to anthropic reasoning. He notes that, in an infinite universe, civilizations can grow to very great size as measured by the number of inhabitants and spatial extent. Most would be much larger than our own young civilization, limited to planet Earth. Since "anthropic reasoning predicts that we are typical," we could be any of the individuals in any of these civilizations. Since, overwhelmingly, most individuals will belong to large civilizations, it follows that anthropic reasoning "predicts with great confidence that we belong to a large civilization" (2). Olum's point is that the fact that we do not belong to such a civilization refutes the conclusion of anthropic reasoning, thereby suggesting this form of reasoning is defective.

My concern here is only to point out that Olum's argument depends essentially on the inductive disjunctive fallacy. Our background evidence is neutral over which individual each of us may be. On Olum's assumption, there are vastly more individuals in large civilizations. Take the disjunctive proposition that you are one of the individuals in a large civilization. Under the correct treatment of neutrality of support, that disjunctive proposition accrues no more support than the proposition that you are any particular individual. Yet Olum infers that the taking of the disjunction has transformed neutral support into "great confidence."

A variant and rather more complicated form of Olum's argument had already been given by Bostrom (2003). Bostrom considers not just the growth of civilizations but that suitably advanced civilizations will have the ability to simulate conscious minds in computers. He considers the case in which the development of the civilizations is such that they are capable of simulating vastly many consciousnesses and they do so. In that case, in an argument similar to Olum's, Bostrom concludes that "we are almost certainly living in a computer simulation" (243). It is the same fallacy.

Another instance of the fallacy has circulated informally, although I have not found it in print.[10] Among multiverses, some will be spatially infinite and others spatially finite. The spatially infinite ones will have infinitely many more observers in them than the spatially finite ones. Since anthropic reasoning allows that we could be any of these observers, it is overwhelmingly likely that we are the observers in a spatially infinite universe. Therefore, our space is overwhelmingly likely to be infinite.

5. Inductive Logics That Tolerate Neutrality of Support. The inductive disjunctive fallacy has depended only on correcting the representation of completely neutral support. A fuller inductive logic will allow this representation to be the starting point of further inductive explorations, much as the Bayesian system uses prior probabilities as its starting point. Since we have no good characterization of the multidimensional space of degrees that accommodates both disfavoring and neutrality of evidence, we have no complete inductive logic that accommodates both.[11]

However, it is possible to discern how such a logic might deal with conditionalization that proceeds from the initial state of completely neutral support. We weaken the Bayesian system so that it comes to tolerate

10. It was described by John Barrow as a test of one's commitment to anthropic reasoning in discussion at the conference "Philosophy of Cosmology: Characterising Science and Beyond" at St. Anne's College, Oxford, September 20–22, 2009.

11. Norton ("Deductively Definable Logics of Induction," forthcoming) tries to survey the terrain of possible logics. It includes a sample "partial neutrality inductive logic."

completely neutral support. We do that by discarding the requirement of additivity of degrees of support. As discussed by Norton (2007b), that additivity is independent of the dynamics of conditionalization encoded in Bayes's theorem. We preserve one consequence of that theorem. For propositions T_1 and T_2, evidence E, and background B, Bayes's theorem in the ratio form asserts

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

where the second equality holds only in the special case in which each of T_1 and T_2 entail the evidence E. 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 $P(T_1|E\&B) = P(T_2|E\&B)$.

We now posit this result independently for a more general inductive logic:[12]

> 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 as knowledge. We replace the probabilistic prior by a neutral prior:

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

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 recover the result at which we should have arrived 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.

6. The Doomsday Argument. A further illustration of these alternative logics can be found in a reanalysis of the doomsday argument. See Bostrom (2002a, chaps. 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. Reanalysis that employs a more

12. This posit is not inevitable but just the simplest. For cases in which it fails, see Norton (2007b, sec. 5) and the "specific conditioning logic" in Norton ("Deductively Definable Logics of Induction," forthcoming, sec. 11.2).

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. There are, of course, many versions of the doomsday argument. My goal here is not to address them all but to show in an example how its result is entirely an artifact of the wrong choice of inductive logic.

6.1. 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 when we find that the process has already persisted for t years? We assign a prior probability density $p(\mathrm{T}\,|\,\mathrm{B})$ to T and a likelihood to our learning that the process has persisted t years:

$$p(\mathrm{t}|\mathrm{T\&B}) = \frac{1}{\mathrm{T}}.$$

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, so our prior is a uniform probability density. We now apply Bayes's theorem in the ratio form for two different values of T, both greater than t:

$$\frac{p(\mathrm{T_1}|\mathrm{t\&B})}{p(\mathrm{T_2}|\mathrm{t\&B})} = \frac{p(\mathrm{t}|\mathrm{T_1\&B})}{p(\mathrm{t}|\mathrm{T_2\&B})} \times \frac{p(\mathrm{T_1}|\mathrm{B})}{p(\mathrm{T_2}|\mathrm{B})} = \frac{\mathrm{T_2}}{\mathrm{T_1}} \times \frac{p(\mathrm{T_1}|\mathrm{B})}{p(\mathrm{T_2}|\mathrm{B})}.$$

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

$$p(\mathrm{T}|\mathrm{t\&B}) \propto \frac{1}{\mathrm{T}}.$$

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 $p(\mathrm{t}\,|\,\mathrm{T\&B}) = 1/\mathrm{T}$. It amounts to saying that 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

$$\mathrm{T}' = n(\mathrm{T}) \qquad \mathrm{t}' = n(\mathrm{t}),$$

where the function $n(\cdot)$ gives the number of people born prior to the time indicated. The new analysis uses a likelihood

$$p(\mathrm{t}'|\mathrm{T}'\&\mathrm{B}) = \frac{1}{\mathrm{T}'}.$$

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 $p(\mathrm{t}\,|\,\mathrm{T}\&\mathrm{B}) = 1/\mathrm{T}$, which varies according to T. The idea the likelihood was trying to express was merely that, even with a T chosen, no value of t in the admissible range from t = 0 to t = T is preferred; our evidence is completely neutral. That uniformity could be expressed by merely setting $p(\mathrm{t}\,|\,\mathrm{T}\&\mathrm{B})$ to a constant. The additivity of probabilities, however, requires that all probability densities integrate to unity. As a result, that constant must vary with different values of T as 1/T so that

$$\int_{\mathrm{t}=0}^{\mathrm{T}} p(\mathrm{t}|\mathrm{T}\&\mathrm{B})\,d\mathrm{t} = \int_{\mathrm{t}=0}^{\mathrm{T}} \mathrm{constant}\,d\mathrm{t} = \mathrm{constant} \times \mathrm{T} = 1.$$

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. That is, the result depends on using the wrong representation for evidential neutrality.

6.2. The Barest Reanalysis. This illusion that we get something for nothing starts to evaporate once we reanalyze 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 start with completely neutral support:

$$[\mathrm{T}_1|\mathrm{B}] = [\mathrm{T}_2|\mathrm{B}] = \mathrm{I}.$$

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 from section 4 and infer that

$$[\mathrm{T}_1|\mathrm{E}\&\mathrm{B}] = [\mathrm{T}_2|\mathrm{E}\&\mathrm{B}] = \mathrm{I}.$$

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 specific t is observed? A return to the Bayesian analysis will show us a way to proceed.

6.3. The Bayesian Analysis Again. The Bayesian analysis of section 6.1 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's theorem:

$$p(\mathrm{T}|\mathrm{t}\&\mathrm{B}) = p(\mathrm{t}|\mathrm{T}\&\mathrm{B}) \times \frac{p(\mathrm{T}|\mathrm{B})}{p(\mathrm{t}|\mathrm{B})} = \frac{1}{\mathrm{T}} \times \frac{p(\mathrm{T}|\mathrm{B})}{p(\mathrm{t}|\mathrm{B})},$$

for $\mathrm{T} > \mathrm{t}$. What seems unknowable is the ratio of priors $p(\mathrm{T}|\mathrm{B})/p(\mathrm{t}|\mathrm{B})$. It turns out, however, 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 is 400–500 years. With this timescale 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 timescale 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 $p(\cdot \mid \cdot)$ 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 $p(\cdot \mid \cdot)$ is used whichever unit is used to measure time. This entails that the probability density $p(\cdot \mid \cdot)$ is invariant under a linear rescaling of the times t and T (that, e.g., corresponds to changing measurements in years to measurements in days):

$$\mathrm{t}' = \mathrm{At} \qquad \mathrm{T}' = \mathrm{AT}.$$

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

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

where $C(\mathrm{t})$ 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:

$$\int_{\mathrm{T}=t}^{\infty} p(\mathrm{T}|\mathrm{t})\, d\mathrm{T} = \int_{\mathrm{T}=t}^{\infty} \frac{C(\mathrm{t})}{\mathrm{T}}\, d\mathrm{T} = \infty.$$

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 from T_1 to T_2:

$$P(\mathrm{T}_1 < \mathrm{T} < \mathrm{T}_2) = \int_{\mathrm{T}_1}^{\mathrm{T}_2} p(\mathrm{T}|\mathrm{t}\&\mathrm{B})\, d\mathrm{T} = \int_{\mathrm{T}_1}^{\mathrm{T}_2} \frac{C(\mathrm{t})}{\mathrm{T}} d\mathrm{T} = \text{finite}.$$

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

$$P(\mathrm{T} > \mathrm{T}_2) = \int_{\mathrm{T}_2}^{\infty} p(\mathrm{T}|\mathrm{t}\&\mathrm{B})\, d\mathrm{T} = \int_{\mathrm{T}_2}^{\infty} \frac{C(\mathrm{t})}{\mathrm{T}} d\mathrm{T} = \infty.$$

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

13. See, e.g., Jaynes (2003, 382). The probability assigned to the small interval $d\mathrm{T}$ must be unchanged when we change units. That is, $p(\mathrm{T}\,|\,\mathrm{t}\&\mathrm{B})\, d\mathrm{T} = p(\mathrm{T}'\,|\,\mathrm{t}'\&\mathrm{B})\, d\mathrm{T}'$. Since $\mathrm{T}' = \mathrm{AT}$, we have $d\mathrm{T}'/d\mathrm{T} = \mathrm{A} = \mathrm{T}'/\mathrm{T}$, so that $p(\mathrm{T}\,|\,\mathrm{t}\&\mathrm{B}) \times \mathrm{T} = p(\mathrm{T}'\,|\,\mathrm{t}'\&\mathrm{B}) \times \mathrm{T}'$, from which the Jeffreys prior follows immediately.

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.

6.4. 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 used in a non-Bayesian analysis. We seek the degree of support $[T_1, T_2 | t]$ for an end time in the interval from T_1 to T_2 given by the observation that the process has progressed to time t. We assume both T_1 and T_2 are greater than t.

The Bayesian analysis of section 6.1 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 timescale. 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 our 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[14]

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

14. 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 $[\cdot, \cdot \mid \cdot]'$ indicates that we are using the rule for computing degrees of support pertinent to the rescaled clock. The invariance, however, tells us

This will still be the case if either interval is 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 oracular powers of prognostication. On that evidence, we have no reason to prefer any finite time interval in the future over any other.[15]

7. Bringing Back Probabilities. There are many cases in which a probabilistic logic is the right one. To know which they are, we need to find a grounding in the facts of the particular case for probabilities of the logic.[16] The simplest case arises when the system is a stochastic one governed by physical chances, such as the decay of a radioactive atom. Then it is natural to conform strengths of support to the chances, for then strengths of support will agree with frequencies of success. A widely applicable example occurs if we assume that the errors entering into the measurement of a quantity arise in a pseudorandom manner. If they are small, independent, and summed in accord with the antecedent conditions of the central limit theorem of probability theory, their pseudorandomness warrants the use of a probabilistic bell curve to model the variations in the measured quantity.[17]

7.1. A Mere Ensemble Is Not Enough. 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.[18] This is the right way to proceed, although there is always scope for the facts invoked

---

that both original and rescaled systems use the same rule, so that the two functions $[\cdot, \cdot \mid \cdot]$ and $[\cdot, \cdot \mid \cdot]'$ are the same. Hence, $[T_1, T_2|t] = [T_3, T_4|t]$ as claimed.

15. 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 that invariance conditions are able to force it.

16. The material theory of induction (Norton 2003, 2005) is an extension of this idea. It asserts that the warrant for an inductive inference is not a universal formal template but a locally obtaining matter of fact.

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 bookie arguments obtain (see Howson and Urbach 2006, chap. 3). These facts warrant one conforming one's inductive reasoning with the probability calculus—but only as long as these facts obtain.

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

to fall short of what is needed. An ensemble is like a deck of cards. We do not have a probability of 1/52 for the ace of hearts when we merely have a deck of cards. We must, in addition, shuffle it and deal a card. Without this randomizer, merely having neutral evidential support for all cards is insufficient to induce the probabilities.

The proposal developed in Gibbons, Hawkings, and Stewart (1987), Hawking and Page (1988), and Gibbons and Turok (2008) supplies an ensemble but no analog of the randomizer. It 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. There, 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.[19] That means that the system will spend roughly equal time 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. (1987) proposal, however, such ergodic-like behavior is not expected. Over time, a single model will not explore a fuller part of the model space of all possible cosmologies. Rather, the proposal is justified by the following remark: "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" (736). 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 above that application of the principle of indifference may lead to the nonprobabilistic representation of completely neutral support.

7.2. The Self-Sampling Assumption. A similar failing arises in connection with the self-sampling assumption of Bostrom (2002a, chaps. 4, 5, and 9; 2002b; 2007). A very large or spatially infinite universe may harbor many observers, and, before consideration of further evidence specific to our circumstances, we might ask which of these many we are. The background evidence considered 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 impose a probabilistic analysis on the problem by stipulation. That is the effect of the self-sampling assumption. It enjoins

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

us as follows: “One should reason as if one were a random sample from the set of all observers in one’s reference class” (Bostrom 2007, 433). Since sampling is probabilistic, we assign equal probability to the outcome that we are each one of the many observers.

Bostrom stresses that these probabilities are not an adaptation of strengths of support to physical chances. “I am not suggesting that there is a physical randomization process, a cosmic fortune wheel as it were, that assigns souls to bodies in a stochastic manner. Rather we should think of these probabilities as *epistemic*” (2002b, 618; see also Bostrom 2002a, 57, for similar remarks). However, if the probabilities are epistemic and thus implement an inductive logic, what grounds do we have for that logic being probabilistic? Bostrom continues to explain that he regards the assumption “as kind of restricted indifference principle” (2002b, 618).

The principle of indifference, however, does not automatically warrant probabilities but only equalities of inductive strength. As we saw in section 3.2 above, if the indifference is extensive enough, the principle can directly preclude a probabilistic logic. Such preclusion arises when indifference persists over redescriptions. This proves to be a problem for the self-sampling assumption. In forming our sampling distribution, should we be indifferent over all people? over individual minutes experienced by people? over groups of people? or over civilizations? Each choice gives a different probability measure. Bostrom (2002a, 69–72) has identified this problem as the “reference class problem” and attempts a solution in subsequent chapters (chaps. 10–11). The attempt depends on the assumption that there is a single correct reference class to be chosen and that poor choices can be eliminated by showing that they have undesirable consequences in a probabilistic analysis. Since the relevant evidence is sufficiently weak to allow indifference to persist over multiple descriptions, assumptions of both a unique correct reference class and the applicability of probabilistic reasoning are in error.

Finally, Bostrom (2002a, 51–58; 2002b; 2007, sec. 24.2) urges that we must employ the self-sampling assumption to save Bayesian analysis of evidence from the following problem. A standard result of Bayesianism is that a good theory is rewarded epistemically for saying that the observed outcome of an experiment is very probable, whereas as a poor theory is punished when it says that the outcome is improbable. Now, a poor theory can still allow the observed outcome to occur as a highly improbable fluctuation, so that its occurrence somewhere in a very big universe is all but assured. As a result, Bostrom believes, we cannot use our observation of the experimental outcome to reward the good theory and punish the poor one in an unsupplemented Bayesian analysis. Both theories allow the observation with high probability. We must invoke the self-sampling assumption to discount the high probability from the poor theory.

If Bostrom is right that the Bayesian analysis has to be saved by an incorrect representation of the inductive import of the evidence, then that seems good reason not to use a Bayesian analysis. The inductive import of the experiments does not have to be explicated by a Bayesian analysis but only by an inductive logic that is properly adapted to the case at hand. Sometimes, as we saw above for the doomsday argument, a nonprobabilistic inductive logic is called for. In this case, however, I do not believe that the problem Bostrom outlines is a challenge to Bayesian analysis.[20] In sum, the self-sampling assumption imposes a stronger probabilistic representation onto the problem than the weaker one warranted by the neutrality of the evidence, thereby risking that conclusions are artifacts of a poorly chosen logic.[21]

8. Conclusion. What the above analysis shows is that there are limits to Bayesian analysis. It is unable to separate neutral evidential support from disfavoring support. If we confuse the two by using a probability measure to represent neutral evidential support, we introduce artifacts into our results that merely reflect the poor choice of inductive logic. These artifacts were illustrated in cosmology in the cases of the inductive disjunctive fallacy and the doomsday argument. These examples are simplified and removed from mainstream cosmological theorizing. They were chosen for analysis precisely because of this simplicity. It gave us enough independent perspective to be able to untangle the faulty reasoning.

Are these same problems a concern in mainstream cosmological theorizing? It takes only a cursory review of the literature to see that it is. The multiverse literature has defined the "measure problem," which is the problem of defining an additive measure over a set of multiverses. If defining an additive measure is merely a mathematical exercise in counting, then the problem would be benign. However, it is not. The measure is supposed to reflect how much we expect the various multiverses to be actualized.[22] In the conditions that largely prevail, our background evi-

20. In brief, the Bayesian needs only that the poor theory makes the outcome of our instantiation of the experiment very unlikely, whereas the good theory makes it likely. These facts are deduced within the poor and the good theories, and no consideration of other observers who may perform the experiment is needed.

21. I pass over one lingering problem arising from the choice of the wrong inductive logic: standard approaches admit no probability measure that is uniform over a countable infinity of observers.

22. Reviewing the articles collected in Carr (2007), e.g., one finds probabilities appearing in full-blown Bayesian analyses, in casual mentions, and in much in between. The idea that assigning these probabilities is an arbitrary and even risky project appears often in the multiverse literature. See, e.g., Aguirre (2007), Page (2007, 422), and Tegmark (2007, 121–22).

dence supplies completely neutral support for the actualization of each multiverse. Therefore, following the principal argument of this article, an additive measure is simply the wrong structure.

This poor choice can cause problems. Our background theories provide no grounds for various cosmic constants to take the values they do. Non-inflationary cosmology provides no reason for us to expect the curvature of the spatial slices to be as close to zero as it is. Fundamental theories simply stipulate values for basic constants like h, c, and G and give no prior reason for why they should have just the very values needed to enable our form of life.

There is a sense that these surprising values demand explanation. What argument can support that sense? The background theories provide no grounds for the parameters to have those specific values. That is, they provide completely neutral evidence. It is easy and common to represent that neutrality by saying that the prior probability of any particular value is very small. However, the redescription of neutrality by the term "low probability" brings connotations. A low-probability event in physics is commonly one that is not to be expected. If it does happen, we normally seek an explanation. By reexpressing neutral support as low probability, we have applied the wrong inductive logic. That brings artifacts. One is an unwarranted demand for explanation.

My point is not that we need no explanation for these parameter values. Rather, it is that we should look elsewhere for a justification of the need for explanation. That raises a difficult question. We cannot insist that everything needs to be explained. Such insistence triggers an unsatisfiable infinite regress. Even if we explain why the parameters have the values they do, we would then need to explain why the equations in which they figure have the form they do, and so on indefinitely. We should surely grant that some things just are the way they are and no further explanation is needed. How do we divide those things that need explanation from those that do not? My sense is that there is little intrinsic to the things that marks them as in pressing need of explanation. Rather, it is a post hoc analysis. Once we find a successor theory, inflationary or anthropic, that can explain some formerly contingent aspect of the world, then we go back and see that aspect anew as one that urgently demanded explanation.

REFERENCES

Aguirre, Anthony. 2007. "Making Predictions in a Multiverse: Conundrums, Dangers, Coincidences." In Carr 2007, 367–86.

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

———. 2002b. “Self-Locating Belief in Big Worlds: Cosmology’s Missing Link to Observation.” *Journal of Philosophy* 99:607–23.

———. 2003. “Are We Living in a Computer Simulation?” *Philosophical Quarterly* 53:243–55.

———. 2007. “Observation Selection Theory and Cosmological Fine-Tuning.” In Carr 2007, 431–43.

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

Gibbons, G. W., S. W. Hawkings, and J. M. Stewart. 1987. “A Natural Measure on the Set of All Universes.” *Nuclear Physics* B 281:736–51.

Gibbons, G. W., and Neil Turok. 2008. “Measure Problem in Cosmology.” *Physical Review* D 77:1–12.

Hawking, S. W., and Don N. Page. 1988. “How Probable Is Inflation?” *Nuclear Physics* B 298:789–809.

Howson, Colin, and Peter Urbach. 2006. *Scientific Reasoning: The Bayesian Approach*. 3rd ed. Chicago: 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* 70:647–70.

———. 2005. “A Little Survey of Induction.” In *Scientific Evidence: Philosophical Theories and Applications*, ed. P. Achinstein, 9–34. Baltimore, MD: Johns Hopkins University Press.

———. 2007a. “Disbelief as the Dual of Belief.” *International Studies in the Philosophy of Science* 21:231–52.

———. 2007b. “Probability Disassembled.” *British Journal for the Philosophy of Science* 58:141–71.

———. 2008. “Ignorance and Indifference.” *Philosophy of Science* 75:45–68.

———. Forthcoming. “Challenges to Bayesian Confirmation Theory.” In *Philosophy of Statistics*. Vol. 7 of *Handbook of the Philosophy of Science*, ed. Prasanta S. Bandyopadhyay and Malcolm Forster. New York: Elsevier.

———. Forthcoming. “Deductively Definable Logics of Induction.” *Journal of Philosophical Logic*, doi: 10.1007/s10992-010-9146-2.

Olum, Ken D. 2004. “Conflict between Anthropic Reasoning and Observation.” *Analysis* 64 (1): 1–8.

Page, Don N. 2007. “Prediction and Tests of Multiverse Theories.” In Carr 2007, 411–29.

Tegmark, Max. 2007. “The Multiverse Hierarchy.” In Carr 2007, 99–125.

Tegmark, Max, Anthony Aguirre, Martin J. Rees, and Frank Wilczek. 2006. “Dimensionless Constants, Cosmology, and Other Dark Matters.” *Physical Review* D 73:1–28.

Van Inwagen, Peter. 1996. “Why Is There Anything at All?” *Proceedings of the Aristotelian Society* 70 (suppl.): 95–120.

Weinberg, Steven. 2000. “*A Priori* Probabilities of the Cosmological Constant.” *Physical Review* D 61:1–4.

Williamson, Jon. 2009. “Philosophies of Probability” In *Philosophy of Mathematics*, ed. A. Irvine, 493–534. Amsterdam: North-Holland.