# Bayesian Inductive Inference and the Anthropic Cosmological Principle

We discuss a methodological framework for the study of cosmology, based on the rules of Bayesian inductive inference. Probabilities are defined as the degree to which hypotheses are believable, given the observations. This definition rids physical science of many of the problems caused by adopting the alternative, frequentist view. We contend that the Bayesian approach is the *only* way to make sensible statements about probability, an assertion highlighted in cosmology by the obvious uniqueness of the Universe. We illustrate our arguments with a detailed discussion of the Anthropic Principles and also refute the so-called Doomsday anthropic argument for the imminent extinction of humanity.

**Key Words:** *cosmology, anthropic principle, Bayesian Inductive Inference, philosophy of science*

## 1. INTRODUCTION

The process of scientific enquiry involves the invention and comparison of theories which compete to explain observational data. For this purpose it is sufficient to use the words theory, model and hypothesis interchangeably; science is then seen to be an exercise in inductive hypothesis testing. The testing of hypotheses on the basis of incomplete or noisy information is the realm of probability theory. In order to understand science, one must therefore understand probability.

Cosmology is the application of physical science to the Universe as a whole, so we need a conception of probability to understand

---

*Comments Astrophys.*
1993, Vol. 17, No. 1, pp. 23–47
Reprints available directly from the publisher

Printed in Singapore

cosmology too. But the Universe is unique, by definition. So how do we apply probabilistic reasoning to one-off events like the Big Bang? Is there a consistent methodological framework for deciding between competing hypotheses using measurements of unique phenomena? We believe the answer to this question is *yes*: not only does such a framework exist, it is the *only* sensible way of treating probability, even when repeated samples of the same noisy process are involved. The main purpose of this paper is to bring to the attention of the astronomer the *Bayesian* view of probabilistic reasoning and show how it leads to a coherent system for formulating and testing hypotheses about the Universe as a whole. Although many astronomers may be familiar with Bayesian methods from, for example, the field of image processing,[1] few are aware of the profound implications of adopting a rigorous Bayesian stance for all problems of inference and hypothesis testing.

A major source of controversy in contemporary cosmology (and indeed physics in general) is the role of the so-called *Anthropic Principles*: the idea that there exists a connection between material properties of the Universe and the presence of human life.[2–6] That cosmology contains much that is far from everyday human experience (quasars, black holes, etc.) is in fact an application of this idea, for human life could not exist close to such objects. In fact the early history of science is another example, since science will begin by developing a theory of the conditions which prevail in everyday life, which are precisely those conditions favourable to everyday life. Reactions to the Anthropic Principle nonetheless range, even among well-known physicists, from enthusiasm (". . . the key to quantum cosmology. . .") to outright hostility (". . . mere tautology . . . the antithesis of the scientific method"). Part of the problem is that anthropic arguments have become associated with the distorted expositions of fundamental physics associated with the New Age movement, but there is a deeper side to this controversy. Physicists and astronomers rarely take time to study the logical structure of their subject and this is the only way to assess the meaning and explanatory value of propositions like the Anthropic Principles. It is up to individuals to state whether they are satisfied with an explanation; if not, further connections must be sought. We shall spend much of this paper showing how the Bayesian stance leads to a clearer understanding of the role of anthropic arguments. We shall also take the opportunity to refute, using Bayesian reasoning, an anthropic argument (known as the Doomsday argument) purporting to show that the human race is in danger of imminent extinction.

We begin with an overview of Bayesian probability.

## 2. BAYESIAN PROBABILITY AND INDUCTIVE INFERENCE

The central principle involved in inductive inference is Bayes' theorem: if $A$, $B$ and $C$ denote propositions which may be true or false then, in standard probability notation,

$$P(A|BC) = \frac{P(A|C)P(B|AC)}{P(B|C)}. \tag{1}$$

The probability $P(A|C)$ that $A$ is true, assuming that $C$ is true, represents the *degree of belief* which it is consistent to hold in the truth of $A$, assuming (for the sake of argument) that $C$ is true. The probability is a number between 0 (false) and 1 (true) and is unique. A common objection to Bayesian probability is that it is in some way ill-defined ("subjective"); this is simply not true. Precise algorithms exist for assigning probabilities, and the same results are obtained whether these are implemented by one person's brain, another person's brain, a monkey's brain or a computer. Misunderstandings arise because different people possess different information and thus assign different probabilities to the same proposition; in general $P(A|C) \neq P(A|C')$ if $C \neq C'$. In the Bayesian interpretation probabilities are *always* conditional on other assumed truths: there is no such thing as absolute or unconditional probability. Even probability 1/6 assigned to each face of a six-sided die is conditional, on our information not discriminating between the faces and on the absence of any information concerning the throwing process.

For example, one often hears the loose statement "evolution of intelligent life was very unlikely." But what can it mean to say that something is unlikely, or improbable, given that it has happened? What is meant is, invariably, "unlikely on the basis of the information *we have*."

Probability should not be confused with proportion, numerically or conceptually, since it is just as applicable to one-off events, for example a football match or the Big Bang, as it is to repeated events like coin-tossing. The criticism that Bayesian theory is subjective, and the consequent desire to make probability an observable phenomenon by identifying it with proportion, have motivated the *frequentist* interpretation of probability. We shall criticise the frequentist view of probability below and demonstrate some of the logical problems to which it leads.

Probability theory, in our view, is not a branch of physics but a branch of logic. Like any branch of mathematics, it cannot be tested by experiment but only by the requirement that it be self-consistent. Probability always applies when there is insufficient information for deductive certainty, and deductive logic—Boolean Algebra—is just a special case; probability *is* inductive logic. Probabilities need to be both assigned and manipulated; they are assigned using the principle of maximum entropy,[7,8] which we shall not discuss here. The rules for manipulating probabilities were originally derived from principles of consistency by Cox[9] and are no more than the familiar product and sum rules:

$$P(AB|C) = P(A|BC)P(B|C), \tag{2}$$

$$P(A|C) + P(\bar{A}|C) = 1, \tag{3}$$

where $AB$ denotes the logical product of $A$ and $B$ ("and") and $\bar{A}$ denotes the negation of $A$. The product rule derives from associativity of the logical product: if $P(AB|C)$ is written as an unknown function of $P(A|BC)$ and $P(B|C)$ and this relation is used to decompose the logical product of *three* propositions into probabilities of single propositions, the two distinct ways in which this can be done must give the same result. This induces an equation for our unknown function whose solution is (2), the product rule. The sum rule derives similarly from associativity of the logical sum, which is related to negation via the logical relation $\overline{A + B} = \bar{A}\bar{B}$. Therefore the laws of probability are simply consequences of associating scalar numbers with propositions. An immediate corollary of the sum and product rules is the *marginalising rule*:

$$P(B|C) = P(AB|C) + P(\bar{A}B|C). \tag{4}$$

This, together with the commutativity of the logical product, $AB = BA$, leads to Bayes' theorem (1). The denominator (4) can be further partitioned:

$$P(B|C) = P(A|C)P(B|AC) + P(\bar{A}|C)P(B|\bar{A}C). \tag{5}$$

What is new here is not the formulae, but the interpretation lent to probability. It is degree of consistent belief, and furnishes a perfectly well-defined theory. All of these formulae apply trivially to *proportions* of the Venn diagram, but this is a conceptually distinct matter. The framework of Bayesian probability furnishes a consistent inductive methodology for science.[10] Hypotheses compete in assigning different probabilities to the results of experiment or observation. The hypothesis which assigns the highest probability to the data (judged by the statistics of the noise) is favoured. Bayes' theorem is a principle of inverse reasoning which tells us how to combine the prior probabilities of the competing hypotheses with the probabilities they predict for the data so as to establish their updated posterior probabilities: denote by $H$ a hypothesis, $D$ the observed data and $I$ all the relevant prior information about the experiment. The posterior probability $P(H|DI)$ is given in terms of the prior $P(H|I)$ (assigned using the maximum entropy principle), and the likelihoods $P(D|HI)$ and $P(D|\bar{H}I)$ of observing these data under the assumptions that the hypothesis is respectively true or false, by

$$P(H|DI) = \frac{P(H|I)P(D|HI)}{P(H|I)P(D|HI) + P(\bar{H}|I)P(D|\bar{H}I)}. \tag{6}$$

Suppose, formally, that we have logically exclusive hypotheses $H_1 \ldots H_n$, which we take to be exhaustive (of our knowledge!), and we wish to decide between them on the basis of some noisy data $D$ and prior information $I$. Suppose our experiment involves the measurement of a parameter; the probability of the proposition that the parameter takes the value $\lambda$ is, by a generalisation of the marginalising rule derived under exclusivity and exhaustivity,

$$\begin{aligned} P(\lambda|DI) &= \sum_i P(\lambda H_i|DI) \\ &= \sum_i P(\lambda|H_iDI)P(H_i|DI) \\ &= \sum_i P(\lambda|H_i)P(H_i|DI) \end{aligned} \tag{7}$$

(since $D$ and $I$ are not used to predict $\lambda$ in any of the theories). This expression is a weighted sum of the probability distributions predicted by each theory for the parameter and the weights are just the probabilities of the respective theories. Likewise (6) generalises to

$$P(H_i|DI) = \frac{P(H_i|I)P(D|H_iI)}{\Sigma_j P(H_j|I)P(D|H_jI)}. \tag{8}$$

Strictly speaking this procedure should always be adopted when predicting a parameter: all known viable theories are to be marginalised over. Usually, however, we find that one $H_i$ has a probability close to unity after some crucial experiment, and is thereafter accepted; all others have probability close to zero and are discarded.

The methodology that emerges involves theory creation—the hardest part, for which we are still totally dependent upon the human brain—theory comparison, and theory selection, in an endless loop reflecting the continuing process of generating, testing and refining theories driven by both theoretical innovation and experimental discovery. The crucial comparative step, however, is *inductive* rather than *deductive*.

An important feature of Bayesian reasoning is that it allows a quantitative statement of Ockham's Razor: that the simplest theory which fits the data closely should be preferred. There is a tradeoff between simplicity of theory and closeness of fit. This principle applies when one theory generalises another by having extra ("floating") parameters which have a definite but unknown value which must be estimated from the data. (These data contain noise whose statistics are taken as given; if the noise did not exist, a single measurement would resolve the issue.) The less general theory by contrast assigns some pre-ordained value (most often zero) to the floating parameters. Clearly, the greater freedom of the generalised theory can be exploited to fit the data better; intuition, however, warns us not to accept a theory with 50 floating parameters just because it manages to fit the data. Bayesian reasoning expresses this quantitatively: the more general theory pays a penalty by distributing some of the prior probability for the test parameter on values where the data and noise statistics subsequently indicate it is very unlikely to lie. The original theory, by contrast, places all of its prior probability eggs in one basket, at a single value of the parameter. Bayesian reasoning then tells us which of the theories to prefer[11,12]; such a formalisation of Ockham's Razor is one particularly gratifying outcome of the Bayesian position.

Astrophysical examples where this analysis would prove useful, once the noise statistics are agreed, include the cosmological constant term in Einstein's field equations and the Brans–Dicke[13] generalisation of these equations, the fifth force, the neutrino rest mass, the cosmological density parameter $\Omega_0$, and the general problem of assessing the presence and strength of real signals in noisy data, for example in searches for faint galaxies (or stars, planets, cosmic microwave background anisotropies). All of these have profound implications for cosmology.

The same principle underlies the drive for the unification of physical theories, by which is meant the equally accurate description, from a single theory, of phenomena which were previously described by distinct theories with separate sets of parameters. This demonstrates also that theoreticians should not stop even if present-day theory is in good accord with every available observation: another theory might do the same with fewer parameters, and therefore be favoured.

## 3. OTHER VIEWS OF PROBABILITY AND SCIENTIFIC METHODOLOGY

Before going any further, it is important to stress the differences between our view of probability and the alternative *frequentist* view.

First, our viewpoint is enormously more general than the frequentist notion that probabilities are relative frequencies (proportions) in some real or imaginary ensemble. In the frequentist view, a proposition is at once true in some elements of the ensemble and false in others, and in this way frequentists erroneously substitute AND for OR. The Bayesian position is also free from problems associated with the failure to incorporate any prior information that cannot be expressed in frequency terms; you would not trust a doctor who told you that 80 per cent of persons with your symptom needed an operation, but who refused to examine your personal medical files!

More importantly it is clear, from the fact that (2) and (3) are derived wholly from the requirement of logical consistency, that any deviation from these rules must be a logical inconsistency. Many of the data analysis techniques collectively known as sampling theory, used when the data comprise repeated noisy samples of a variable having definite (but unknown) value, are not equivalent to Bayesian reasoning. These methods give good answers—meaning, with hindsight, close to the Bayesian—in the kind of problems for which they were evolved; nevertheless, they can all be made to look arbitrarily illogical by suitable choice of dataset, and Jaynes[14] gives examples. Frequentists often talk about *random processes* or *random variables*; to a Bayesian there are no random variables, only variables whose (fixed) values we do not know. A "random" process is simply one about which we only know sufficient to specify probability distributions for its parameters. Repeated samples comprise proportions, not probabilities.

In contrast, no paradox has ever been found in the *correct* application of Bayesian methods. Furthermore, in order to deal with unique events like the weather, frequentists introduce the notion of an *ensemble*, a (perhaps infinite) collection of imaginary possibilities. This allows them to retain the notion that probability is a proportion within this collection. Provided the calculations are done correctly, numerical results should agree with the Bayesian results, but frequentists often talk about the ensemble as if it were real and this is dangerous: there is only one, imperfectly known, system.

Still more disturbing is the influence that frequentist and other non-Bayesian views of probability have had upon the development of the philosophy of science. Carnap,[15,16] a logical positivist, attempted a theory of inductive reasoning but failed to employ Bayes' theorem in the correct way. We reject Carnap's idea that there are two kinds of probabilities—logical and factual—in favour of a single coherent definition. Other prominent philosophers of science reject the notion that inductive arguments have any epistemological value. This anti-inductivist stance, often somewhat misleadingly called *deductivist*, is evident in the thinking of Karl Popper, Thomas Kuhn and, latterly, Paul Feyerabend. Regardless of the ferocity of their arguments with each other they have in common that at the core of their philosophies lies the rejection of inductive reasoning. This line of thought began with the work of the empiricist philosopher, David Hume (1711–1776). For a thorough analysis of the anti-inductivist philosophers just mentioned, and for their debt to Hume, see David Stove[17]; we shall restrict ourselves to some brief comments here.

Karl Popper began as a frequentist, but changed to a *propensity* view of probability that nonetheless remained physical and testable.[18,19] Popper has asserted that all observations are "theory-laden" and that "sense-data, untheoretical items of observation, simply do not exist."[20] This is not true. Data are *numbers*, and can be incorporated as propositions into any theory we wish. Popper also believes that no theory becomes more probable when evidence in its favour is discovered; and that each theory begins by being infinitely improbable (and presumably remains so). The first assertion is nonsensical: "evidence in favour" can only mean "makes the theory more probable." The second assertion carries the implicit assumption that there are an infinite number of theories waiting around to be discovered; but probabilistic reasoning can only involve those theories which *have* been found, and with which we can generate testable predictions. This is precisely what we would wish to do.

A more serious problem is Popper's demand that theories should be *falsifiable*, a deductive notion which should logically mean that the theory can be decisively rejected. The disparaging implication that scientists live only to prove themselves wrong comes from concentrating exclusively on the possibility that a theory might be found to be *less* probable than its challenger. In fact evidence does not confirm, or discount, a theory; it either makes the theory more probable (supports it) or makes it less probable. For a theory to be useful, it must be capable of having its probability altered by incoming data, and the appropriate criterion is therefore *testability*. Bayes theorem tells us, by inverse reasoning, that a testable theory will not predict all things with equal facility.

Thomas Kuhn, another anti-inductivist, in his popular and influential book *The Structure of Scientific Revolutions*[21] neglected in his methodology the crucial inductive step by which theories are compared, and never formulated a coherent way of doing this. Indeed Kuhn himself states "we may . . . have to relinquish the notion that changes of paradigm carry scientists . . . closer and closer to the truth." (It is Kuhn who popularised the word *paradigm* to describe the mode of thought corresponding at any instant to the prevailing theory.) This reasoning is taken to an extreme by Feyerabend[22,23] who describes himself as an epistemological anarchist and asserts that "anything goes." Yet the only proper respect in which anything goes is in the formulation of new theories and experiments; everything else is a well-defined process. Feyerabend, however, asserts that normal science is a "fairytale" and is qualitatively no different from "astrology, acupuncture or witchcraft."[24] He suggests further that scientific theories should be chosen by majority vote rather than experiment and comparative hypothesis testing. Few practitioners of science have any sympathy for such excesses; one might more reasonably view this as a *reductio ad absurdum* argument against the rejection of induction.

A specific field within contemporary science which would benefit greatly from the application of Bayesian reasoning is quantum mechanics. The formalism of quantum mechanics does not predict the outcome of single measurements of observables, but only the probability of each eigenvalue. Quantum theory still stands as one of the century's premier intellectual achievements; however, its dominant, "Copenhagen," interpretation arbitrarily instructs scientists not to ask what might happen in any one run of the experiment, denying even the meaning of the question.

Now, it cannot be denied that there may be no underlying reason. But, if you accept that, how would you ever know? You will never go out and see if you can do any better. And, if you did succeed, you would have a new theory which beats quantum mechanics in the hypothesis-testing game—precisely the aim of science. The Copenhagen view is, literally, anti-scientific. Here is the source of all the mysticism which has been written about quantum theory.

As an illustration of Copenhagen logic, suppose I drive two cars, which with the tools available to me I cannot distinguish from each other, round the same track at the same speed. One breaks down; the other doesn't. Copenhagen asserts that there is no meaning in asking why this is so, and no point in looking more closely at the cars beforehand to try to distinguish differences which cause one of the cars to fail. Motorists would not be content with this standard of logic, and nor should physicists.

It follows that any new theory which seeks to predict individual measurements of an observable within a run must be based on differences, which present-day technology cannot discern, within the system at each measurement. The variables of this theory are, at present, hidden from us, and are therefore termed *hidden variables*. They have long been unfashionable, but they are the only way to make progress; even quantum field theory merely defers the problem. Although a hidden variable theory will have more parameters than quantum theory, its capacity to fit the results of individual measurements renders it indefinitely more (less, if a bad theory) probable than quantum theory as observations accumulate. Prior biases become overwhelmed by the data.

One thing we can say about the hidden variables is that they are nonlocal; this is Bell's result.[25] (A Bayesian exposition of Bell's theorem, treating spin measurements of one particle as giving information about its internal hidden variable, and through the interparticle correlation about a second particle, is given by Garrett.[26] Spin measurements on the second particle cannot be accounted for in this way, and so the hidden variables are not internal to the particles—they are nonlocal.) Nonlocality in turn explains why the hidden variables are so difficult to influence, since experimental science concentrates on isolating something in order to study it—an impossibility with a nonlocal entity. The extreme abstraction of quantum gravity research is, by all precedent, another signal that something more is needed; and, of course, nonlocality has extreme implications for quantum, or post-quantum, cosmology.

In summary, Copenhagen says "the system itself does not know what will happen," while hidden variables say "the system does know—today's experimenters don't know" (because they lack the necessary information). There is no "random" process; only an unknown one. The so-called Many-Worlds interpretation of quantum mechanics, in which the Universe is held to branch into an ensemble of worlds corresponding to the different eigenvalues, is simply a more sophisticated version of Copenhagen: it provides an (untestable) rationalisation for why prediction supposedly cannot be improved, and why we should give up. It is also clearly motivated by the incorrect frequentist viewpoint of probability.

Bayesian analysis also leads to a coherent understanding of the origin of irreversibility in thermodynamics, another subject of concern for cosmologists wishing to understand the "Arrow of Time"; see the Appendix.

## 4. THE ANTHROPIC PRINCIPLE(S)

Now let us apply a Bayesian inductive scheme to the Anthropic Principles, following Garrett.[27] Suppose we have a model of the Universe $M$ that contains various parameters which can be fixed by some form of observation. Let $U$ be the proposition that these parameters take values $U_1, U_2, \ldots$. Crucial in the discussion of anthropic selection arguments is the role of life; let $L$ be the proposition that intelligent life evolves in our Universe. In terms of Bayes' theorem,

$$P(U|LM) = \frac{P(U|M)P(L|UM)}{\Sigma_{U'}P(U'|M)P(L|U'M)},$$

where $U'$ is a dummy variable of summation. The dependence of $P(U|LM)$ on $P(L|UM)$ demonstrates that the values of $U$ for which

$P(L|UM)$ is larger correspond to larger values of $P(U|LM)$: since life *is* observed, those parameter values which make life more probable are clearly preferred. To go any further we have to say something about the likelihood $P(L|UM)$ and the prior $P(U|M)$. Although we know the general principles for doing this,[7] in this case it is too difficult to do exactly. We can be sure, however, that the prior will be broad rather than sharply peaked. If now the likelihood is sharply peaked in $U$, this peak is projected directly to the posterior $P(U|LM)$. The likelihood is assigned from our knowledge of how stars form and shine; how planets are created in orbits around stars; what conditions are needed for life on a planet; how evolution creates intelligence, and so on. There are gaps in our knowledge here; nevertheless, if any one step in the chain is sharply peaked in $U$, we can marginalise over any intermediate steps using (4) and still end up with a sharp peak in the likelihood $P(L|UM)$, and so also in $P(U|LM)$. There are good reasons for supposing that intelligent life must be carbon-based, and therefore evolve on a planet. It is therefore reasonable to suppose that $P(U|LM)$ should be peaked, so that there is a *correlation* between the propositions $U$ and $L$. This is exactly the type of correlation that Bayesian reasoning can successfully exploit.

Let us suppose that the relevant parameters in $U$ include Newton's constant of gravitation $G$, the charge on the electron $e$, the mass of the proton $m_p$. Our analysis above indicates that there must be a correlation between the presence of life and the values that these constants take, in full accord with our methodological scheme. Moreover, there is no reason why such arguments could not be used to find the values of fundamental constants ("parameters") in advance of their direct measurement; in fact, Hoyle[28] successfully predicted, from the existence of carbon-based life, a resonance in the spectrum of carbon-12. (In any case, the time ordering of experiment and theory is no more than historical accident.) An illustration of the logic is a plant whose seeds germinate only after prolonged rain. A newly germinated (and intelligent) specimen could observe dampness in the soil directly, or predict it from this self-knowledge coupled with the observation of its own germination. We hope this convinces the reader that the Anthropic Principle, used properly, *is* predictive. To illustrate further the explanatory power of this kind of argument, consider the explanation by Dicke[29] of the apparent coincidence between the dimensionless ratio $Gm_p^2/\hbar c \simeq 0.6 \times 10^{-38}$ and the inverse square root of the number of nucleons in the Universe. If the Universe has mean density ρ and has radius $ct$ then the number of nucleons will be

$$N \simeq \frac{4\pi}{3}\,\rho\,\frac{(ct)^3}{m_p}. \tag{9}$$

For a matter-dominated Universe, whose evolution follows the Friedman equation, the density varies with time as $\rho \simeq (Gt^2)^{-1}$ so that $N \simeq c^3 t/Gm_p$ (ignoring constants of order unity). But the elements found in our bodies (such as iron, required for the existence of haemoglobin) are created mainly in supernova explosions. Now the time taken for supernovae to form in an expanding Universe can be estimated from the mass and luminosity of typical stars, since the luminosity is related to the mass loss rate through photon emission. The result is

$$t_s = \left(\frac{\hbar c}{Gm_p^2}\right)\left(\frac{\hbar}{m_p c^2}\right) \simeq 10^{10}\ \text{years}. \tag{10}$$

Hence,

$$N \simeq \frac{c^3}{Gm_p}\left(\frac{\hbar c}{Gm_p^2}\right)\frac{\hbar}{m_p c^2} \simeq \left(\frac{Gm_p^2}{\hbar c}\right)^{-2}. \tag{11}$$

This argument is just one example of a number that exist of this type and it has clear explanatory power.[2–5] To see how this kind of argument fits into the methodology we have discussed above, consider the question: How *surprised* should we be that the constants of nature take their particular values? This question clearly requires a probability-based answer: the smaller the probability of observed values given our prior knowledge, the more surprised we should be to find them. But this surprise must be bounded in some way: we know that the values have to lie *somewhere*. (In fact, we can usefully define surprise to be the logarithm of the reciprocal of the probability, whereupon the sum rule (3) expresses the bounding idea.) Obviously, to be surprised at both the truth and the falsehood of the same proposition is to reason inconsistently.

Arguments of this type were named the *Weak Anthropic Principle* by Brandon Carter,[2] as a corrective to the Copernican principle that humanity does not occupy a privileged place in the Universe. In a subsequent paper, Carter[4] goes some of the way towards elucidating the Bayesian nature of the reasoning and gives further examples.

We now move on to a more speculative version of the argument, called the *Strong Anthropic Principle*, that the Universe *must* have those properties which allow life to evolve within it; in other words the existence of life is itself a law of nature. It is clear that this type of argument has a different logical status from the previous ones, which were simply applications of inductive reasoning; it is a *proposition*, something whose probability should be *assigned* using Bayesian inductive reasoning. So, given that the Universe *does* have those properties conducive to life, does it make any sense to say that it *must*?

The Strong Anthropic Principle (SAP) asserts that if $U_L$ and $U_{\bar{L}}$ are the sets of parameter values that are conducive and detrimental to life respectively, then

$$P(U \in U_L | SAP, I) = 1, \quad P(U \in U_{\bar{L}} | SAP, I) = 0, \qquad (12)$$

where $I$, as above, denotes any prior information we have apart from the existence of life. Now, obviously,

$$P(U \in U_L | I) = 1, \quad P(U \in U_{\bar{L}} | I) = 0. \qquad (13)$$

The similarity of (12) to (13) does not imply the truth of the SAP from the existence of life. All we can do is manipulate these statements using the sum and product rules. We can derive, for example, that

$$P(SAP | L, I) = P(SAP,\ U \in U_L | L, I) \\ + P(SAP,\ U \in U_{\bar{L}} | L, I) \qquad (14)$$

whence, with some manipulation,

$$P(SAP | L, I) = P(SAP | U \in U_L, L, I), \qquad (15)$$

which is clearly of no help. The logical reason is that two observations which separately lead to the same conclusion need not imply each other: during daylight hours, both rain and the absence of direct sunlight imply the existence of clouds; but the absence of direct sunlight does not imply rain.

As far as Bayesian logic is concerned it makes no difference what motivates the SAP, since the role of probability theory is to work with given propositions and hypotheses, not to generate them or query their origins. Nothing prevents us from advancing a new, perhaps deeper, theory at any stage. But the SAP is unsatisfactory: it does not tell us how to redistribute the prior probability of those values of $U$ conducive to life among the values not conducive to life. To do this we would need more details of the reasoning leading to the SAP in the first place; only then can we make a useful assignment of the probability and its predictions. The SAP is therefore not in itself predictive to any useful degree. In fact the first example, given by Carter,[2] is not even of SAP type: it is a weak anthropic argument. The same is true of Carr and Rees' discussion[3] of cosmological models in which the fundamental "constants" $U$ vary with spatial and temporal position in the Universe. In such models, humanity must exist in a region of space-time where $U \in U_L$. A specific realisation of this is Linde's eternal inflationary scenario,[30] in which spontaneous symmetry-breaking is invoked so that the "constants" describing physics below the Grand Unification energy take different values in mutually incommunicable regions of the Universe. Such models do, however, involve extremely speculative physical assumptions about high-energy physics, and no compelling testable theory of this type has yet been advanced.

We have adopted the definition of SAP used by Barrow and Tipler,[5] but there are other kinds of argument. We can find arguments based on deeper theories of fundamental physics; if the fundamental constants were different there might then arise a fundamental inconsistency in the theory. Superstring theory, for example, constrains the number of dimensions of space-time in this way. This argument makes no explicit reference to life. Another kind of argument is teleological or *causative* in the human, psychological sense. But probability theory treats only *correlations*—between the presence of life and the conditions needed for it. And the formulae of a physical theory tell only of correlations between variables. The only valid meaning of *causality*, in physics, is that which enters the formulae, of time-ordering. We contend that much of the confusion surrounding the SAP stems from failure to distinguish between the human, purposive meaning of *causation* and its adoption to mean *correlation*. Teleological theories typically unite physics and biology. Since physics and biology are complementary and work on different levels of description, such theories must possess quite extraordinary features. The role of consciousness in collapsing the quantum mechanical wavefunction of a system is a particularly grotesque example. In fact, wavefunction collapse has been satisfactorily *derived*, in a simple problem, from the quantum mechanics of the joint system of observer-and-observed and the requirement that the observer be treated in the classical approximation[31] (we do not agree with all of van Kampen's comments about probability). We are therefore highly skeptical about claims that conscious observers in a sense *create* the Universe they observe, as has been suggested by, for example, Patton and Wheeler[32] and dubbed the *Participatory Anthropic Principle* by Barrow and Tipler.[5] Other SAP arguments based on quantum branching into an ensemble of Universes[4,5] involve the frequentist-inspired confusion we discussed above.

A different version of this argument, which may be held at a subconscious level, is to see the SAP as a facet of God ($G$): the theological axiom that $P(L|G) = 1$ leads through Bayesian reasoning to evidence for the existence of God. This argument, however, advances neither science nor theology.

## 5. REFUTING THE DOOMSDAY ARGUMENT

Now we analyse an argument related to the anthropic selection effects discussed above. This argument, outlined below, purports to show that it is a reasonable inference, from our present existence, that all life is about to end: a Doomsday argument.[33,34] In neither of these references[33,34] (nor in subsequent work by Leslie, the main advocate) is the Doomsday argument phrased in Bayesian terms by calculating the probability of our existence at a particular human epoch assuming Doom at some later epoch, and then using this as a likelihood in Bayes' theorem to incorporate our observed existence into the distribution for Doom's date. The absence of such a formulation means that the argument is incompletely specified, and that any criticism can be countered by saying "that's not what I meant!" Below, we translate the Doomsday argument into Bayesian reasoning, and show that it does not survive the process; we also present a quantitative counter-example. Bayesian logic clarifies the discussion and highlights any errors, in line with our discussion of Section 2 (Leslie[34] seems implicitly to doubt this). While we accept that we may have missed some relevant conditioning information we do require arguments addressed to us to be phrased in Bayesian language, with the chosen conditioning information made explicit at every point. The most common cause of discrepancy between intuition and Bayesian calculation is the absence, from the calculation, of information which is implicitly exploited by our intuition. All relevant prior information should be included.

The Doomsday argument, as set out by Leslie,[34] is this: that one should prefer the scenario which is least surprising, and that it is less surprising to find oneself a fair way through the span of humanity—as is the case if there are fewer people following—than to find oneself born very early in our race. Therefore our existence, at the present point in time, tends to favour fewer descendants.

Denote by $I$ the propositions taken to be true to start with—the prior information—by $N$ the number of humans who will ever exist, and by $n_R$ the information that you, the reader, are the $n$th human to be born; $R$ stands for your own name. Then Bayes' theorem tells us that

$$P(N|n_R I) = KP(N|I)P(n_R|NI), \tag{16}$$

where the normalising constant $K$ is determined *a posteriori*:

$$K^{-1} = P(n_R|I) = \sum_N P(N|I)P(n_R|NI).$$

We can, if we prefer, eliminate $N$ (or $n_R$) in favour of $f_R \equiv n_R/N$, the fraction of humanity born up to and including yourself. The first part of Leslie's assertion is that one should prefer values of

$N$ for which the posterior surprise $s(N|n_R I)$ is smaller; with $s = -\log P$ this means preferring those values having larger probability. Such choices are decision-theoretic: Bayesian logic has as its output only probability distributions but, if a single number is demanded, we wholeheartedly agree with Leslie's choice. Second, Leslie states that the surprise at being in a particular fraction of humanity is larger, so that the probability is smaller, if the fraction is an early one rather than a central one. This is an incomplete statement because it does not specify upon what conditioning information the reasoning is based; your surprise at learning a friend has been knocked over by a car is less if you know he was blackmailing prominent underworld figures, for example. Unless the conditioning information is made explicit, no definite statement is made. Nor can one reason backwards from the probability so as to reveal the conditioning information: a glance at Bayes' theorem indicates that this procedure is horribly non-unique. Consequently, there is no more to be said about the Doomsday argument *per se*.

The Doomsday argument has been compared by Leslie[34] to an urn model, in which the balls in an urn represent all humans born up to the end of the species and withdrawal of a ball from the urn corresponds to being born. The lottery is conducted in such a way that no ball has any information distinguishing the remaining balls for selection purposes; the balls do not know how numerous they are. This is a perfectly good exercise in probabilistic reasoning but as an analogy to the Doomsday argument it is very poor, for any ball is equally likely, viewed in advance, to be withdrawn at any stage. (This remains true without replacement of the balls, as here.) Therefore the model fails to encapsulate Leslie's intuition that the probability of being withdrawn in any fixed fraction is less early on. Perhaps Leslie believes this because it is common for probability distributions to fall off away from the mean; but this is untrue for the uniform distribution derived here. A more serious objection is that the mechanism of human life provides us with relevant conditioning information of a different sort from the urn model: information about names, family trees, generations, meetings, pregnancies and so on. For example, the ball corresponding to John, son of Jack, should have zero probability of being selected if Jack's ball has not yet been withdrawn. In fact, if the prior information $I$ contains lineal data of this sort, then the likelihood

$P(n_R|NI)$ is assigned entirely from this and is not dependent on $N$ at all: it depends on such matters as which, of many roughly equally pregnant women, gives birth first. This independence of the likelihood from $N$ implies that the posterior distribution for $N$ is equal to the prior so that nothing is learned about $N$.

Leslie himself[34] casts doubt on the urn analogy by asserting that it is not as if pre-existing souls are placed, one by one, into successive human bodies. Whether this is true is a matter for theologians; Leslie means (and we agree) that, even if we knew this to be so, it would make no difference for the purpose of inferring $N$. The reason is that labels ("pre-names") on the souls convey no information to us about the bodies they would enter.

Perhaps Leslie means that it is less surprising to be born later because one then has the impression of a greater choice of possible parents. A probability analysis can be based on this. But because Leslie does not state his conditioning information we cannot be certain what he means, and it seems he would argue that being born very late in humanity is as surprising as being born very early, in which case this is not what he has in mind.

Another point is that if the drawing of a red ball from an urn containing one red and $N - 1$ green balls takes place early on, say, at the seventh draw, Leslie inclines (correctly) to believe that there are more likely to be 20 than 10,000 green balls. This is because the likelihood, given no knowledge discriminating between the balls in the drawing process, is proportional to $1/N$, weighting the prior distribution to smaller values of $N$ *en route* to the posterior. But, once again, the conditioning information which we have does not correspond to this model.

A warning that the Doomsday argument is incorrect comes from the absence, from it, of any Doomsday *mechanism*. One might also expect that the longer humanity survives, the better, not the worse, is the evidence that it will continue to survive. Let us translate this insight into quantitative Bayesian terms. Suppose (for simplicity) that according to present knowledge there is a constant probability $\alpha dt$ that the human race will be obliterated in a time interval $(t, t + dt)$ by, for example, a large meteorite strike. Then the probability that humanity will still exist at time $t$ after its origin is $e^{-\alpha t}$, and we use our information that it has survived to the present time $t_0$ to make an estimate of $\alpha$ as above. We can trans-form freely between $t$ and $n(t)$, the number of births up to time $t$, using any population growth model we choose. If we take $n(t)$ as our variable we highlight the disagreement between our analysis and the Doomsday argument. From Bayes' theorem, with density $P(t_0|\alpha, I) = e^{-\alpha t_0}$, we have

$$P(\alpha|t_0, I) = \frac{P(\alpha|I)e^{-\alpha t_0}}{\int_0^\infty d\alpha' P(\alpha'|I)e^{-\alpha' t_0}}. \tag{17}$$

We see that, as $t_0$ increases, the posterior density $P(\alpha|t_0, I)$ becomes weighted towards smaller $\alpha$; for example, using a constant prior $P(\alpha|I)$ we find $P(\alpha|t_0, I) = t_0 e^{-\alpha t_0}$. The expected lifetime of the human race is

$$\langle t\rangle = \int_{t=0}^{\infty} t\, dP(\text{extinction in } (t, t+dt) \text{ and no extinction up to } t|t_0, I), \tag{18}$$

which here simplifies, using the laws of probability, to

$$\langle t\rangle = t_0\left[1 + \frac{\int_0^\infty d\alpha P(\alpha|I)(1 - e^{-\alpha t_0})/\alpha t_0}{\int_0^\infty d\alpha P(\alpha|I)e^{-\alpha t_0}}\right]. \tag{19}$$

The second term in the square brackets represents the expectation beyond the present time $t_0$. For a constant prior $P(\alpha|I)$ this term diverges, but for more realistic prior distributions it is well-behaved. The expected time beyond $t_0$ *increases* as $t_0$ increases, in accordance with our intuition and contrary to the Doomsday argument. This model also invokes a mechanism which appears both in the prior and the likelihood, and in which time plays a central role, in contrast to the Doomsday idea. Our model also highlights another shortcoming of the Doomsday argument: the probability of a Doomsday occurring sometime is unity in our model, so that the question is no longer *if* but *when*. A more sophisticated model would include the possibility that Doomsday is not inevitable.

Some religions assert that certain individuals have been, or will be, chosen by a process not known (to us) to discriminate between humans. This *is* similar to the urn, and believers in these religions will generate predictions about Doomsday, consistent with their religious beliefs, which differ from the predictions of unbelievers. (Of course there is further religious information about Doomsday which might also be included, a trivial example being that Doomsday cannot take place before the coming of a promised person.)

## 6. CONCLUSIONS

We believe that a methodology based on Bayesian inductive inference is the appropriate methodology for science in general and cosmology in particular. The Bayesian framework incorporates a definition of probability that has no difficulties of interpretation when applied to unique events. A Bayesian interpretation of probability should form the proper basis for attempts to study quantum cosmology.

We have emphasised that Bayesian inductive inference is nothing other than the generalisation of deductive logic to the case where knowledge is incomplete, for example when the data suffer from noise. The procedure is unambiguous and is fully defined (when combined with the principle of maximum entropy for the assignment of prior probabilities) and is in full accord with intuition. It further allows a quantitative statement of Ockham's Razor: a principle that ought to be applied more frequently than at present in modern-day theoretical cosmology! The extraordinarily speculative nature of modern, quantum cosmology has been chronicled by Oldershaw[35] and, with comments on the Bayesian status of the Anthropic Principle, by Newton-Smith.[36] Silence is a wiser counsel than unbridled speculation; there is nothing disreputable in "I don't know," for if scientists knew everything then the research side of their profession would no longer exist!

Looking at the Anthropic Principles from a Bayesian point of view reveals clear differences between arguments of the Weak and Strong Anthropic type. Weak Anthropic arguments have clear explanatory power, whereas arguments of the Strong Anthropic type are either weak ones in disguise or ideas having no explanatory or predictive value. We have also taken the opportunity to refute a common argument for the imminent extinction of the human race by giving a correct Bayesian analysis of the inductive logic involved.

## APPENDIX: THE THERMODYNAMIC ARROW OF TIME

Define the microstate of a system as specified by the positions and velocities of all particles. Define the macrostate as specified by macrovariables such as pressure, density, temperature and any others needed to generate a relation between them which is reproducible to the experiment's satisfaction. The microstates of a system consistent with the initial macrostate all evolve reversibly, by Hamilton's equations, to be wholly contained within the set of microstates consistent with the final macrostate; for otherwise the experiment would not be reproducible in the macrovariables. This set of final consistent microstates is therefore not smaller than the set of initial consistent microstates, and for a non-adiabatic process it is in fact larger: denoting the number of consistent microstates by $W$, we have $W_{\text{final}} > W_{\text{initial}}$, whence if $S \equiv k \log W$, $S_{\text{final}} \geq S_{\text{initial}}$. This relation reflects a loss of *information* about the microscopic details, and $k \log W$ can be shown to be the constrained maximum value of the information entropy $-\mathrm{Tr}\,(\rho \log \rho)$ where $\rho$ is the phase-space density, subject to the values of the macrovariables, when the number of particles is large.

The crucial point is that there is no paradox between reversible microscopic evolution and irreversible macroscopic evolution because the latter is not derivable *dynamically* from the former: the concept of information is also needed, and information loss is irreversible. Ergodic theory, while valuable in few-body problems such as the solar system, is simply not relevant to statistical mechanics. The correct idea is, more fully, an application of the principle of maximum entropy[7]; for more detail of the present outline see Garrett.[8]

The so-called Heat Death of the Universe applies strictly to macrovariables and, since macroscopic reproducibility is a user-defined concept, caution must be exercised in such arguments.

ANTHONY J. M. GARRETT
*Inference Consultants,*
*Byron's Lodge, 63 High Street,*
*Grantchester,*
*Cambridge CB3 9NF,*
*United Kingdom*

PETER COLES
*School of Mathematical Sciences,*
*Queen Mary & Westfield College,*
*Mile End Road, London E1 4NS,*
*United Kingdom*

## References

1. S. F. Gull and G. J. Daniell, Nature **272**, 686 (1978).
2. B. Carter, in *Confrontation of Cosmological Theories with Observational Data*, ed. M. S. Longair, IAU Symposium **63** (Reidel, Dordrecht, 1974), pp. 291–298.
3. B. J. Carr and M. J. Rees, Nature **278**, 605 (1979).
4. B. Carter, Phil. Trans. R. Soc. Lond. A **310**, 347 (1983).
5. J. D. Barrow and F. J. Tipler, *The Anthropic Cosmological Principle* (Oxford University Press, Oxford, 1986).
6. Y. V. Balashov, Am. J. Phys. **59**, 1069 (1991).
7. E. T. Jaynes, *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, ed. R. D. Rosenkrantz, Synthese Series **158** (Reidel, Dordrecht, 1983).
8. A. J. M. Garrett, in *Maximum Entropy in Action*, eds. B. Buck and V. A. Macaulay (Clarendon Press, Oxford, 1991), pp. 139–170.
9. R. T. Cox, Am. J. Phys. **14**, 1 (1946).
10. A. J. M. Garrett, Analytical Proc. R. Chem. Soc. **29**, 426 (1992).
11. A. J. M. Garrett, in *Maximum Entropy and Bayesian Methods, Laramie, Wyoming, 1990*, eds. W. T. Grandy and L. Schick (Kluwer, Dordrecht, 1991), pp. 357–364.
12. A. J. M. Garrett, Physics World **4**(5), 39 (1991).
13. C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
14. E. T. Jaynes in *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, eds. W. L. Harper and C. A. Hooker (Reidel, Dordrecht, 1976), pp. 175–257. [Reprinted as Chapter 9 of Ref. 7.]
15. R. Carnap, *Logical Foundations of Probability* (University of Chicago Press, Chicago, 1950).
16. R. Carnap, *The Continuum of Inductive Methods* (University of Chicago Press, Chicago, 1952).
17. D. C. Stove, *Popper and After: Four Modern Irrationalists* (Pergamon Press, Oxford, 1982).
18. K. R. Popper, *The Logic of Scientific Discovery* (Hutchinson, London, 1959). [Translation of *Logik der Forschung* (Springer-Verlag, Vienna, 1934).]
19. K. R. Popper, *Postscript to the Logic of Scientific Discovery (3 Vols): I Realism and the Aim of Science 1983; II The Open Universe: an Argument for Indeter-*

*minism 1982; III Quantum Theory and the Schism in Physics 1982*, ed. W. W. Bartley (Hutchinson, London).

20. K. R. Popper, in *Problems in the Philosophy of Science*, eds. I. Lakatos and A. Musgrave (North-Holland, Amsterdam, 1968), p. 163.
21. T. S. Kuhn, *The Structure of Scientific Revolutions* (2nd Edition, enlarged), (University of Chicago Press, Chicago, 1970).
22. P. K. Feyerabend, *Against Method: Outline of an Anarchistic Theory of Knowledge* (New Left Books, London, 1975).
23. P. K. Feyerabend, *Farewell to Reason* (Verso, London, 1987).
24. P. K. Feyerabend, quoted in Science **206**, 534 (1979).
25. J. S. Bell, Physics **1**, 195 (1964).
26. A. J. M. Garrett, Found. Phys. **20**, 1475 and **21**, 753 (1990).
27. A. J. M. Garrett, in *Maximum Entropy and Bayesian Methods, Seattle, Washington, 1991*, eds. G. J. Erickson, P. Neudorfer and C. R. Smith (Kluwer, Dordrecht, 1992), pp. 153–164.
28. F. Hoyle, Astrophys. J. Suppl. **1**, 121 (1954).
29. R. H. Dicke, Nature **192**, 440 (1961).
30. A. D. Linde, Int. J. Mod. Phys. **2A**, 561 (1987).
31. N. G. van Kampen, Physica A **153**, 97 (1988).
32. C. M. Patton and J. A. Wheeler, in *Quantum Gravity: an Oxford Symposium*, eds. C. J. Isham, R. Penrose and D. W. Sciama (Clarendon Press, Oxford, 1975), pp. 538–605.
33. H. Nielson, Acta Physica Polonica B **20**, 454 (1989).
34. J. Leslie, Philosophical Quarterly **40**, 65 (1990).
35. R. L. Oldershaw, Am. J. Phys. **56**, 1075 (1988).
36. W. H. Newton-Smith, *The Origin of the Universe*, preprint (1992).