Predicting future duration from present age: A critical assessment

Carlton M. Caves

Predicting future duration from present age: a critical assessment

Carlton M. Caves

*Using a temporal version of the Copernican principle, Gott has proposed a statistical predictor of future longevity based on present age (Gott III, J. R., 1993, Nature, **363,** 315) and applied the predictor to a variety of examples, including the longevity of the human species. Although Gott's proposal contains a grain of truth, it does not have the universal predictive power that he attributes to it.*

1. Introduction

Returning from a five-week residence at the Isaac Newton Institute this past summer, I found on my desk the 12 July issue of *The New Yorker*, containing a provocative story by the well known science writer Timothy Ferris [1]. The story, entitled 'How to Predict Everything', describes how J. Richard Gott, a Princeton astrophysicist, makes universal probabilistic predictions for a phenomenon's future duration based on knowing how long the phenomenon has lasted. The justification for Gott's rule is said to be a *temporal* version of the Copernican principle: when you observe a phenomenon in progress, your observation does not occur at a special time.

Here is Gott's account, as related to Ferris, of how he conceived his rule while contemplating the Berlin Wall.

Standing at the Wall in 1969, I made the following argument, using the Copernican principle. I said, Well, there's nothing special about the timing of my visit. I'm just travelling—you know, Europe on five dollars a day—and I'm observing the Wall because it happens to be here. My visit is random in time. So if I divide the Wall's total history, from the beginning to the end, into four quarters, and I'm located randomly somewhere in there, there's a fifty-per-cent chance that I'm in the middle two quarters—that means, not in the first quarter and not in the fourth quarter.

Let's suppose that I'm at the beginning of that middle fifty per cent. In that case, one quarter of the Wall's ultimate history has passed and there are three quarters left in the future. In that case, the future's three times as long as the past. On the other hand, if I'm at the other end, then three quarters have happened already, and there's one quarter left in the future. In that case, the future is one-third as long as the past.

…

(The Wall was) eight years (old in 1969). So I said to a friend, 'There's a fifty-per-cent chance that the Wall's future duration will be between (two and) two-thirds of a year and twenty-four years.' Twenty years later, in 1989, the Wall came down, within those two limits that I had predicted. I thought, Well, you know, maybe I should write this up.

Ferris goes on to recount how Gott applies his method to the longevity of the human species.

The question that Gott has been asking lately is how long the human species is going to last. Since scientists generally make predictions at the ninety-five-per-cent confidence level, Gott begins with the assumption that you and I, having no reason to think we've been born in a special time, are probably living during the middle ninety-five per cent of the ultimate duration of our species. In other words, we're probably living neither during the first two and a half per cent nor during the last two and a half per cent of all the time that human beings will have existed.

'*Homo sapiens* has been around for two hundred thousand years,' Gott said ···. 'That's how long our past is. Two and half per cent is equal to one-fortieth, so the future is probably at least one-thirty-ninth as long as the past but not more than thirty-nine times the past. If we divide two hundred thousand years by thirty-nine, we get about fifty-one hundred years. If we multiply it by thirty-nine, we get 7.8 million years. So if our location in human history is not special, there's a ninety-five-per-cent chance we're in the middle ninety-five per cent of it. Therefore the human future is probably going to last longer than fifty-one hundred years but less than 7.8 million years.

Now, those numbers are interesting, because they give us a total longevity that's comparable to that of other species.'

†*Author's address:* Center for Advanced Studies, Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131-1156, USA.

These glib predictions astonished me, not because Gott concludes from them that *homo sapiens* is unlikely to last longer than other species—that is a legitimate subject for inquiry and debate—but because they are put forward as a universal rule, applicable no matter what other information one has about the phenomenon in question. In making statistical predictions of future longevity, Gott dismisses the entire process of assembling and organizing information about a phenomenon, evaluating that information critically, and if possible, formulating laws that describe the phenomenon. Put succinctly, he rejects as irrelevant the process of rational, scientific inquiry, replacing it with a single, universal statistical rule. That has to be wrong.

I decided it was important to find the flaws in Gott's reasoning: flawed thinking is an inevitable, even necessary part of the scientific enterprise, but when it makes its way into *The New Yorker*, the time has come to find the flaws and draw attention to them. I began by requesting from the UNM Library a copy of the *Nature* article [2] where Gott proposes his rule and applies it to the above examples and others. A citation search turned up two other pieces in which Gott adds to the content of his *Nature* article: a Letter to *Nature* [3] responding to letters criticizing the original article and a chapter in the proceedings of an Astronomical Society of the Pacific (ASP) Symposium [4]. The present paper analyses what I found in Gott's papers and reports my conclusions.

2. Gott's delta-*t* argument

Gott justifies his probabilistic predictions by making what he calls the delta-*t* argument [2]. Suppose there is a phenomenon that has a beginning, or birth, at time t_0 and an end, or death, at time $t_0 + T$, T being the duration of the phenomenon. You observe the event at a time t between the beginning and the end, corresponding to a present age, $t_p = t - t_0$, and a future duration, $t_f = T - t_p$. If there is nothing special about the observation time—this is the content of the temporal Copernican principle—Gott reasons that $t_p = t - t_0$ is a random variable uniformly distributed between 0 and T. This means that t_p lies between aT and bT, $0 \leqslant a \leqslant b \leqslant 1$, with probability $b - a \equiv f$; in symbols, we write

$$P(aT < t_p < bT) = b - a = f . \tag{1}$$

Gott's next step is to infer from equation (1) that the duration T lies between the corresponding bounds, t_p/b and t_p/a, with the same probability f. Translated to future duration, this says that t_f lies between $(b^{-1} - 1)t_p$ and $(a^{-1} - 1)t_p$ with probability f, i.e.

$$G\left(\frac{1-b}{b} t_p < t_f < \frac{1-a}{a} t_p \right) = b - a = f . \tag{2}$$

All of Gott's predictions flow from this probability rule. I use the letter G to distinguish probabilities based on this rule.

Gott phrases his predictions in terms of particular $f \times 100\%$ confidence levels, which he obtains by letting a and b be equidistant from 0 and 1, i.e. $a = 1 - b$. The resulting choices, $a = \pm (1 - f)$ and $b = \pm (1 + f)$, lead to Gott's confidence-level prediction:

$$\frac{1-f}{1+f} t_p < t_f < \frac{1+f}{1-f} t_p \tag{3}$$
$$(f \times 100\% \text{ confidence level}) .$$

For example, in his encounter with the then ($t_p = $)8-year-old Berlin Wall, Gott used $f = 1/2$, with $a = 1/4$ and $b = 3/4$, which led him to predict with 50% confidence that the total duration of the Wall would lie between $4t_p/3 = 10\frac{2}{3}$ yr and $4t_p = 32$ yr or, equivalently, that the future duration would lie between $t_p/3 = 2\frac{2}{3}$ yr and $3t_p = 24$ yr. In most of his work, Gott uses a 95% confidence level, corresponding to $f = 0.95$.

Another form of Gott's rule arises from letting $b = 1$ and $a = (1 + Y)^{-1}$. Inserting these choices into equation (2), one finds that $t_f < Y t_p$ with probability $Y/(1 + Y)$; equivalently, the probability that the future duration is not less than $Y t_p$ is $(1 + Y)^{-1}$, i.e.

$$G(t_f \geq Y t_p) = \frac{1}{1 + Y} . \tag{4}$$

In his *Nature* article, Gott derives equation (4) independently of the delta-*t* argument by assuming that the phenomenon of interest is an exponential decay (see Gott's equation (6) and preceding discussion). There being no hint in the delta-*t* argument that Gott restricts his method to exponential decays, this derivation must be intended as an *example* of his method. I defer discussion of this derivation, since its status can be appreciated only after exposing and correcting the flaws in Gott's reasoning.

The delta-*t* argument implies that Gott's rule provides a universal method for predicting the future duration of any phenomenon, the only assumption being that the observation time is not special. Moreover, it is clear from the variety of phenomena to which Gott applies his rule—durations of the Berlin Wall, Stonehenge, and the Soviet Union, the publication lifetime of *Nature*, longevity of the human species, and in his ASP contribution and in his conversations with Ferris, running times of plays in New York—that he places no restrictions on the applicability of his rule.

It is not hard to find an error in the delta-*t* argument: *the step from equation (1) to Gott's rule (2) has no justification in probability theory*. This error has been pointed out by Buch in a letter to *Nature* criticizing Gott's method [5]. The total duration T (or the future duration t_f) is unknown and thus must be treated as a random variable described by a

prior probability distribution. This prior distribution expresses whatever information one possesses that can be used to make probabilistic statements about the phenomenon's duration. After collecting the data that the phenomenon's present age is t_p, the only procedure authorized by probability theory is to update the prior distribution to a new, posterior distribution for T (or t_f), which reflects both the prior information and the present age. The formal procedure for this updating is called Bayes's theorem [6].

The error just identified is suÉcient to invalidate the delta-t argument. To correct it requires an analysis that uses Bayes's theorem to update probabilities. Indeed, Gott has endorsed [3,4] a Bayesian analysis suggested by Buch [5]; this Bayesian analysis, said to be based on the temporal Copernican principle, leads to Gott's rule, provided one uses a particular prior distribution, dT/T, called the Jeffreys prior [7]. The reader should be aware, however, that the Bayesian analysis suggested by Buch and endorsed by Gott is also flawed. In considering the Buch–Gott Bayesian analysis below, we will uncover this flaw, thus revealing a second error in the delta-t argument, just as serious as the first, but more insidious because it is more subtle: *equation (1) is an incorrect mathematical formulation of the temporal Copernican principle.* The pay-off for identifying this second flaw is that it clarifies the meaning and status of the temporal Copernican principle. In developing a proper Bayesian analysis based on the temporal Copernican principle, we will discover that Gott's rule is a universal consequence of the Copernican principle, *in the situation where one knows the phenomenon to be in progress, but does not know its present age.* Not knowing the present age, one cannot make Gott's predictions of future duration.

Before turning to the Bayesian analysis, however, I introduce a few examples that show that Gott's rule cannot be a universal predictor and also serve to put some flesh on the dry bones of the subsequent Bayesian analysis.

3. Examples of using Gott's rule

I advise my students to test the solution to a homework problem by considering special cases where the solution is already known. This common-sense technique, a good rule in scientific thinking and in everyday life, provides compelling evidence that Gott's predictions cannot have the universal validity that he attributes to them.

Exponential decay. Consider an atom that is excited to a metastable energy level at some unknown time and then decays exponentially to the ground state with a decay constant $\tau^{-1} = (20 \text{ min})^{-1}$. You come along at time t and are told that the atom is in the metastable level, having been excited a time $t_p = 15$ min ago. According to Gott, you can predict with 95% confidence that the

decay will occur between $t_f = t_p/39 = 23.1$ s and $t_f = 39t_p = 9.75$ h into the future; more telling is that equation (4) predicts that $t_f \geqslant 4t_p = 60.0$ min with probability 1/5. These predictions contradict the defining property of an exponential decay: being informed that the atom is still in the excited state at time t simply resets the clock so that your expectations for its future decay are the same as though it had been initially excited at time t. Specifically, you predict that it will survive a further time t_f without decaying with probability $\exp(-t_f/\tau)$, corresponding to 95% confidence of decay between $t_f = 0$ and $t_f = \tau \ln 20 = 3.00\tau = 59.9$ min. Although the numerical discrepancies between Gott's predictions and the predictions of an exponential decay are important, they are only a symptom of the real problem: Gott's rule, by including present age in the prediction of future duration, is inconsistent with the very notion of an exponential decay.

Buch [5] has pointed out that Gott's rule is inconsistent with the properties of an exponential decay. In his reply to Buch [3] and in his ASP contribution [4], Gott admits that his method does not apply to an exponential decay whose decay constant is known. Instead, he says that it applies to an exponential decay whose decay constant is unknown and distributed according to the Jeffreys prior $d\tau/\tau$; this leads to the Jeffreys prior dT/T for total duration T and is not an exponential decay at all. Gott [3] also reasserts his exponential-decay derivation of equation (4), to be discussed below. All this leaves one thoroughly confused—does Gott regard his rule as universal or not?—but his subsequent conversations with Ferris [1] make clear that he does not acknowledge any restrictions on the use of his rule.

Longevity of an individual. Suppose you are going to a meeting of your book club, to be held at a member's house that you've never been to before. You find the right street, but having forgotten the street address, you choose between two houses where there is evident activity. Knocking at one, you are told that the activity within is a birthday party, not a book-club meeting. Your friendly enquiry about the age of the celebrant elicits the reply that she is celebrating her ($t_p =$) 50th birthday. According to Gott, you can predict with 95% confidence that the woman will survive between $t_p/39 = 1.28$ yr and $39t_p = 1950$ yr into the future. Since the wide range encompasses reasonable expectations regarding the woman's survival, it might not seem so bad, until one realizes that equation (4) predicts that with probability 1/2 the woman will survive beyond 100 years old and with probability 1/3 beyond 150. Few of us would want to bet on the woman's survival using Gott's rule.

One might object at this point that Gott probably did not intend his rule to apply to an individual's longevity, but in his ASP contribution [4], Gott applies the rule to himself: 'At the time my (*Nature*) paper was published on May 27, 1993, I was 46.3 yr old, so the 95% delta-*t* argument predicted that my future longevity would be at least 1.2 yr but less than 1806 yr. I have survived past the lower limit already and so if I do not make it past the upper limit, then that prediction will indeed prove correct for me!'

Deterministic phenomena. The best testing ground for ideas comes from extreme cases, and here the most extreme case is a deterministic phenomenon. Putting the example in a dramatic context, suppose you are captured by terrorists, who confine you to a small room. You are told that at some time in the next 24 h, a timer will be set and that after it has ticked for 30 min, poison gas will fill the room, killing you. You are then drugged and wake up to find the timer ticking and reading 20 min since being set. According to Gott, you can predict with 95% confidence that the time to release of the gas lies between $t_f = 20$ min/39 = 30.8 s and $t_f = 39 \times 20$ min = 13.0 h; even worse, equation (4) predicts that with probability 2/3 the time to release is 10 min or more. These reassuring predictions provide scant comfort, since you *know* you have 10 min to live.

These examples demonstrate that Gott's rule cannot be a universal method for predicting future durations. If the rule has any validity, it must involve other information than the present age of a phenomenon. As E. T. Jaynes taught us [8], when probabilistic predictions violate one's intuition, the proper response is neither to accept the nonintuitive predictions without question nor to dismiss them out of hand, but rather to identify the information underlying the prediction. You will either find the information inapplicable to the situation at hand, thereby confirming your intuition and allowing you to discard the predictions, or you will sharpen your intuition.

The objective of this paper is to identify the prior information that underlies Gott's rule. The tool is Bayesian analysis. We will discover that the temporal Copernican principle contains a grain of truth, but that grain of truth does not include Gott's predictions of future duration.

4. A flawed, but instructive Bayesian analysis

Return to the general situation introduced above, that of a phenomenon with a birth time t_0 and a duration T. You observe the phenomenon at time t. It is often useful to replace one or both quantities, t_0 and T, by the present age, $t_p = t - t_0$, and the future duration, $t_f = T - t_p$. In developing the Bayesian analysis, I first formulate and analyse a flawed

approach, advanced by Buch [5] and endorsed by Gott [3], which is modelled on Gott's delta-*t* argument. For this purpose, it is most convenient to use t_p and T as the primary variables. The reason for going through this flawed analysis is that it turns up the second error in Gott's delta-*t* argument.

Your prior information about the phenomenon is expressed in a prior probability $p(t_p, T) dt_p dT$, the joint probability that the phenomenon has lasted a time between t_p and $t_p + dt_p$ at the time of observation and that the phenomenon will last a total time between T and $T + dT$. The joint probability density can be written as $p(t_p, T) = p(t_p|T) w(T)$, where $p(t_p|T)$ is the *conditional* probability density for the present age, given a total duration T, and $w(T)$ is your prior probability density for the total duration. Throughout I use upper-case letters for probabilities and lower-case letters for probability densities.

Before going further, it is useful to introduce two quantities related to $w(T)$: $\lambda(T)$ is the death rate—i.e. $\lambda(T) dT$ is the probability that the phenomenon, having lasted a time T, ends in the next dT—and $Q(T)$ is the survival probability—the probability that the phenomenon lasts at least a time T. These quantities are related by

$$w(T) = -\frac{dQ}{dT} = Q(T)\lambda(T) \tag{5}$$

or, equivalently, by

$$Q(T) = \int_T^\infty dT' \, w(T') = \exp\left(-\int_0^T dT' \, \lambda(T')\right). \tag{6}$$

An exponential decay is characterized by a constant death rate, $\lambda(T) = \lambda_0$, in which case $Q(T) = \exp(-\lambda_0 T)$ and $w(T) = \lambda_0 \exp(-\lambda_0 T)$.

Gott's formulation of the temporal Copernican principle is the following: if there is nothing special about the observation time, the present age is a random variable uniformly distributed between 0 and T, i.e.

$$p(t_p|T) dt_p = dt_p/T, \quad 0 \le t_p \le T. \tag{7}$$

This is the probability-density version of equation (1). We now use Bayes's theorem,

$$P(X|Y)P(Y) = P(X, Y) = P(Y|X)P(X), \tag{8}$$

to find your posterior probability density for the total duration, given the present age:

$$p(T|t_p) = \frac{p(t_p|T)w(T)}{p(t_p)} = \begin{cases} 0, & T < t_p, \\ w(T)/T \, p(t_p), & T \ge t_p. \end{cases} \tag{9}$$

The unconditional probability density $p(t_p)$ for the present age, which is a normalization constant in this expression, is given by

$$p(t_p) = \int_{t_p}^\infty dT \, \frac{w(T)}{T}. \tag{10}$$

One can easily verify that Gott's rule, embodied in equations (2)–(4), is equivalent to a posterior density

$$p(T|t_{\mathrm{p}}) = \begin{cases} 0, & T < t_{\mathrm{p}}, \\ t_{\mathrm{p}}/T^2, & T \geq t_{\mathrm{p}}. \end{cases} \quad (11)$$

To get this posterior from the present analysis, one must assume the (unnormalizable) Jeffreys prior,

$$w(T) = \frac{1}{T}. \quad (12)$$

Buch [5] concludes that Gott's rule is unreasonable because it corresponds to an unnormalizable prior density. Gott [3] replies (correctly, I think) that there is nothing wrong with an unnormalizable prior, since the posterior density for T can be normalized. He defends the Jeffreys prior as being the appropriate 'vague Bayesian prior' to use in a situation where one initially knows nothing about the magnitude of the duration [3,4].

Jaynes [9] has delineated the conditions for using the Jeffreys prior, showing that it should be used when one's prior information is unchanged by a rescaling of the total duration, $T' = \alpha T$ ($\alpha > 0$). If one's prior information is unchanged by the rescaling, then the density for T', $w'(T') = w(T') \, \mathrm{d}T/\mathrm{d}T' = w(T)/\alpha$, should have the same functional form as the original density, i.e. $w'(T') = w(T')$ This gives $= w(\alpha T) = w(T)/\alpha$, which implies that $w(T) \propto 1/T$. This might seem to be progress in identifying the information that underlies Gott's rule—use it when one has no prior information about time scales associated with the phenomenon—but it turns out not to be, because the present Bayesian analysis is wrong. The reason for presenting it is not to consider its consequences, but to identify where it goes wrong.

5. A straightforward Bayesian analysis

That something is wrong is made apparent by a different analysis of the same situation, this time a straightforward Bayesian analysis that does not invoke the Copernican principle. Your prior information about the total duration is expressed in the prior density $w(T)$. You observe the phenomenon still to be in progress at a time t_{p} after its beginning. The conditional probability for this observation, given a total duration T, is 0 if $t_{\mathrm{p}} > T$ and 1 if $t_{\mathrm{p}} \leq T$. Thus Bayes's theorem implies, with O denoting the observation,

$$p(T|O) = \frac{P(O|T)w(T)}{P(O)} = \begin{cases} 0, & T < t_{\mathrm{p}}, \\ w(T)/Q(t_{\mathrm{p}}), & T \geq t_{\mathrm{p}}. \end{cases} \quad (13)$$

Here the normalization constant is the survival probability, i.e. $P(O) = Q(t_{\mathrm{p}})$.

The posterior density (13) is so eminently reasonable that one could have written it down without using the formal apparatus of Bayes's theorem. It says that the effect of discovering the present age is to rule out durations shorter than the present age; your posterior expectations for durations longer than the present age are the same as your prior expectations, with appropriate renormalization. Notice that this inference updates sensibly: subsequent observations that find the phenomenon still in progress simply exclude wider interval of durations. Yet putting this simple inference in the context of the Copernican principle apparently yields a different posterior density (9) for the total duration. How can that be? There is nothing wrong with the Bayesian inference in either analysis, so the culprit must be Gott's formulation of the temporal Copernican principle. Thus we arrive at the second error in the delta-T argument: *the uniform density (7) for t_{p}—and, by extension, equation (1)—is not the correct mathematical formulation of the temporal Copernican principle.*

Where the uniform density goes wrong is in *assuming* that your observation occurs while the phenomenon is in progress. If your observation does not occur at a special time, then it is very likely that it occurs before the phenomenon begins or after it has ended. Including these other possibilities leads to a proper Bayesian formulation of the temporal Copernican principle, which is consistent with the inference expressed in equation (13).

6. A proper Bayesian analysis of the temporal Copernican principle

In formulating a proper Bayesian analysis, it is convenient to choose the birth time t_0 and the total duration T as the primary variables. Your prior knowledge about these two quantities is incorporated in two probability densities: (i) $\gamma(t_0)$ gives the probability $\gamma(t_0)\mathrm{d}t_0$ that the phenomenon begins between times t_0 and $t_0 + \mathrm{d}t_0$; (ii) $p(T|t_0)$ gives the probability $p(T|t_0)\mathrm{d}T$ that the phenomenon lasts a time between T and $T + \mathrm{d}T$, given that it began at time t_0. The corresponding joint probability density is $p(t_0, T) = p(T|t_0)\gamma(t_0)$.

The temporal Copernican principle—that your observation does not take place at a special time—is a time-translation symmetry that restricts the form of the prior densities [9]. To say that your observation time is not special is to say that your prior information is unchanged if the entire phenomenon is displaced in time while your observation time remains fixed. To be consistent with this translation symmetry, your prior probability density should be unchanged by such a time translation; i.e. $p(t_0, T)$ should be independent of the birth time t_0. Thus the temporal Copernican principle can be captured precisely in the following two statements.

(1) The phenomenon is equally likely to begin at any time. This means that $\gamma(t_0)$ is a constant. In order to work with normalizable probabilities, I replace the exact symmetry with the approximate one that

$\gamma(t_0)$ has a constant value, $1/\Delta$, at all times within a very long time interval. The duration Δ of this very long time interval exceeds all other times relevant to the problem, particularly typical durations.

(2) Probabilities for total duration are independent of birth time. This means that the conditional probability density $p(T|t_0)$ does not depend on t_0 and can be written as $p(T|t_0) = w(T)$, where $w(T)$ is the probability density introduced above.

Should you be dissatisfied with these restrictions on the prior probabilities, it means that you do not accept the temporal Copernican principle as applying to your prior information. Dissatisfaction should not be surprising, for one would not expect the Copernican principle to apply to all situations. The three examples introduced earlier illustrate considerations that arise in using the temporal Copernican principle. In all three examples, it is easy to accept that ignorance of the birth time is described by the time-translation symmetry of the temporal Copernican principle: the atom can be excited at any time during an interval much longer than the decay time; for the woman at the birthday party, the situation could be phrased in terms of an individual whose birth could occur at any time over a period much longer than a typical human lifetime; the timer can be set at any time within a 24 h period, a period somewhat longer than the 30 min that the timer ticks. Moreover, in the cases of the atom and the poison gas, duration probabilities are independent of the birth time. In contrast, in the case of the longevity of an individual, the prior conditional probability for the individual's lifetime would depend on the time of birth. Your prior expectation for the longevity of an individual born, say in Britain, would depend on whether the individual was born in the second half of the 20th Century, at the beginning of the 19th Century, or 10 000 years ago, at the end of the last Ice Age.

At time t you make your observation. In Gott's formulation the observation yields the present age, but we now understand that getting the present age presupposes that your observation finds the phenomenon in progress. The first result of the observation is simply to determine whether the phenomenon has not yet begun, is already over, or is in progress. Only the last of these possibilities, denoted by I for 'in progress', is of interest to us. The conditional probability to find the phenomenon in progress, given a birth time t_0 and a duration T, is

$$P(I|t_0, T) = \begin{cases} 1, & t_0 \le t \le t_0 + T, \\ 0, & \text{otherwise}. \end{cases} \quad (14)$$

The unconditional probability to find the phenomenon in progress is given by

$$\begin{aligned} P(I) &= \int_{-\infty}^{\infty} dt_0 \int_0^{\infty} dT\, P(I|t_0, T)\gamma(t_0)w(T) \\ &= \int_{-\infty}^{t} dt_0\, \gamma(t_0) \underbrace{\int_{t-t_0}^{\infty} dT\, w(T)}_{= Q(t-t_0)} \\ &= \int_0^{\infty} dt'\, \gamma(t-t')Q(t'). \end{aligned} \quad (15)$$

The assumption that $\gamma(t_0)$ is constant for all times of interest means that $\gamma(t-t') = 1/\Delta$ for all times t' such that the survival probability $Q(t')$ is significantly different from zero. This allows us to put $P(I)$ in the form

$$P(I) = \overline{T}/\Delta, \quad (16)$$

where

$$\overline{T} = \int_0^{\infty} dT\, Tw(T) = \int_0^{\infty} dT\, Q(T) \quad (17)$$

is the mean total duration with respect to the prior density $w(T)$. The present analysis assumes that $\overline{T}$ is finite, which requires, for large durations T, that $Q(T)$ go to zero faster than $1/T$ or, equivalently, that $w(T)$ go to zero faster than $1/T^2$. For an exponential decay, $\overline{T}^{-1} = \lambda_0$ is the decay constant. Notice that the probability to find the phenomenon in progress is very small.

Bayes's theorem gives the posterior probability density for t_0 and T, given that the phenomenon is occurring:

$$\begin{aligned} p(t_0, T|I) &= \frac{P(I|t_0, T)\gamma(t_0)w(T)}{P(I)} \\ &= \begin{cases} \gamma(t_0)w(T)/P(I), & t_0 \le t \le t_0 + T, \\ 0, & t_0 > t \text{ or } t_0 + T < t. \end{cases} \end{aligned} \quad (18)$$

If $t - t_0$ is large enough in this expression that $\gamma(t_0)$ does not have its constant value, then $T \ge t - t_0$ is so large that $w(T)$ is negligible. Thus we can again replace $\gamma(t_0)$ by the constant value $1/\Delta$, leaving

$$p(t_0, T|I) = \begin{cases} w(T)/\overline{T}, & t_0 \le t \le t_0 + T, \\ 0, & t_0 > t \text{ or } t_0 + T < t. \end{cases} \quad (19)$$

It is instructive to consider equation (21) from a variety of perspectives. A first question asks how the probability density for total duration changes on learning that the phenomenon is occurring:

$$p(T|I) = \int_{-\infty}^{\infty} dt_0\, p(t_0, T|I) = \frac{Tw(T)}{\overline{T}}. \quad (20)$$

Notice that $p(T|I)$ is biased toward longer durations than the prior density $w(T)$. This is because the phenomenon is very unlikely to be in progress at a random time selected from the long time interval Δ, so finding it in progress

prejudices you to think that it has a longer duration than your original expectations.

A useful, equivalent form for equation (21) comes from changing variables to present age and future duration. The Jacobian of the transformation from (t_0, T) to (t_p, t_f) is -1, which implies that $dt_0 \, dT = dt_p \, dt_f$. Hence, the probability density for present age and future duration, given that the phenomenon is occurring, is

$$p(t_p, t_f|I) = p(t_0, T|I) = \begin{cases} w(t_p + t_f)/\overline{T} \,, & t_p \geq 0 \text{ and } t_f \geq 0 \,, \\ 0 \,, & \text{otherwise} \,. \end{cases} \quad (21)$$

Knowing the phenomenon is in progress is equivalent to saying that both the present age and future duration are non-negative, so we can regard that condition as implicit and omit it from subsequent expressions. The content of equation (21) is the following: if you know the phenomenon is in progress, but do not know its present age, you treat uniformly the split of total duration into past and future; more precisely, you assign the same probability, governed by $w(T)$, to all ways of splitting T into past and future. That is the temporal Copernican principle. Indeed, *equation (21) is the mathematical embodiment of the temporal Copernican principle for phenomena known to be in progress.*

Equation (21) has three immediate consequences that highlight the connection between the Copernican principle and Gott's rule. We proceed by noting that once the phenomenon is known to be in progress, the total duration is the sum of the present age and the future duration, i.e. $p(T|t_p, t_f, I) = \delta(T - t_p - t_f)$. Another application of Bayes's theorem then gives

$$p(t_p, t_f|T, I) = \frac{p(T|t_p, t_f, I)p(t_p, t_f|I)}{p(T|I)} = \frac{1}{T}\delta(T - t_p - t_f) \,. \quad (22)$$

This is a conditional version of equation (21), with the same content.

An obvious consequence is that if you know the phenomenon is in progress and also know its total duration, then you conclude that the present age is uniformly distributed between 0 and T:

$$p(t_p|T, I) = \int_0^\infty dt_f \, p(t_p, t_f|T, I) = \begin{cases} 1/T \,, & t_p \leq T \,, \\ 0 \,, & t_p > T \,. \end{cases} \quad (23)$$

This is the precise statement of what Gott is trying to capture in his initial assumption (1) about the present age. The starting point (7) of the flawed Bayesian analysis also asserts that t_p is uniformly distributed between 0 and T, but it is different from equation (23) in a subtle, but crucial way: because $p(t_p|T, I)$ is conditioned on knowing the phenomenon is occurring, further statistical inference uses the conditional density $p(T|I) = Tw(T)/\overline{T}$, instead of the prior density $w(T)$; we find below (see equation (28)) that

this is how the present Bayesian analysis comes into agreement with the straightforward inference of the preceding section.

A second obvious, but important consequence of equation (21) is that

$$P\left(\frac{1-b}{b}t_p < t_f < \frac{1-a}{a}t_p \,\middle|\, T, I\right)$$
$$= \int_0^T dt_p \int_{(b^{-1}-1)t_p}^{(a^{-1}-1)t_p} dt_f \, p(t_p, t_f|T, I) \quad (24)$$
$$= \int_{aT}^{bT} \frac{dt_p}{T} = b - a \,.$$

Since the condition on future duration in the probability on the left is equivalent to $aT < t_p < bT$, this is just the statement that in dividing the total duration into past and future, possibilities satisfying the condition are a fraction $b - a$ of all the possibilities. Furthermore, since the conditional probability (24) is independent of T, the same result holds no matter what the prior density for T:

$$P\left(\frac{1-b}{b}t_p < t_f < \frac{1-a}{a}t_p \,\middle|\, I\right)$$
$$= \int_0^\infty dT \, P\left(\frac{1-b}{b}t_p < t_f < \frac{1-a}{a}t_p \,\middle|\, T, I\right) \quad (25)$$
$$p(T|I) = b - a \,.$$

Setting $b = 1$ and $a = (1+Y)^{-1}$ in this result yields the third consequence of equation (21):

$$P(t_f \geq Yt_p|I) = \frac{1}{1+Y} \,. \quad (26)$$

Equations (25) and (26) are precise statements of Gott's rule in forms (2) and (4). Indeed, they look just like Gott's rule, with the crucial difference that they are conditioned on knowing the phenomenon is in progress, without knowing its present age.

We are now in the curious position of affirming that *for a phenomenon known to be in progress, but whose present age is unknown, the temporal Copernican principle leads to universal statistical predictions, which are described by Gott's rule.* Indeed, all the manipulations in Gott's delta-*t* argument are valid in this situation. The down side for Gott is that this conclusion does not authorize his predictions: *in these circumstances, Gott's rule has no power to predict future durations from present ages, for the simple reason that the present age is unknown.*

The results of the present Bayesian analysis make perfect sense in the three examples introduced above. Oddly enough, the deterministic example is the simplest: if you find the timer ticking, *but there is nothing to indicate how long it has been*

ticking, it is reasonable to assign the same probability to all ways of dividing the 30 min interval into past and future.

The case of the women's longevity is a bit more complicated. If you encounter an individual, *but are given no clue as to the individual's age*, a first cut might treat the past–future split uniformly. For a person born in Britain, a more careful analysis would give greater weight to the future than to the past, because of the increase in life expectancy in this century. I have already indicated that the case of an individual's longevity does not fit into the temporal Copernican principle for just this reason. It should be emphasized that the problem is not the use of Bayesian analysis: the increase in life expectancy could be incorporated into a more complicated Bayesian analysis, which would automatically produce a bias toward the future.

The case of the atom is particularly interesting because of Gott's exponential-decay derivation of equation (4). If you find the atom in the excited state, *but you are not told when it was excited*, it is reasonable to assign the same probability to all ways of splitting a particular duration T into past and future and to weight the result by an exponential $\exp(-\lambda_0 T) = \exp[-\lambda_0(t_p + t_f)]$, which expresses the probability for duration T. The final result, properly normalized, is the probability density (21) specialized to an exponential decay:

$$p(t_p, t_f|I) = \lambda_0^2 \exp[-\lambda_0(t_p + t_f)].\qquad(27)$$

This allows us to understand Gott's exponential-decay derivation of equation (4) (see Gott's equation (6)): he starts with equation (27) (see Gott's equations (3) and (4)), from which he immediately derives equation (26), all without realizing that equation (27) applies to an exponential decay whose present age is unknown. Having gone through a proper Bayesian analysis, we now understand that equation (26) does not depend at all on assuming an exponential decay, but rather is a universal consequence of the temporal Copernican principle, valid no matter what the prior density $w(T)$, provided the present age is unknown.

The next task is to find out what happens if you do discover the present age. When you determine the present age of the phenomenon, your Bayesian posterior for the total duration is given by

$$p(T|t_p, I) = \frac{p(t_p|T,I)p(T|I)}{p(t_p|I)} = \begin{cases} 0, & T < t_p, \\ w(T)/Q(t_p), & T \geq t_p, \end{cases}\qquad(28)$$

where

$$p(t_p|I) = \int_0^\infty dT\, p(t_p|T,I)p(T|I) = Q(t_p)/\overline{T}.\qquad(29)$$

This posterior density is identical to the one that emerged from the straightforward Bayesian analysis that wholly ignored the Copernican principle. This is as it should be, because in the language of this section, the straightforward

Bayesian inference corresponds to first learning the birth time t_o and then discovering that the phenomenon has survived at time t_p, a situation that is equivalent to first learning that the phenomenon is in progress and then discovering its present age. Once you are informed of the present age or, equivalently, of the birth time, you *are* at a special time, the time t_p since the phenomenon began. The temporal Copernican principle becomes irrelevant. It just gets in the way of the obvious inference expressed in equation (13).

At this point it is profitable to re-read Gott's account of his 1969 encounter with the Berlin Wall. If Gott had not known when the Wall was built, the logic of the first two paragraphs of his account would be impeccable. Under those circumstances, it would be reasonable to assign probability 1/2 to the encounter's occurring during the middle two quarters of the Wall's total history. Since he did know that the Wall was built in 1961, however, his encounter did occur at a special time, the time eight years after the Wall's construction. The predictions made in the third paragraph of his account do not follow from the argument in the first two paragraphs. Indeed, his posterior expectations for the Wall's duration should have been a renormalized version of his prior expectations, whatever those were, with durations up to eight years excluded.

We can now give a succinct account of how Gott's delta-t argument goes awry: the first two steps are wrong. The step from equation (1) to Gott's rule (2) is a non-Bayesian inference having no justification in probability theory; just as important, equation (1) is itself an incorrect expression of the temporal Copernican principle, because it assumes that an observation at a random time will find the very unlikely result that the phenomenon is in progress. In repairing these errors, we discovered that Gott's rule for relating future duration to present age is indeed a universal consequence of the temporal Copernican principle, but only in a situation—not knowing the present age—which leaves the rule shorn of predictive power. Gott's predictions require knowing how long a phenomenon has lasted, but once you obtain this information, the temporal Copernican principle no longer has any impact, because you are at a special time within the lifetime of the phenomenon.

7. Gott's rule as a predictor

All of Gott's predictions—from the future duration of the Berlin Wall to the longevity of the human species—are now detached from their original mooring in the temporal Copernican principle and left to float free of justification. Yet a flawed analysis might lead to reasonable predictions. There might be some justification for Gott's predictions other than the Copernican principle. Both the straightforward Bayesian analysis and the analysis based on the Copernican principle culminate in the same inference (equations (13) and (28)): once you know the present age,

your expectations about total duration are the same as your prior expectations, except that durations shorter than the present age are excluded. Thus all questions about the applicability of Gott's predictions reduce to determining what prior density underlies his predictions.

As noted above, Gott's rule follows from a posterior density

$$p(T|O) = \begin{cases} 0, & T < t_{\mathrm{p}}, \\ t_{\mathrm{p}}/T^2, & T \geq t_{\mathrm{p}}. \end{cases} \tag{30}$$

Within the correct Bayesian analysis, this posterior comes from an unnormalizable prior density

$$w_{\mathrm{g}}(T) = \frac{1}{T^2}. \tag{31}$$

This prior density, distinguished by a subscript g, corresponds to a survival probability $Q_{\mathrm{g}}(T) = 1/T$ and to a death rate $\lambda_{\mathrm{g}}(T) = 1/T$. One way to characterize $w_{\mathrm{g}}(T)$ is that the characteristic time associated with the death rate, $\lambda_{\mathrm{g}}^{-1}(T)$, is always the same as the age T.

The prior density $w_{\mathrm{g}}(T)$ is different from the Jeffreys prior that Gott [3,4] identifies with his predictions, the reason being that Gott uses the flawed Bayesian analysis given above. Yet within the Bayesian analysis using the temporal Copernican principle, $w_{\mathrm{g}}(T)$ has a scale-free status similar to that found by Jaynes [9] for the Jeffreys prior. Suppose that once you know the phenomenon is in progress, anything else you know, coming from the prior information about T, is unchanged by a simultaneous change in the scale of the past and the future. Under such a scale change, $t_{\mathrm{p}}' = \alpha t_{\mathrm{p}}$ and $t_{\mathrm{f}}' = \alpha t_{\mathrm{f}}$, the new and old probability densities are related by

$$p'(t_{\mathrm{p}}', t_{\mathrm{f}}'|I) = p(t_{\mathrm{p}}, t_{\mathrm{f}}|I)\, \mathrm{d}t_{\mathrm{p}}\, \mathrm{d}t_{\mathrm{f}}/\mathrm{d}t_{\mathrm{p}}'\, \mathrm{d}t_{\mathrm{f}}' = p(t_{\mathrm{p}}, t_{\mathrm{f}}|I)/\alpha^2. \tag{32}$$

To say that all your information is unchanged by this scale change is to say that the old and new densities should have the same functional form, i.e. $p'(t_{\mathrm{p}}', t_{\mathrm{f}}') = p(t_{\mathrm{p}}', t_{\mathrm{f}}')$, which implies that $p(\alpha t_{\mathrm{p}}, \alpha t_{\mathrm{f}}|I) = p(t_{\mathrm{p}}, t_{\mathrm{f}}|I)/\alpha^2$. Using equation (21) to write this in terms of the prior density, one finds that $w(\alpha T) = w(T)/\alpha^2$, which implies that the prior density has the form (31).

As discussed above, the Jeffreys prior applies when your prior information about the duration, before any observation, is scale-invariant. Once you know the phenomenon is in progress, however, $w_{\mathrm{g}}(T)$ captures the notion of scale invariance, because it corresponds to invariance of $p(t_{\mathrm{p}}, t_{\mathrm{f}}|I)$ under simultaneous rescaling of the past and future. In contrast, the Jeffreys prior corresponds to invariance of $p(t_{\mathrm{p}}, t_{\mathrm{f}}|I)$ under rescaling of t_{p} or t_{f}, but not both simultaneously.

We have now uncovered the prior information that underlies the use of Gott's rule as a predictor of future duration; namely, knowing that a phenomenon is in progress, you cannot identify any time scales associated

with the phenomenon either into the past or into the future. One way of thinking about this is that for a phenomenon that has no time scales, discovering the present age does *not* put you at a special time in the phenomenon's history, so some consequences of the temporal Copernican principle survive. Whether the scale-free prior information is appropriate must be judged case by case; it is not a universal rule. The scale-free prior certainly does not apply to the three examples introduced in this article, each of which has an obvious time scale: for the atom, the scale is the decay time; for an individual, the scale is a typical human lifetime; for the deterministic phenomenon, the scale is the 30 min that the timer ticks. Ignoring these time scales is the reason that Gott's rule leads to absurd predictions for these examples.

The examples Gott discusses at the beginning of his *Nature* article all have readily identifiable time scales that make application of Gott's rule problematic. The survival of a human institution—a political institution such as the government of the former Soviet Union or a cultural institution such as a periodical like *Nature*—is influenced by the 30 yr time scale of a generation or by a typical human lifetime, since loyalty to and management of such institutions change on these time scales. Physical manifestations of human institutions, such as the Berlin Wall or Stonehenge, are influenced by these same human time scales and, in addition, by the time scale over which erosion leads to disintegration.

8. The success of Gott's rule

Even though there is little reason to adopt Gott's rule, he portrays his predictions as successful [1–4]. Consider, for example, his 95%-confidence prediction that *Nature*, given its 123 yr history of publication in 1993, would continue to publish for a period between 3.15 years and 4800 years. Gott would consider this prediction successful because *Nature* has already surpassed the lower bound and is very unlikely to exceed the upper bound. Yet there's the hitch: the upper bound is far too large; without doing any analysis, anyone could have written down a similar very large 95% confidence interval and achieved the same 'success'. To assess Gott's rule, one should direct attention not at the the 95% confidence predictions, but at the high probabilities the rule assigns to very long future durations. Gott's rule in form (4) predicts that with probability 1/2, *Nature* will continue to publish for more than 123 years after 1993, with probability 1/5 for more than 492 years, with probability 1/10 for more than 1107 years, and with probability 1/20 for more than 2337 years. These probabilities posit a great deal of faith in the durability of human institutions.

To make this point more quantitatively, it is useful to consider a particular form of the probability that future duration exceeds some multiple of present age:

$$P(t_f \geq Yt_p|O) = P\left(T \geq (1+Y)t_p \mid O\right)$$
$$= \int_{(1+Y)t_p}^{\infty} dT\, p(T|O) = \frac{Q[(1+Y)t_p]}{Q(t_p)} \quad (33)$$
$$= \exp\left(-\int_{t_p}^{(1+Y)t_p} dT\, \lambda(T)\right).$$

This form makes clear that $P(t_f \geq Yt_p|O)$ depends only on the death rate during the interval between the present age and the lower bound for longevity. For a death rate $\lambda_g(T) = 1/T$, one gets Gott's rule.

Now let us apply this to the example of a periodical like *Nature*. At start-up a new publication confronts a variety of short-term, rapid-death scenarios. Should it survive these initial hazards and become established like *Nature*, the next time scale it faces might be roughly a human lifetime. If this time scale is modelled by a constant death rate $\tau^{-1} = (60 \text{ yr})^{-1}$, then one finds from equation (33) that $P(t_f \geq Yt_p|O) = \exp(-Yt_p/\tau)$. For *Nature*, this gives predictions quite different from Gott's: for example, a probability 0.129 to continue publishing for more than 123 years beyond 1993 and a probability 2.75×10^{-4} to continue publishing for more than 492 years.

Should these predictions seem unduly pessimistic, it is because the constant decay rate does not recognize a long publication record as providing evidence for continued success. A prejudice that success begets success can be incorporated, without discarding the time scale, by choosing, for example,

$$\lambda(T) = \frac{1}{T}\left(\frac{T}{\tau}\right)^{\beta}, \quad (34)$$

where $\beta \geq 0$. For $\beta = 1$, this gives a constant death rate τ^{-1}, and for $\beta = 0$, it gives Gott's rule. For intermediate values, it gives a death rate that decreases with age, but with the time scale τ still having an effect. The resulting probability (33) is

$$P(t_f \geq Yt_p|O) = \exp\left[-\left(\frac{t_p}{\tau}\right)^{\beta}\frac{(1+Y)^{\beta}-1}{\beta}\right]. \quad (35)$$

For *Nature* this gives, assuming $\beta = \pm$, a probability 0.305 to continue publishing for more than 123 years beyond 1993, a probability 2.90×10^{-2} for more than 492 years, and a probability 2.05×10^{-3} for more than 1107 years. The point here is not the particular values nor even the death-rate model, but rather that there is one or more time scales, which can and should be incorporated in the prior distribution.

Gott stresses the success of his predictions [1,4] for the 44 Broadway and off-Broadway plays listed in *The New Yorker* on 27 May 1993, the day his original *Nature* article was published. For example, Gott's 95%-confidence rule predicted that *Cats*, having played for 3885 days, would continue to play for a period between 100 days and 415 years. Gott regards this prediction as a success because

the production continues today, thereby surpassing the lower bound, and is unlikely to exceed the upper bound [1]. Yet since *Cats* had run 6263 days through 30 November 1999, when I determined that it was still running, the same rule predicts that with probability 1/5, it will continue to run for at least another 68.6 years, with probability 1/10 for at least another 154 years, and with probability 1/20 for another 326 years. Such predictions ignore obvious time scales. A new production faces a variety of short- to medium-term scales, including the time to the first reviews, the time over which a producer is willing to back a losing production, the annual cycle of openings and closings, and the time over which a star performer tires of a particular part and moves on to other challenges. An established production like *Cats*, having survived these initial hurdles, must deal with the decade- to generation-long scale over which taste and fashion change substantially and the production experiences a nearly 100% turnover of personnel. Including this long-term scale would temper Gott's predictions for extraordinarily long runs.

The problems with Gott's long-term predictions show up more dramatically in phenomena, such as the longevity of an individual, where an initial period of low death rate is followed by relatively rapid extinction. We do not need a detailed model to tell us whether we should believe Gott's prediction, based on his age of 46.3 years on 27 May 1993, that he has a 1/3 chance to survive to more than 139 years old.

There are two reasons, in my view, why Gott is able to get away with making his scale-free predictions for the survival of governments and plays and periodicals. First, statistical models for the longevity of these phenomena are not well developed, so Gott is protected from the absurdities that arise immediately in the three examples used in this article. Although there are readily identifiable time scales associated with the phenomena Gott considers, how to incorporate them into prior probabilities for duration has not been much investigated. There is a good reason for this: to assess the viability of an established government or play or periodical, readily available current data about the particular phenomenon in question—data such as the popularity of the government, the balance sheet of the play or periodical, trends in attendance at the play or the number of subscribers of the periodical—are far more cogent than prior information about longevity together with the present age. Second, the intervals that Gott finds for survival times are so wide that he is likely to be right, until he is forced to place bets based on the high probabilities he assigns to long survival times. A negative feature of such bets, however, is that the bettors might not survive until the bets are settled. Even for the case of human longevity, where one could easily formulate bets that Gott would almost certainly lose, the time scales are long enough that one might not get much personal satisfaction from winning.

A way to overcome this diÉculty is to bet on the survival of creatures with a shorter lifetime than humans, but for which

data on present age and future survival are readily obtainable. For this purpose, I sent an e-mail on 21 October 1999, and again on 2 December 1999, to my department's most comprehensive e-mail alias, which includes faculty, staff, and graduate students, requesting information on pet dogs. The responses were compiled and checked for accuracy on 6 December; a notarized list of the 24 dogs, including each dog's name, date of birth, and breed, and the caretaker's name, was deposited in my departmental personnel file on 21 December 1999. Gott's rule predicts that each dog will survive to twice its present age with probability 1/2. For each of the 6 dogs above 10 years old on the list, I am offering to bet Gott $1000 US, at odds of 2:1 in his favour, that the dog will not survive to twice its age on 3 December 1999. The reason for weighting the odds in Gott's favour is to test his belief in his own predictions: given the odds, his rule says that his expected gain, at $1000 per bet, is $6000; moreover, the probability that he will be a net loser (by losing five or more of the six bets) is $7/64 = 0.109$.

9. Discussion

The stated objective of this article is to determine what prior information underlies Gott's rule. Gott proposed his rule as a predictor of future duration based on knowing the present age and nothing else. What we have discovered is that the actual prior information underlying Gott's rule is both less and more than he thought. On the one hand, Gott's rule is a consequence of the temporal Copernican principle for a phenomenon whose age is unknown, but this universal form of Gott's rule has no predictive power for future durations. On the other hand, Gott's rule as a predictor of future durations is a consequence of discovering the present age of a phenomenon that has no identifiable time scales in the past or future.

What about the focus of Gott's *Nature* article, the longevity of the human species? A species's survival depends on its ability to adapt to short- and long-term environmental changes produced by other species in its ecosystem and by climatological and geological processes. The adaptations are made possible by existing genetic variability in the gene pool and by random mutation. How *homo sapiens* fits into this picture is a complicated question, certainly not amenable to a universal statistical rule. As Ferris [1] puts it, '···in my experience most people either think we're going to hell in a handbasket or assume that we're going to be around for a very long time'. Both views are a reflection of advancing technology. The first comes from alarm at technology's increasing impact—changes might be so rapid that we (and certainly other) species could not adapt. The second comes from a belief that technology can save us—by controlling the environment or by making possible remarkable adaptations such as escaping our earthly environment or changing our genetic constitution.

Gott dismisses all such thinking as the illusions of those who do not appreciate the power of the Copernican principle. He contends that everything relevant to assessing our future prospects is contained in the statement that we are not at a special time. This article shows that the Copernican principle is irrelevant to considerations of the longevity of our species. Perhaps we are still subject to the factors that determine the survival of other species. More likely, our survival—and the survival of many other species along with us—depends on what we do now and in the future. We better think hard about it.

References

[1] Ferris, T., How to predict everything: Has the physicist J. Richard Gott really found a way? *The New Yorker,* **75** (No. 18), 35–39 (12 July 1999).

[2] Gott, J. R., III, 1993, *Nature* **363,** 315.

[3] Gott, J. R., III, 1994, *Nature,* **368,** 108.

[4] Gott, J. R., III, 1996, *Clusters, Lensing, and the Future of the Universe*, edited by V. Trimble and A. Reisenegger (San Francisco: Astronomical Society of the Pacific), pp. 140–151.

[5] Buch, P., 1994, *Nature,* **368,** 107.

[6] Bernardo, J. M., and Smith, A. F. M., 1994, *Bayesian Theory* (Chichester: Wiley).

[7] Jeffreys, H., 1939, *Theory of Probability* (Oxford: Clarendon Press).

[8] Jaynes, E. T., 1986, *Maximum Entropy and Bayesian Methods in Applied Statistics*, edited by James H. Justice (Cambridge: Cambridge University Press), pp. 1–25.

[9] Jaynes, E. T., 1968, *IEEE Trans. Syst. Sci. Cybernetics,* **SSC–4,** 227; reprinted in Rosenkrantz, R. D. (ed.), 1983, *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics* (Dordrecht, The Netherlands: Reidel), pp. 116–130.

Carlton M. Caves has been Professor of Physics and Astronomy at the University of New Mexico since 1992. He received a PhD in Physics from the California Institute of Technology in 1979, working mainly on the theory of gravitational-wave detection. His subsequent career has proceeded from relativity theory through quantum optics to quantum information science, the unifying strand being an interest in how quantum mechanics impacts what we can measure and what we can know. His interest in probability theory springs from the conviction that to have any hope of understanding the weird features of quantum mechanics, one must first have a clear idea of what probability means. He maintains that the only consistent way of thinking about probabilities is the Bayesian interpretation, which holds that probabilities, far from being physical properties, are a measure of credible belief based on what one knows. His main accomplishment is to have had a series of excellent PhD students and postdocs, who continue to teach him most of what he knows.