Past Longevity as Evidence for the Future*

Ronald Pisaturo†‡

Gott (1993) has used the 'Copernican principle' to derive a probability distribution for the total longevity of any phenomenon, based solely on the phenomenon's past longevity. Leslie (1996) and others have used an apparently similar probabilistic argument, the 'Doomsday Argument', to claim that conventional predictions of longevity must be adjusted, based on Bayes's Theorem, in favor of shorter longevities. Here I show that Gott's arguments are flawed and contradictory, but that one of his conclusions is plausible and mathematically equivalent to Laplace's famous—and notorious—'rule of succession'. On the other hand, the Doomsday Argument, though it appears consistent with some common-sense grains of truth, is fallacious; the argument's key error is to conflate future longevity and total longevity. Applying the work of Hill (1968) and Coolen (1998, 2006) in the field of nonparametric predictive inference, I propose an alternative argument for quantifying how past longevity of a phenomenon does provide evidence for future longevity. In so doing, I identify an objective standard by which to choose among counting time intervals, counting population, or counting any other measure of past longevity in predicting future longevity.

1. Introduction. Since the 1990s, several leading philosophic and scientific journals have published numerous items regarding scientific prediction, such as prediction of the longevity of the human race, based on the "Copernican principle" (variants of which are also known in the literature as 'anthropic' principles) and what has become known as the 'Doomsday Argument'. (See, e.g., Gott 1993, 1994; Caves 2000; Sowers 2002; Sober 2003; Monton and Kierland 2006; Adams 2007. See also Leslie 1996 and Bostrom 2002. For a good survey article, see Richmond 2006.)

The Copernican principle, as defined by Gott (1993, 315) states that

*Received May 2007; revised October 2008.

†To contact the author, please e-mail: pos@ronpisaturo.com.

‡I would like to thank Frank Coolen, Glenn D. Marcus, Christopher McKinlay, Michael Dickson, Shayne Wissler, and an anonymous referee of this journal for their helpful comments on earlier drafts, and acknowledge Ayn Rand, whose identification of characteristics as ranges of measurement ([1966] 1990, 6–11) gave me a philosophical foundation for exploring the topic of this paper.

Philosophy of Science, 76 (January 2009) pp. 73–100. 0031-8248/2009/7601-0004$10.00
Copyright 2009 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/599273 Published online by Cambridge University Press

"it is a mistake to assume, without sufficient reason, that we occupy a privileged position in the universe." The 'delta *t* argument' in Gott 1993 seems to interpret this principle as a form of the familiar 'principle of indifference' and as implying directly, without further argument, that we are at the present moment equally likely to be any fraction of the way through the full lifetime of the human race. For example, we are as likely to be, at the present moment, in the first half of humanity's lifetime as we are to be in the second half. More generally, Gott claims that t_p/t_t, the ratio of any extant phenomenon's past longevity t_p to its total longevity t_t, has a probability distribution that is uniform between 0 and 1. Equivalently, the probability density function of the total longevity t_t of a phenomenon in progress is $t_p t_t^{-2}$ for all $t_t > t_p$.

The delta *t* argument as presented in Gott 1993 is not so much an argument as a conjecture. (The supporting argument [1993, 315] that Gott does offer is based on the premise that intelligent species are spawning and becoming extinct at constant rates, and that humanity is one intelligent species selected randomly from those species alive today.) Moreover, as noted often in the literature (e.g., Caves 2000), Gott 1993 can be interpreted as committing the fallacies of hasty generalization and neglected aspect, applying the delta *t* argument to cases—such as a Broadway play, the Berlin Wall, and *Nature* magazine—for which we have much additional information beyond the mere longevity of the phenomenon. Indeed, even for the case of mankind, we have much scientific and humanistic knowledge—well beyond the knowledge of how long mankind has survived—that bears on the question of how long mankind *will* survive.

Interestingly, in the same article, Gott (1993, 316–317) presents and endorses another conjecture, also based on the Copernican principle, claiming that it is the ratio of the number of humans (more precisely, 'intelligent observers') already born to the total number of humans that will be born that has a probability distribution that is uniform between 0 and 1. Gott calls this idea the 'Copernican anthropic principle'. Of course, Gott's delta *t* argument and his Copernican anthropic principle contradict each other whenever birth rates are not constant. Gott seems to imply that his latter idea supersedes the former, but he does not offer an argument for this preference. Of course, having two contradictory conclusions casts doubt on each of them.

Nevertheless, the conclusion of the delta *t* argument has enough plausibility to have invited several arguments in favor (e.g., Gott 1994; Monton and Kierland 2006), as well as against (e.g., Caves 2000; Sober 2003) in the philosophic and scientific literature.

It is common sense that past success is evidence for future success. For example, the more times that a drug cures a disease, the more confident

https://doi.org/10.1086/599273 Published online by Cambridge University Press

a doctor reasonably becomes in the drug. An automobile insurance company reasonably surmises that a driver who has not had an accident for 10 years is a better insurance risk than is a driver who has not had an accident for 10 days. Even regarding the case of human longevity, common sense suggests that the fact that mankind has survived for hundreds of thousands of years is more evidence for the robustness of the race than would be the fact that mankind has survived for only a few thousand years.

Bertrand Russell (1912, 166–167) noted that even in cases for which we have much additional or deeper knowledge beyond knowledge of the mere past success of some phenomenon, we are still relying ultimately on the counting of some kind of past instances. For example, in predicting the longevity of mankind, suppose we rely on physical knowledge regarding the behavior of our solar system and on biological knowledge regarding our genetics and environment; but, if so, then instead of simply counting years that mankind has survived, we are implicitly counting the much greater numbers of years or instances in which principles of physics and biology have held. After all, how do we know that the laws of motion will hold next time except through knowing that they have held many, many times before?

Gott's Copernican principle is, in effect, one attempt to validate and quantify the common-sense assumption that past success is evidence for future success. What is at stake in validating this assumption is much more than the seemingly-frivolous prediction of human longevity. At stake is any claim that something will *continue* to happen based on the fact that it has *been* happening.

A variant of Gott's Copernican anthropic principle is what Leslie (1996) calls the 'Doomsday Argument'. This argument asserts no explicit claim regarding the actual probability distribution of the human race's longevity. Rather, it claims that there is a 'Bayesian shift' making the probability of imminent extinction (what Leslie calls 'Doom Soon', as opposed to 'Doom Delayed' or what we will call 'Doom Late') greater than initially thought before considering the Doomsday Argument. The argument goes like this: Suppose K, the number of humans born so far, is less than N. Further suppose that you initially estimate that the probability that there will be a total of N humans born is p, and that the probability that there will be $m \cdot N$ humans born (where $m > 1$) is q. Then invoke a version of the Copernican anthropic argument: Given that there will be a total of N humans, the probability that exactly K humans would have been born so far is $1/N$; but given that there will be mN humans, the probability that exactly K humans would have been born so far is only $1/mN$. In other words, there having been K humans born so far would be m times more likely if there will ultimately be N humans than if there will ultimately

https://doi.org/10.1086/599273 Published online by Cambridge University Press

be mN humans. Therefore, invoking Bayes's theorem, the initial ratio of p/q has to be adjusted by multiplying it by m. Gott (1994) seems implicitly to endorse the 'Bayesian shift' idea when he uses Bayes's theorem in this very way to derive his delta t argument starting from the Jeffreys (1932, 48) prior, t_i^{-1} for all $t_i > 0$.

Like Gott's delta t argument, the Doomsday Argument relies on a form of the principle of indifference to infer a uniform probability distribution. Nevertheless, as we shall see in Section 4, the Doomsday Argument's uniform distribution and Gott's uniform distribution, though seemingly identical, actually differ. (For an interesting discussion of the principle of indifference, and whether the principle implies any numerical probabilities at all, see Norton 2008.)

Much of the literature on the Doomsday Argument consists of one side attacking the argument (e.g., Korb and Oliver 1998, 1999; Bartha and Hitchcock 1999; Olum 2002; Sowers 2002; Sober 2003) and the other side rebutting (Bostrom 1999, 2002; Bostrom and Cirkovic 2003; Bradley and Fitelson 2003; Adams 2007) those attacks. It can be argued that the two main sides of this debate are not pro-Doomsday and anti-Doomsday, but rather anti-Doomsday and anti-anti-Doomsday. Even Leslie, perhaps the Doomsday Argument's main philosophic champion, rebutting at least a dozen objections to the argument (1996, 206–236), expresses his own doubt about the argument if the future has not yet been determined (1996, 18, 147–152, 233–234). Bostrom, the other most prolific anti-anti-Doomsday philosopher, in turn dismisses (2002, 108) Leslie's indeterminism objection and rebuts (2002, 109–126) about another dozen objections too. But Bostrom (2002, 104–107) raises his own objection, a version of which Leslie (1996, 258–263) had previously rebutted: the indefiniteness of the 'reference class' (e.g., Should we be counting past and future humans or past and future 'intelligent observers'?). Adams (2007) rebuts an objection by Sowers (2002) and, like Bostrom and Cirkovic (2003), concludes that the Doomsday Argument is inconclusive.

The arguments on both sides (anti vs. anti-anti) of this debate often become very abstract and rely heavily on thought experiments. One reason for the abstractness is that the Doomsday Argument does not argue for specific final (posterior) probabilities, but only for Bayesian shifts.

Unlike the delta t conjecture, the conclusion of the Doomsday Argument seems implausible on its face. After all, how could it be that knowledge of past longevity always shortens our estimate of future longevity? Yet a definitive refutation has been elusive.

If the Doomsday Argument is to be rejected, it must be rejected for the right reason, lest a hidden baby be thrown out with the bathwater—especially since that hidden baby might be the ability to estimate the future from the past. In this article, I conclude that:

https://doi.org/10.1086/599273 Published online by Cambridge University Press

1. On the surface, Gott's delta *t* argument and the Doomsday Argument appear consistent with some common-sense grains of truth:
    - The past longevity of a phenomenon does provide evidence for future longevity.
    - In the absence of any evidence regarding a specific time period or a specific longevity, any phenomenon is more likely to end in an imminent time period than in a later one of equal duration. That is, Doom Soon is more likely than Doom Late if the time periods denoted by Soon and Late are of equal duration.
2. Nevertheless, the Doomsday Argument's claim that there is a Bayesian probability shift that must be applied to probability assessments of longevity, making Doom Soon more likely than one would think initially, is false. The argument's invocation of Bayes's theorem conflates future durations and total durations. Moreover, the argument's version of the Copernican principle is often contradicted by the prior probability distributions for past and future duration.
3. The conclusion of Gott's delta *t* argument turns out to be mathematically equivalent to Laplace's famous—and notorious—'rule of succession'. Though Gott's own arguments—including his derivation of the delta *t* formula from the Jeffreys prior using the Doomsday Argument—are invalid, contradictory, or incomplete, there is a valid way to derive Gott's delta *t* formula from the Jeffreys prior.
4. There is an objective standard by which to choose between counting time intervals vs. counting population (or counting any other units of measure) in predicting future longevity. But this standard requires putting aside Gott's Copernican arguments and the Doomsday Argument and identifying a different argument, based on the work of Hill (1968, 1988) and Coolen (1998), in their place.

2. Doom Soon vs. Doom Late. It is trivial common sense that Doom Soon will be far more likely than Doom Late given that Soon and Late are of equal duration, and given the facts in most situations. Let us consider some examples.

Example 1A. Suppose, for the simplicity of round numbers, that the human race is beginning its one millionth year of existence, and that the human birth rate has been and always will be constant (until the bitter end). Suppose further that the only serious threats to the ongoing existence of the race are of a physical nature—such as a giant asteroid hitting the Earth and destroying all human life—and have probabilities that have been estimated by scientists. Suppose that scientists, tracking the density and characteristic movements of asteroids, and so on, calculate that at the start

https://doi.org/10.1086/599273 Published online by Cambridge University Press

of any given year, this probability is one in ten thousand for the coming year.

Thus we have a geometric distribution for the probability of doom in a given year: $P(\text{Doom in Year } t_t) = (1 - q)^{k-1} \cdot q$, where q, the probability that the human race will end in the current year given that the human race is alive at the beginning of that year, is equal to 10^{-4}, and $k = t_t - t_p$, where again t_p = past longevity and t_t = total longevity. That would mean that the probability of 'Doom Soon', that is, doom in the coming year—call it Year 1,000,000—is 10^{-4}. That is, $P(\text{Doom in Year } 1{,}000{,}000) = (1 - q)^0 \cdot q = 10^{-4}$.

Let us calculate the probability of doom in future years, given that the present time is the beginning of Year 1,000,000. The probability of doom in Year 1,000,001 is the probability that we survive Year 1,000,000 times the probability that an asteroid hits in Year 1,000,001: $P(\text{Doom in Year } 1{,}000{,}001) = (1 - 10^{-4}) \cdot 10^{-4} \approx 10^{-4}$. Similarly, the probability of doom in Year 10^9, which we will call Doom Late, is as follows: $P(\text{Doom in Year } 10^9) = (1 - 10^{-4})^{998{,}999{,}999} \cdot 10^{-4} \approx e^{-99{,}900} \cdot 10^{-4} < 10^{-43{,}400} \cdot 10^{-4}$.

In other words, in this straightforward example (a geometric distribution), the probability of Doom Soon is greater than the probability of Doom Late by the astronomical factor of $10^{43{,}400}$. But that does not mean that there is a probability shift that has made the probability of Doom Soon greater than the original 10^{-4} that the scientists originally assigned. It simply means that the probability of Doom Late is much *less* than 10^{-4}, for this trivially simple reason: For there to be Doom in the later year, not only does an asteroid (or the equivalent) have to hit in that year, but we first have to live through nearly a billion years *without* an asteroid hitting.

Note that the factor of $10^{43{,}400}$ is much greater than the Bayesian shift factor asserted by the Doomsday Argument; according to that argument, the shift factor would be 10^3. Suppose there were a Bayesian shift applied on top of the $10^{43{,}400}$, so that Doom Soon were $10^{43{,}403}$ more likely than Doom Late. How would that change the probability of Doom Soon?

The first thing to notice is that we would no longer have a geometric distribution or its continuous analogue, an exponential distribution. The exponential distribution is $P(\text{Doom at time } t_t) = \lambda \exp\{-\lambda t_f\} dt_t$, where λ = probability of doom in the coming year, and t_f = the number of years from now = $t_t - t_p$. To adjust this distribution to account for a Bayesian shift, we would have to introduce an additional factor of t_t^{-1}, and then normalize the distribution:

$$\text{Doomsday-adjusted } P(\text{Doom at time } t_t) = \left[\int_{t_p}^{\infty} R(t_t) dt_t\right]^{-1} R(t_t) dt_t,$$

where $R(t_t) = t_t^{-1} \exp\{-\lambda(t_t - t_p)\}$.

https://doi.org/10.1086/599273 Published online by Cambridge University Press

In our numerical example above, $\lambda = 10^{-4}$ and $t_p = 999{,}999$. To calculate the Doomsday-adjusted probability of Doom in Year 1,000,000, we integrate the above probability distribution for $t_t = 999{,}999$ to 1,000,000, yielding a probability of .0001001. This probability is a 0.1% increase over the prior probability of 10^{-4}. In other words, the purported Bayesian shift in this particular example is negligible.

Thus far, I have considered Doom Soon as the coming time period (e.g., the coming year) and Doom Late as some time period of equal duration (e.g., some one year) far in the future. Such single-unit time periods are used by Doomsday writers such as Leslie in their basic Bayesian-shift arguments, described above. However, Leslie (1996, 200–205) sometimes builds on this basis, going on to treat Doom Soon as, say, doom in the next 150 years or so, and Doom Late as doom much later than that. Let us apply these revised definitions of Doom Soon and Doom Late to our example just above. To calculate the Doomsday-adjusted probability of Doom by the end of Year 1,000,149 of humanity's life (again assuming that we are currently at the start of Year 1,000,000), we integrate the above probability distribution for $t_t = 999{,}999$ to 1,000,149, yielding a probability of 0.01503 or 1.503%. This is a 0.1% increase over the prior probability of 0.01489 or 1.489%—still negligible.

Now of course Doomsday proponents such as Leslie do not usually assume a constant birth rate (though Leslie does so in the example we are now considering), and they claim that the Bayesian shift should be based on numbers of births, not numbers of years. But the Bayesian shift, if real, should apply to any example; and one purpose of the following example will be to reveal a bizarre implication of the Doomsday Argument.

Example 1B. Under his scenario referenced above, Leslie hypothesizes an abrupt change in the probability of doom per year. For instance: through the Year 2150, the probability of doom in the coming year might be .0001 at the start of each year; but then, because of the likelihood that man will settle throughout the galaxy, the yearly probability of doom will plummet to a very small number. Suppose that this number is so small that, if mankind survives past 2150, then it is almost certain that mankind will survive for more than 10 billion years. Then the Bayesian shift has an enormous effect. Prior to the Bayesian shift, the ratio of Doom Soon (before Year 2150) to Doom Late (after Year 10 Billion) is about 1.5 to 98.5, or 1 to 66. But the Bayesian shift increases that ratio by a factor of more than 10^4 (the ratio of 1 million plus 10 billion years vs. only 1 million plus 150 years), making Doom Soon almost certain! And if we count births instead of years, and if the birthrate is increasing, then the Bayesian shift is even greater.

This is the kind of reasoning that Leslie uses to argue that the Bayesian

https://doi.org/10.1086/599273 Published online by Cambridge University Press

shift is very significant. He concludes that we should be much more concerned than we would otherwise have been about Doom Soon, and therefore we should pay more attention to finding ways to reduce the risks to the human race.

But wait. Let us revisit Example 1A, the simple case of the geometric or exponential distribution. But this time suppose that scientists and engineers figure out how to build a defense system that would protect us from asteroids and other major physical threats, the only problem being that it would take 150 years to build the system. According to the Doomsday Argument, it would be folly to build the system! Without the system, there would be a 1.503% chance of doom in the next 150 years; but if it were decided to build the system, then—by the calculation in Example 1B—humanity would be virtually certain to die before the system was completed. Indeed, the same argument would hold even if it would take only one year to build the system. And the same argument holds regarding whether mankind should try to settle throughout the galaxy. Instead of rousing us out of complacency, the Doomsday Argument resigns us to it.

The literature (e.g., Bostrom 2002, 142–157) contains other hypothetical examples in which the Doomsday Argument implies weird conclusions; but these other examples are highly contrived, even fantastic. In contrast, the example above is just the kind of example that Doomsday proponents offer as a likely scenario.

Thus we have raised doubts about the idea of a Bayesian shift, but we have not yet refuted the idea. A diehard (pardon the pun) advocate of the Doomsday Argument might reply, "Aha! This explains why there are no alien races colonizing the galaxy! By the time alien races know enough to colonize space, they also know the Doomsday Argument, which tells them that they would almost certainly perish before implementing any space-colonization program that would make their races far more resistant to extinction. And the foolhardy races that tried to colonize space, defying the Doomsday Argument, perished."

In any event, with at least some doubt raised, and with our basic understanding of why Doom Soon is more likely than Doom Late even without a Bayesian shift, let us put aside Example 1B and build upon Example 1A (the geometric or exponential distribution) in a different way.

Example 2. Let us return to the geometric distribution of Example 1A, but this time let $q = 10^{-6}$ instead of 10^{-4}. Therefore, $q^{-1} = t_p + 1$; that is, the probability of doom in the current year happens to equal (approximately) the reciprocal of the number of years that mankind has survived so far. Suppose that, if many millions of years go by without an asteroid hitting, scientists start to think that the probability of an asteroid hitting in a given year is really less than one in a million. Suppose they actually discover physical evidence that explains why the chance of an

https://doi.org/10.1086/599273 Published online by Cambridge University Press

asteroid hitting is less than they previously thought. (Indeed, that is what one would expect if millions of years went by without an asteroid hitting.) Suppose further that the change in the scientists' probability estimate correlates very closely with the number of years that have passed without an asteroid ever hitting; for example, at the start of Year 2,000,000, the scientists assess the probability of Doom in Year 2,000,000 as one in two million; and at the start of Year 1 Billion, the probability of doom in the coming year is 10^{-9}. Thus we no longer have a geometric distribution; students of probability theory will recognize this new assumption as essentially equivalent numerically to repeatedly assuming Laplace's famous—and in many circles notorious—Rule of Succession (Laplace 1812, xii–xiii; [1825] 1995, 10–11; Jaynes 2003, 155, 279, 339, 553–588), which states, for example, that if the sun has risen for N consecutive days, *and we have no other relevant knowledge*, then the probability that the sun will not rise tomorrow is $1/(N + 2)$.

Under this scenario, the probability of Doom Soon (Doom in Year 1,000,000) is 10^{-6}. The probability of Doom Late (Doom in Year 1,000,000,000) is

$$\begin{aligned} 10^{-9} \cdot \prod_{n=1}^{10^9-10^6} [(10^6 - 1 + n)/(10^6 + n)] \\ = 10^{-9} \cdot [(10^6/(10^9)] \\ = 10^{-9} \cdot 10^{-3} = 10^{-12}. \end{aligned}$$

That is, Doom Soon is a million times more likely than Doom Late. But should this fact make us more concerned than we were before about Doom Soon? Of course not. The chance of Doom Soon is still 10^{-6}, just as unlikely as it was before. It's just that the chance of Doom Late is even less likely than Doom Soon, for two reasons:

1. As each year goes by without an asteroid hitting, new evidence tells us to lower the probability that an asteroid will hit in the coming year. The magnitude of this factor is 10^3, which is t_L/t_S, where t_L is the time of Doom Late and t_S is the time of Doom Soon.
2. The second reason is the same as the reason for the factor of $10^{43,400}$ in our first example. In order for there to be even a chance that an asteroid will hit in a year far in the future, we first must get through many years in which an asteroid does *not* hit. Note that this factor is diminished by the existence of the first factor; it is diminished from 10^{434} (assuming a geometric distribution with $p = 10^{-6}$) to 10^3, which is t_L/t_S.

Thus we have two factors of t_L/t_S by which Doom Soon is greater than

https://doi.org/10.1086/599273 Published online by Cambridge University Press

Doom Late. The delta t argument, which asserts the probability distribution $t_p t_t^{-2}$, also claims two such factors of t_L/t_S, since $t_p t_S^{-2} = t_p t_L^{-2} \cdot (t_L/t_S)^2$.

The Doomsday Argument claims that the 'Bayesian shift' should make us increase the probability of Doom Soon. But in our numerical Example 2, we were led to do something that might seem similar but yet is different: we left the probability of Doom Soon unchanged, and we *decreased* the probability of Doom Late. On the other hand, there does seem to be a Bayesian shift from the Jeffreys prior to the delta t formula. We will address this paradox later.

Now let us generalize the mathematics of our numerical Example 2 a bit. Again letting t_p be past longevity and t_t be total longevity, we have

$$P(t_t|t_p) = t_t^{-1} \cdot \lim_{\Delta t \to 0} \prod_{n=0}^{N} [(t_p - 1 + n)/(t_p + n)], \text{ where } N = (t_t - t_p)/\Delta t.$$

$$P(t_t|t_p) = t_t^{-1} \cdot t_p/t_t$$

$$= t_p t_t^{-2},$$

which is exactly the probability distribution in Gott's delta t argument. Thus Gott's delta t argument is a new argument for an old rule, Laplace's rule of succession!

3. The Past as Evidence for the Future. To many, however, Gott's *new* argument is no more convincing than Laplace's *old* argument for his rule of succession. Therefore, let us consider another approach, a variation on the reasoning used by Jeffreys (1932): "Two measures are made: what is the probability that the third observation will lie between them? The answer is easily seen to be one-third." Jeffreys then derives the Jeffreys prior as the only probability distribution that satisfies this simple, 'easily seen' condition. Hill (1968, 1988, 1993; see also Fisher 1939 and Dempster 1963) generalizes the Jeffreys premise as follows: Given n exchangeable observations ordered by value (with no ties), the probability that the next observation will be in the interval between a given consecutive pair of the n past observations is $1/(n + 1)$. Of interest to us here is this special case: the probability that the next observation will be greater than all n previous observations is less than or equal to $1/(n + 1)$, with the less-than condition accounting for the possibility of ties.

Coolen (1998, 2006), working in the field of nonparametric predictive inference, builds on Hill in order to apply these ideas to calculating probabilities for future successes. A special case of Coolen's general result (which applies as well when the number of past successes and the number of past failures are both nonzero) applies to our example. Define the

https://doi.org/10.1086/599273 Published online by Cambridge University Press

function $G(y)$ as the danger of extinction in the yth year of the life of the human race. Assume that we have no knowledge of how the value of the function G changes with y; that is, we have no reason to believe that the danger increases or decreases with time, and indeed we know of no pattern to G at all. (In other words, the values of G for each value of y are exchangeable.) Suppose that G is defined in such a way that if and only if $G(y)$ reaches or exceeds a threshold value Z, then the human race will perish in Year y.

We do not need to know the actual values of $G(y)$ for each year y. But if mankind has survived through Year Y, then we do know that $G(y) < Z$ for all $y < Y$. Therefore, using Hill's reasoning, the probability that G will exceed Z in the coming year is less than or equal to $1/(Y + 1)$. That is:

$$P\{G(Y + 1) > Z | G(y) < Z \text{ for all } y \leq Y\} \leq 1/(Y + 1). \quad (1)$$

If we ignore the inequality sign, we have the exact same result as the result of Laplace and Gott. Thus Laplace's Rule of Succession and Gott's delta t argument are a worst-case estimate of the risk to the human race as calculated by Coolen's approach, if we consider no information other than mankind's past longevity. Let us call this argument, which applies Coolen's results to the problem of predicting future longevity based on past longevity, the 'danger-function argument'. (Some philosophical underpinning for Coolen's approach can be found in Ayn Rand [1966] 1990, 6–11, which identifies all characteristics as ranges of measurement.)

Now let us consider how we might build on Coolen's approach to reach an exact probability instead of a probability range; that is, let us try to eliminate the inequality sign. Suppose that $G(Y + 1)$ turns out to be the new maximum value for G. What then is the probability that $G(Y + 1)$ exceeds Z? It might be argued (though not without controversy: see, e.g., Norton 2008) that in the absence of any other information, the principle of indifference suggests that this probability is 1/2. Then: $P\{G(Y + 1) < Z | G(y) < Z \text{ for all } y \leq Y\} = 1/(Y + 1) \cdot 1/2 = 1/2(Y + 1)$.

Let us call this special case of the danger-function argument 'max-half-doom'. Using our Example 2 above, if we replace the delta t argument with max-half-doom,

$$P(\text{Doom Soon}) = P(\text{Doom in Year } 1{,}000{,}000) = \frac{1}{2} \cdot 10^{-6}, \text{and}$$

$$P(\text{Doom Late}) = P(\text{Doom in Year } 1{,}000{,}000{,}000)$$
$$= \left(\frac{1}{2} \cdot 10^{-9}\right) \cdot \prod_{n=0}^{10^9 - 10^6 - 1} [(10^6 - 1 + 2n)/(10^6 + 2n)]$$

https://doi.org/10.1086/599273 Published online by Cambridge University Press

$$= \left(\frac{1}{2} \cdot 10^{-9}\right) \cdot [(10^6 - 1)/(10^9 - 1)]^{1/2}$$

$$\approx \left(\frac{1}{2} \cdot 10^{-9}\right) \cdot 10^{-3/2} = \frac{1}{2} \cdot 10^{-10.5}.$$

Again generalizing the math:

$$\text{For } t_t > t_p, P(t_t|t_p) = 1/2 \cdot t_t^{-1} \cdot \lim_{\Delta t \to 0} \prod_{n=0}^{N} [(t_p - 1 + 2n)/(t_p + 2n)],$$

$$\text{where } N = (t_t - t_p)/\Delta t;$$

$$P(t_t|t_p) = 1/2 \cdot t_t^{-1} \cdot (t_p/t_t)^{1/2}$$

$$= 1/2 \cdot t_p^{1/2} t_t^{-3/2}.$$

Thus we have a result that is more optimistic than Gott's for the longevity of the human race or of any phenomenon for which we have past history. Instead of tailing off at the rate of t^{-2}, our distribution tails of at the rate of $t^{-3/2}$. Thus, while delta t predicts an even chance that the total longevity of the human race will exceed twice the past longevity, max-half-doom predicts an even chance that the total longevity of the human race will exceed four times the past longevity. That is, if we solve this equation for x,

$$\int_{t_p}^{x} \frac{1}{2} t_p^{1/2} t_t^{-3/2} dt_t = .5,$$

we get $x = 4t_p$.

One might argue that even max-half-doom is overly pessimistic and provides only a worst-case bound on the probability distribution of future longevity: for if a new maximum for the danger function is reached year after year without doom, it might be argued that the risk of doom given a new high in the danger function is less than one half. But even if one rejects the max-half-doom argument altogether, the reasoning of Coolen bolsters the argument that past success is evidence for future success, and it establishes $t_p t_t^{-2}$ as a worst-case distribution.

Is it mere coincidence that Gott's result is exactly equal to the worst-case bound in Coolen's result? And if more than mere coincidence is at work, what accounts for one result being an equality and the other being a bound? In particular, what accounts for the asymmetry between past and future duration in Coolen's result? We will return to these questions toward the end of the next section.

https://doi.org/10.1086/599273 Published online by Cambridge University Press

4. The 'Bayesian Shift' Debunked. Comparing the formulas for delta t and max-half-doom reveals a pattern. Suppose the probability of Doom in the current year Y, given survival the past $Y - 1$ years is m/Y for all $m > 0$, $Y > 0$. Then for $t_t > t_p$,

$$P(t_t|t_p) = m \cdot t_t^{-1} \cdot \lim_{\Delta t \to 0} \prod_{n=0}^{N} [(t_p - 1 + n/m)/(t_p + n/m)],$$

$$\text{where } N = (t_t - t_p)/\Delta t; \tag{2}$$

$$P(t_t|t_p) = m t_p^m t_t^{-1-m}.$$

In Gott's delta t argument, $m = 1$. In max-half-doom, $m = 1/2$. If $m = 2$, then $P(t_t|t_p) = 2t_p^2 t_t^{-3}$. In other words, to shift the probability distribution by another factor of t_t^{-1}, the probability of Doom Soon needs to be increased by no more than a factor of 2. If the probability of Doom Soon were already very high, say, with the value of m at 100 (e.g., the probability of doom next year for the human race being 10^{-4}), then a Bayesian shift would increase that probability only by a negligible factor of 1.01.

But is there a Bayesian shift at all? Gott claims that his delta t argument entails a Bayesian shift applied to the Jeffreys prior. But let us look at the Jeffreys prior in the context of Equation (2), which yields the Jeffreys prior when $m = 0$ (and t_p is undetermined) or m is arbitrarily close to 0. In other words, the Jeffreys prior implies that the probability of doom in any given year (after the first year) is virtually zero. Let us suppose that we start with the Jeffreys prior before we know anything about the past longevity of the human race. Then we discover that the human race has lived for a million years. Gott's delta t argument asserts that we should let t_p be one million, and we should change the value of m to 1. The argument does not assert that we should *increase* the value of m by 1; rather, it says that whatever m was before is irrelevant, that the *correct value* of m is 1.

In short, the changing of m from 0 to 1 masquerades as a Bayesian shift. But it is not a shift; it is a superseding. Or, to state the point another way, the notion of a Bayesian shift is like a broken clock that tells the right time once a day: there is an apparent Bayesian shift only when the prior probability is the Jeffreys prior.

To see this point from another perspective, consider the case in which the prior probability distribution is the distribution described in our numerical Example 2 above. In other words, suppose our prior distribution is in the form of Equation (2) with $m = 1$, and the probability of doom in the coming year is 10^{-6}. Then suppose we discover that the human race has lived for a million years; the delta t argument would then suggest

https://doi.org/10.1086/599273 Published online by Cambridge University Press

the exact same distribution that we already had from our scientific evidence. In particular, the delta t argument would also set the probability of doom in the coming year at 10^{-6}. Would the delta t argument thus be implying that we should increase the value of m from 1 to 2, and thus increase the probability of doom in the coming year to $2 \cdot 10^{-6}$? Of course not.

Consider a more extreme case, on the extremity opposite that of the case of the Jeffreys prior. Suppose that we do not consider the past longevity of mankind, or of any species for that matter, but instead our scientific evidence leads us to conclude that there is a 1/100 chance each year that an asteroid will hit the earth and destroy the human race, with the chances of such destruction diminishing only slightly each year. Then suppose that we do consider, for the first time, the fact that mankind has existed for a million years, thus having withstood one million alleged one-in-a-hundred chances of doom. Will this new fact make us increase our probability of doom this year? Of course not. The new fact will have an opposite effect. It will make us suspect that perhaps we made an error in our calculations, which imply a value of 10^4 for m, and that perhaps the correct value for m is closer to the value of 1 suggested by delta t, or the value of 1/2 suggested by max-half-doom.

Thus the delta t argument does not imply a Bayesian shift toward Doom Soon, but rather a specific final probability distribution. Moreover, even in the case of the Jeffreys prior, the apparent Bayesian shift does not imply a shift to Doom Soon. Let us see why.

The Jeffreys (nonnormalizable) prior distribution for *future* duration t_f is $P_J(t_f) = (ct_f)^{-1}$, where c is some constant, and Gott's delta-t distribution for *total* duration t_t is $P_G(t_t) = t_p t_t^{-2}$.

But let us compare apples to apples. Gott's delta-t distribution for *future* duration t_f is

$$P_G(t_f) = t_p(t_f + t_p)^{-2} = t_p(t_f + t_p)^{-2} \cdot (t_f t_f^{-1}) = t_p t_f/(t_f + t_p)^2 \cdot cP_J(t_f).$$

Therefore,

$$\text{if } t_f \approx t_p, \text{ then } P_G(t_f) \approx (c/4)P_J(t_f);$$

$$\text{if } t_f \ll t_p, \text{ then } P_G(t_f) \approx (t_f/t_p) \cdot P_J(t_f) \ll (c/4)P_J(t_f);$$

$$\text{if } t_f \gg t_p, \text{ then } P_G(t_f) \approx (t_p/t_f) \cdot P_J(t_f) \ll (c/4)P_J(t_f).$$

Thus, compared to the Jeffreys prior, the delta-t distribution is not a shift toward Doom Soon, but rather a shift toward $t_f = t_p$, in accordance with common sense. For example, the cumulative Jeffreys prior is infinite for $0 < t_f \leq K$ as well as for $K \leq t_f < \infty$, for any $K > 0$; therefore, essentially half of the cumulative density for t_f is in the first future interval of time.

https://doi.org/10.1086/599273 Published online by Cambridge University Press

In other words, according to the Jeffreys prior, the probability of doom within the next year is essentially 1/2. In contrast, once we account for the fact that mankind has lived for a million years, the delta-t probability of doom within the next year shifts *down* from 1/2 to 10^{-6}.

But what about the Doomsday Argument, which claims not to assert specific probabilities but *only* a Bayesian shift? As we shall see, the conflating of t_f (i.e., future duration) and t_t (i.e., total duration) is the key error in the Doomsday Argument's application of Bayes's theorem.

Before analyzing the Doomsday Argument's unorthodox use of Bayes's theorem, it is instructive to consider a more conventional Bayesian analysis, based on our Example 1A.

Example 3. Let us continue to assume that the human birth rate has been and always will be constant, so that it does not matter whether we count births or units of time. Let us start with the 'product rule', which is even slightly more basic than Bayes's theorem, from probability theory. I will use the notation of Jaynes (2003, 88): "X = prior information; H = some hypothesis to be tested; D = the data." One good thing about Jaynes' notation is that it recognizes that even the 'prior' probability generally relies on some information. And so

$$P(H|DX)P(D|X) = P(HD|X) = P(H|X)P(D|HX).$$

We will use the left and right sides of this equation twice, for two hypotheses. Let H_{FS} = the hypothesis that doom will be soon—that is, that t_f, the phenomenon's future longevity, $< t_{FS}$. Let H_{FL} = the hypothesis that doom will be late—that is, that t_f, the phenomenon's future longevity, $> t_{FL}$, with $t_{FL} > t_{FS}$. Let D_p = the fact that the phenomenon's past longevity $= t_p$. Then

$$P(H_{FS}|D_pX)P(D_p|X) = P(H_{FS}|X)P(D_p|H_{FS}X),$$

and

$$P(H_{FL}|D_pX)P(D_p|X) = P(H_{FL}|X)P(D_p|H_{FL}X).$$

Divide the first equation by the second, assuming that $P(H_{FL}|D_pX)$ is nonzero. The term $P(D_p|X)$ drops out:

$$P(H_{FS}|D_pX)/P(H_{FL}|D_pX)$$
$$= [P(H_{FS}|X)/P(H_{FL}|X)] \cdot [P(D_p|H_{FS}X)/P(D_p|H_{FL}X)].$$

The left side of the equation is the ratio of the posterior probabilities. The first bracketed factor on the right side is the ratio of prior probabilities, and the second bracketed factor on the right side is the purported 'Bayesian shift'.

In our Example 1A above, we used a geometric prior distribution; this

https://doi.org/10.1086/599273 Published online by Cambridge University Press

time we will use the continuous analog, the exponential distribution. And so our prior information X states that $P(\text{Doom at time } t_f) = \lambda \exp\{-\lambda t_f\} dt_f$, where λ = probability of doom (from asteroids) in the coming year = 10^{-4}.

Let us suppose this time that mankind's past longevity, t_p, is 200,000 years. Let us further suppose that $t_{FS} = 100$ years, and $t_{FL} = 10^6$ years. Then

$$[P(H_{FS}|X)/P(H_{FL}|X)] = \left[\int_0^{100} \lambda \exp\{-\lambda t_f\} dt_f\right] \Big/ \left[\int_{10^6}^{\infty} \lambda \exp\{-\lambda t_f\} dt_f\right].$$

Now we must evaluate $P(D_p|H_{FS}X)$ and $P(D_p|H_{FL}X)$. But the problem as stated does not give us enough information to do so. We do not have enough information even to evaluate $P(D_p|X)$. In other words, X does not tell us about the prior distribution for t_p. To ascertain such a prior, we must ask: Were the relevant causal forces that are at work from the present onward also at work in the past?

If the answer is no, then the future duration is irrelevant as evidence for the past; and then $P(D_p|H_{FS}X)/P(D_p|H_{FL}X) = 1$, and thus there is no Bayesian shift. But if the answer is indeed no, then we have been wasting our time with this problem and with this entire subject; for then not only is the future irrelevant as evidence for the past, but also the past is irrelevant as evidence for the future. The very idea that past longevity helps us assess the prospects for future longevity rests on the premise that whatever causal factors, for better (e.g., man's reasoning ability) or worse (e.g., the risk of asteroids) were at work in the past will also be at work in the future. This premise is also the key to dealing with the so-called 'reference-class' problem discussed by Leslie (1996, 258–263), Bostrom (2002, 104–107), and others. The correct reference class consists of those instances to which the same causal factors apply. For example, if the threat to human life is and always will be the same as what has always been the threat to primates, then the past longevity of primates should be used as the past duration in projecting the future duration of the human race.

If there is some mixture of old and new causal factors in a given problem, then the problem must be decoupled accordingly. Therefore, let us continue this example with the assumption that the answer is yes, the causal factors from the present onward were also at work in the past.

Note the consequent symmetry between our assessment of the past and our assessment of the future. For example, for mankind to survive at least 200,000 years into the future, we would have to withstand 200,000 10^{-4} chances of doom by asteroids; such a chance is $\exp\{-20\}$. Likewise, for mankind to have been alive for more than 200,000 years prior to

https://doi.org/10.1086/599273 Published online by Cambridge University Press

today, each of the past 200,000 years must have been free of catastrophic asteroids; the chance of that occurrence also is $\exp\{-20\}$.

However, there is an asymmetry. Another condition must have existed in the past: man must have been spawned 200,000 or more years ago, not later than 200,000 years ago. In other words, the absence of a catastrophic asteroid in the past 200,000 years is a necessary condition for mankind's age being more than 200,000 years; but it is not a sufficient condition. Overall, therefore, the expected future life of mankind is greater than the expected past life. More precisely, the distribution for the past longevity is a worst-case bound on the distribution for future longevity. *This result is consistent with Gott's delta* t *result being equal to the worst-case bound in Coolen's result.*

What would it take to move closer to symmetry between the probabilities of past duration and future duration? One scenario would be that there is a kind of built-in 'expiration date' along with a spawning date. Such a scenario would hold for the longevity of a mortal individual as opposed to that of an entire race. Such an individual has a built-in biological clock programmed for eventual death. Another example would be a vehicle's fuel tank, which must become empty eventually. In these cases, there is in effect a second danger function that is cumulative over time.

Another scenario would achieve exact symmetry: Spawning occurs immediately when there is no danger. For example, we could assume that mankind was spawned immediately after the last catastrophic asteroid hit. Of course, this kind of assumption is unrealistic, but it is the kind of assumption implicit in the delta t argument.

In our current example, to make our point, it will not matter whether or not we think that there is some kind of biological clock ticking toward the inevitable extinction of the entire human race, or whether or not there is immediate spawning in the absence of danger. But the immediate-spawning assumption makes the math a little easier, so let us assume that. Therefore, given that $\lambda = 10^{-4}$, the prior probability that mankind has lived 200,000 years or more is $\exp\{-20\}$; and the prior probability that mankind has lived exactly 200,000 years is $\exp\{-20\}(10^{-4})$. In other words, for t_p = 200,000 years, $P(D_p|X) = \exp\{-20\}(10^{-4})$.

What then can we say about the Bayesian shift factor $P(D_p|H_{FS}X)/P(D_p|H_{FL}X)$? Well, it depends; and here we encounter what Bayesian analysis typically is used for in science and engineering: for dealing with a probability distribution of a parameter in a probability distribution. Our answer depends on our probability distribution for λ, our parameter in the exponential probability distribution for t_f. For example, are we virtually certain that the value of λ is 10^{-4}? If so, then whether H_{FS} turns out to be true or whether instead H_{FL} turns out to be

https://doi.org/10.1086/599273 Published online by Cambridge University Press

true is virtually irrelevant for our assessment of λ, and so our assessment of the past continues to use virtually the same value of 10^{-4} for λ given either H_{FS} or H_{FL}, and so $P(D_p|H_{FS}X)/P(D_p|H_{FL}X) \approx 1$, and there is no Bayesian shift.

The more interesting case is the one in which 10^{-4} is merely a maximum likelihood or expected value for λ, and the probability distribution for λ is somewhat flat over a significant range of values. Then

$$P(D_p|H_{FS}X) = P(D_p H_{FS}|X) \Big/ P(H_{FS}X)$$
$$= \left[\int_0^\infty p(\lambda)\lambda \exp\{-\lambda t_p\}\lambda \exp\{-\lambda t_{FS}\} d\lambda\right] d_{tp}\, dt_{FS} \Big/ \lambda \exp\{-\lambda t_{FS}\} dt_{FS},$$

and

$$P(D_p|H_{FL}X)$$
$$= \left[\int_0^\infty p(\lambda)\lambda \exp\{-\lambda t_p\}\lambda \exp\{-\lambda t_{FL}\} d\lambda\right] dt_p\, dt_{FL} \Big/ \lambda \exp\{-\lambda t_{FL}\} dt_{FL},$$

where $p(\lambda)$ is the probability density function for λ. In this case, whether H_{FS} turns out to be true or whether instead H_{FL} turns out to be true is very significant.

Suppose H_{FS} turns out to be true. That suggests that λ probably is larger than 10^{-4}; perhaps the correct value is closer to 10^{-2}. Then $P(D_p|H_{FS}X)$ is more like $\exp\{-2000\}(10^{-2})$, which is much smaller than $\exp\{-20\}(10^{-4})$.

Now suppose H_{FL} turns out to be true. That suggests that λ probably is much smaller than 10^{-4}; perhaps the correct value is closer to 10^{-6}. Then $P(D_p|H_{FL}X)$ is more like $\exp\{-1/5\}(10^{-6})$, which is much larger than $\exp\{-20\}(10^{-4})$. Therefore, $P(D_p|H_{FS}X)/P(D_p|H_{FL}X) \ll 1$, and the 'Bayesian shift' is far in the opposite direction from what the Doomsday Argument asserts.

This is the result expected from common sense. If past duration turns out to be much longer than what would have been expected from the prior probability of doom, then the probability of Doom Late will increase. Conversely, if past duration turns out to be much shorter than what would have been expected from the prior probability of doom, then the probability of Doom Soon will increase.

Now let us turn to the Doomsday Argument's basic equation (see, e.g.,

https://doi.org/10.1086/599273 Published online by Cambridge University Press

Leslie 1996,199–200 for one equivalent version of this equation):

$$P(H_{TS}|D_p X)/P(H_{TL}|D_p X) = [P(H_{FS}|X)/P(H_{FL}|X)] \cdot [P(D_p|H_{TS}X)/P(D_p|H_{TL}X)], \quad (3)$$

where H_{FS} = the hypothesis that the future duration of the phenomenon will be short; H_{FL} = the hypothesis that the future duration of the phenomenon will be long; H_{TS} = the hypothesis that the *total* duration of the phenomenon will be short—that is, that t_t, the phenomenon's *total* longevity, = t_{TS}; H_{TL} = the hypothesis that the *total* duration of the phenomenon will be long—that is, that t_t, the phenomenon's *total* longevity, = t_{TL}, with $t_{TL} > t_{TS}$.

Clearly, this is an invalid application of Bayes's theorem, as it conflates future duration and total duration. This error does not affect the left side of Equation (3), since $P(H_{FS}|D_p X) = P(H_{TS}|D_p X)$ and $P(H_{FL}|D_p X) = P(H_{TL}|D_p X)$, but the error does affect the right side. If we correct the error by considering only future durations, then we get the example that we just completed above. If instead we correct the error by considering only total durations, then we get this equation:

$$P(H_{TS}|D_p X)/P(H_{TL}|D_p X) = [P(H_{TS}|X)/P(H_{TL}|X)] \cdot [P(D_p|H_{TS}X)/P(D_p|H_{TL}X)]. \quad (4)$$

Let us now use this equation to go through a simple numerical example of the type often presented by Doomsday writers.

Example 4. Suppose: t_p = 199,900 years; H_{TS} = 200,000 years; that is, H_{FS} = 100 years, that is, Doom in the 100th year from now; H_{TL} = 10,000,000 years; that is, H_{FL} = 9,800,100 years, that is, Doom in the 9,800,100th year from now. Suppose further these prior probabilities implied by X, the prior information: $P(H_{FS}|X) = .01$; $P(H_{FL}|X) = .99$.

Notice that the Doomsday Argument always starts with prior probabilities of *future* duration, not *total* duration, in keeping with the argument's conflating of these two ideas. The Doomsday Argument wants us to plug in .01 and .99 for $P(H_{TS}|X)$ and $P(H_{TL}|X)$, respectively. But that would be wrong. We will have to *derive* the prior probabilities of these total durations. To help us do so, let us first turn to the Bayesian-shift factor.

According to the Doomsday Argument's version of the Copernican principle, $P(D_p|H_{TS}X) = 1/200{,}000$, and $P(D_p|H_{TL}X) = 1/10{,}000{,}000$. (These values would make the Bayesian shift equal to 50, and the final probability of Doom Soon would then be about .33 instead of .01.) Note that this version of the Copernican principle is different from the version used by Gott! According to Gott, for fixed *past* longevity t_p, the ratio

https://doi.org/10.1086/599273 Published online by Cambridge University Press

t_p/t_t of any extant phenomenon's past longevity t_p to its total longevity t_t has a probability distribution that is uniform between 0 and 1; that is, $P(t_p/t_t|t_p) = U(0,1)$. But for the Doomsday Argument, this uniform distribution holds for fixed *total* longevity t_t; that is, $P(t_p/t_t|t_t) = U(0,1)$. But in many cases, including our current example, the Doomsday Argument's version of the uniform-distribution premise contradicts the prior information!

In our current example, if $t_t = 200{,}000$, then our prior information tells us that $t_p = 199{,}900$ with probability 1; and if $t_t = 10{,}000{,}000$, then there are only two possible values for t_p: 199,900 and 9,999,900. And so, for $t_p = 199{,}900$ years, $P(D_p|H_{TS}X) = 1$, not 1/200,000; $P(D_p|H_{TL}X) = P(t_p = 199{,}900;\ H_{FL}|X) \ / \ [P(t_p = 199{,}900;\ H_{FL}|X) + P(t_p = 9{,}999{,}900;\ H_{FS}|X)]$. By similar reasoning, we have these expressions for the prior probabilities of the total durations: $P(H_{TS}|X) = P(t_p = 199{,}900;\ H_{FS}|X)$; $P(H_{TL}|X) = P(t_p = 199{,}900;\ H_{FL}|X) + P(t_p = 9{,}999{,}900;\ H_{FS}|X)$.

To condense our notation, let H_{PS} = the past was short; that is, $t_p = 199{,}900$; H_{PL} = the past was long; that is, $t_p = 9{,}999{,}900$. Then

$$\begin{aligned} &P(H_{TS}|D_pX)/P(H_{TL}|D_pX) \\ &\quad = \{P(H_{PS}H_{FS}|X)/[P(H_{PS}H_{FL}|X) + P(H_{PL}H_{FS}|X)]\} \\ &\qquad \cdot\{1/P(H_{PS}H_{FL}|X)/[P(H_{PS}H_{FL}|X) + P(H_{PL}H_{FS}|X)]\} \\ &\quad = P(H_{PS}H_{FS}|X)/P(H_{PS}H_{FL}|X). \end{aligned}$$

In order to evaluate the joint probabilities $P(H_{PS}H_{FS}|X)$ and $P(H_{PS}H_{FL}|X)$, we must know: How does knowledge of past duration affect our assessment of future duration? The Doomsday Argument claims to deduce the answer to this question from its version of Bayes's theorem. But now we see that Bayes's theorem, when applied correctly, does not answer on this question.

Suppose the prior information had told us that knowledge of past duration has no effect on our assessment of future duration. Then

$$\begin{aligned} &P(H_{TS}|D_pX)/P(H_{TL}|D_pX) \\ &\quad = P(H_{PS}|X)P(H_{FS}|X)/P(H_{PS}|X)P(H_{FL}|X) \\ &\quad = P(H_{FS}|X)/P(H_{FL}|X) \\ &\quad = .01/.99, \end{aligned}$$

and so there is no Bayesian shift. Of course this result makes sense, since we just assumed that knowledge of the past does not affect our assessment of the future.

https://doi.org/10.1086/599273 Published online by Cambridge University Press

Let us take another example, another variant of our Example 1A, based on the exponential distribution.

Example 5. Let us again use Equation (4), the corrected version of the Doomsday Argument's basic equation, with all the hypotheses—H_{FS}, H_{FL}, H_{TS}, and so on—and other notation defined as before. Suppose the prior information X tells us the following: The danger of asteroids has existed throughout the history of the human race, and will continue to exist, and let λ = the annual risk of doom from this danger; also, human life was spawned immediately after the last catastrophic asteroid hit Earth; also, from now on, there will be a new, additional risk of doom from nuclear war, and let μ = that annual nuclear risk of doom. Then the prior probability $P(\text{Doom will be at time } t_f|X) = (\lambda + \mu)\exp\{-(\lambda + \mu)t_f\}dt_f$, where t_f = the number of years into the future. Similarly, $P(\text{Human race was born at time } t_p|X) = \lambda\exp\{-\lambda t_p\}dt_p$.

Let us see whether the Doomsday Argument's uniform-distribution premise holds in this example. Recall, the Doomsday Argument claims that $P(D_p|H_{TS}X) = 1/t_{TS}$ and $P(D_p|H_{TL}X) = 1/t_{TL}$. But in fact,

$$P(D_p|H_{TS}X) = P(D_pH_{TS}|X)/P(H_{TS}|X)$$

$$= R_{TS}(t_p)dt_p\Big/\Big[\int_0^{t_{TS}} R_{TS}(t_p)dt_p\Big], \text{ where } R_{TS}(t_p) = P(D_pH_{TS}|X).$$

As in the previous examples, we are not given the joint probability $P(D_pH_{TS}|X)$ in the prior information, and so this problem also is underdetermined. But if we again assume that past duration and future duration are independent, so that $R_{TS}(t_p) = P(D_p|X)P(H_{TS}|X)$, then

$$R_{TS}(t_p) = \lambda\exp\{-\lambda t_p\}(\lambda + \mu)\exp\{-(\lambda + \mu)(t_{TS} - t_p)\}$$

$$= \lambda(\lambda + \mu)\exp\{-(\lambda + \mu)t_{TS} + \mu t_p\}.$$

The value of this expression increases as t_p increases, and so the Doomsday Argument's uniform-distribution premise is contradicted by the prior information. This is common sense. If future risk is greater than past risk, then it is likely that future longevity will be shorter than past longevity. Moreover, if the risk of doom by nuclear war did not exist in the past, then past longevity is irrelevant to assessing the future risk of doom by nuclear war; all that past longevity could help us with is in assessing the risk of doom by asteroids. But this example reveals a case in which the Doomsday Argument's uniform-distribution premise does not contradict the prior information, but is instead merely redundant.

Example 6. Suppose $\mu = 0$ so that $P(\text{Doom at time } t_f|X) = \lambda\exp\{-\lambda t_f\}dt_f$. Then $R_{TS}(t_p) = \lambda^2\exp\{-\lambda t_{TS}\}$, a constant value for all t_p, and $P(D_p|H_{TS}X) = 1/t_{TS}$, as the Doomsday Argument claims; likewise,

https://doi.org/10.1086/599273 Published online by Cambridge University Press

$P(D_p|H_{TL}X) = 1/t_{TL}$. Therefore, the 'Bayesian shift' factor $[P(D_p|H_{TS}X)/P(D_p|H_{TL}X)] = t_{TL}/t_{TS}$. But what about the factor $[P(H_{TS}|X)/P(H_{TL}|X)]$? Again, the Doomsday Argument wants us to plug in $\lambda \exp\{-\lambda t_{FS}\}dt_f$ for $P(H_{TS}|X)$ and $\lambda \exp\{-\lambda t_{FL}\}dt_f$ for $P(H_{TL}|X)$. But in fact, the probability density function for total duration is not an exponential distribution. A total duration of t_t is achieved when any combination of t_p and t_f adds up to t_t. And so,

$$P(H_{TS}|X) = \int_0^{t_{TS}} \lambda^2 \exp\{-\lambda t_{TS}\}dt_p = \lambda^2 \exp\{-\lambda t_{TS}\}t_{TS}; \tag{5}$$

$$P(H_{TL}|X) = \int_0^{t_{TL}} \lambda^2 \exp\{-\lambda t_{TS}\}dt_p = \lambda^2 \exp\{-\lambda t_{TL}\}t_{TL}. \tag{6}$$

Note the factors t_{TS} and t_{TL} at the ends of Equations (5) and (6) respectively. These factors will cancel out the 'Bayesian-shift' factor. These factors in Equations (5) and (6) highlight the difference between dealing with probabilities of future duration and probabilities of total duration. For example, a total duration of t_{TS} can be achieved by the whole-number combinations $t_p = 1$ and $t_f = t_{TS} - 1$, $t_p = 2$ and $t_f = t_{TS} - 2$, . . . $t_p = t_{TS} - 1$ and $t_f = 1$; on the other hand, a total duration of t_{TL}, though less likely than a total duration of t_{TS} for sufficiently large t_{TL}, can be achieved by proportionally more combinations of t_p and t_f.[1]

Plugging in values to Equation (4), we have

$$\begin{aligned} P(H_{TS}|D_pX)/P(H_{TL}|D_pX) &= [\lambda^2 \exp\{-\lambda t_{TS}\}t_{TS}/\lambda^2 \exp\{-\lambda t_{TL}\}t_{TL}] \cdot [t_{TL}/t_{TS}] \\ &= \exp\{\lambda(t_{TL} - t_{TS})\} \\ &= \exp\{\lambda(t_{FL} - t_{FS})\} \\ &= P(H_{FS}|X)/P(H_{FL}|X) \end{aligned}$$

and again there is no Bayesian shift.

In short, the Doomsday Argument commits three rather elementary errors. First, the argument conflates future longevity and total longevity

1. Readers of the literature on the Doomsday Argument might notice a superficial resemblance between Equations (5) and (6) and the mathematical expressions of what Bostrom (2002, 66) has dubbed the 'Self-Indication Assumption'. This assumption holds that, in the absence of any information other than that the phenomenon in question is currently ongoing, the probability of t_{TL} is greater than the probability of t_{TS}. Clearly, this assumption is contrary to my argument here.

https://doi.org/10.1086/599273 Published online by Cambridge University Press

in its attempt to apply Bayes's theorem. Second, the argument's uniform-distribution premise contradicts the prior information in many cases. Third, even if we correct these two errors, the statement of the problem does not specify the degree to which past duration is evidence for future duration, and vice versa; in other words, the statement of the problem does not specify the joint probabilities of past and future durations. If we assume that past and future durations are independent, so that the joint probability $P(t_p t_f|X) = P(t_p|X)P(t_f|X)$, then there is no 'Bayesian shift'.

Since the Doomsday Argument is invalid, Gott's use of the Doomsday Argument to derive his delta t argument from the Jeffreys prior is also invalid. There is, however, a valid way to derive the delta t argument from the Jeffreys prior. But we must use the Jeffreys prior in the right way: as the prior probability distribution of a parameter in a probability distribution. Suppose the prior probability distribution for future duration $P(t_f|X) = \lambda \exp\{-\lambda t_f\} dt_f$, and likewise for past duration $P(t_p|X) = \lambda \exp\{-\lambda t_p\} dt_p$. Suppose further that λ itself is a random variable whose prior probability distribution is the Jeffreys prior. In other words, $P(\lambda|X) = 1/\lambda$.

Then

$$P(t_f|X) = \int_0^\infty \lambda^{-1} \lambda \exp\{-\lambda t_f\} d\lambda = 1/t_f, \text{ and } P(t_p|X) = 1/t_p,$$

which is the version of the Jeffreys prior that Gott starts with. However, by starting where Gott does, with a prior distribution for t_f, we have no specification of how knowledge of past duration affects our knowledge of future duration. On the other hand, by starting with the prior distribution for λ, we do have this specification, through the product rule: $P(t_t|t_p X) = P(t_t, t_p|X)/P(t_p|X)$.

We have just seen that $P(t_p|X) = 1/t_p$. What about $P(t_t, t_p|X)$? Since t_p and t_f both are exponentially distributed, and $t_f = t_t - t_p$,

$$P(t_t, t_p|X) = \int_0^\infty \lambda^{-1} \lambda^2 \exp\{-\lambda t_t\} d\lambda = t_t^{-2}.$$

Thus $P(t_t|t_p X) = t_p t_t^{-2}$, which is the delta t formula! We have not used any version of the Copernican principle in this derivation of the delta t formula from the Jeffreys prior.

To summarize, we have refuted the Doomsday Argument and its myth of the Bayesian shift, which has been a source of pessimism regarding the outlook for mankind. At the same time, we have shown that Gott's delta t formula, though Gott's own arguments are flawed and incomplete, and

https://doi.org/10.1086/599273 Published online by Cambridge University Press

Coolen's danger-function argument are both plausible predictors of future longevity in the absence of evidence other than past longevity. Moreover, we have identified how slightly differing assumptions (regarding an expiration date or a spawning date) account for the slight difference between Gott's result and Coolen's. But we have not yet addressed critically the greatest source of pessimism of all: the idea that instead of counting years, we should be counting births.

5. Time vs. Population. As noted, Gott's Copernican anthropic argument is exactly parallel to Gott's delta t argument and simply replaces counts of years with counts of human births. Thus, according to the anthropic argument, if the nth person in history has just been born, the probability that he will be the *last* person born is $1/(n + 1)$.

Gott's population-based (anthropic) argument is much more pessimistic than is his time-based (delta t) argument, because current birth rates are much higher than they were in the past, and they keep increasing. By the time-based argument, the risk of extinction decreases slightly with each passing year. By the population-based argument, in contrast, the risk of extinction decreases slightly with each passing *birth*; but since there are many more births each year, the risk of extinction each year increases substantially.

By the reasoning of Gott (and Leslie as well), there is no logical basis for choosing between these two arguments. Thus, Gott's and Leslie's arguments, if not all their conclusions, are refuted by their own indefiniteness. (This problem is a variant of the reference-class problem.)

In contrast, let us see how the danger-function argument does admit a logical, fact-based criterion for choosing between time and population as predictors of longevity. Using Equation (1) and replacing numbers of years with numbers of births, we have

$$P\{G(N + 1) < Z | G(n) < Z \text{ for all } n \leq N\} \leq 1/(N + 1),$$

where N is the number of humans born so far, and G is the danger to the human race during person n's tenure as the last person to have been born. That is, a population-based danger-function argument implies that the probability distributions for $G(n)$ for each value of n are identical in our context of knowledge.

In the early years of the human race, there might have been a new birth every day. Therefore, the person just born would have enjoyed a full day as the last human born so far. If an asteroid—or the equivalent—had hit at any time during that day, he would have been the last person born, period. In the present era, in contrast, there is a new birth about every fifth of a second. Only if an asteroid or the equivalent hits in the next fifth of a second can the person just born be the last ever. Thus the

https://doi.org/10.1086/599273 Published online by Cambridge University Press

population-based argument implies that the risk of extinction in the next fifth of a second is equivalent to the risk that we might have become extinct in a whole day in prehistoric times.

Here is where we can see clearly how each danger-function argument—one based on time and one based on population—makes very definite assumptions about reality. They are not arguments based on an assumption of no evidence. Each argument does assume certain evidence. We decide between them based on the evidence that actually exists.

Consider the following analogous problem. Suppose we have a new kind of machine that has not failed in its 1,000 days of operation. Suppose that the machine has produced 50,000 widgets in that time. Suppose the output of the machine was only 10 widgets per day at the start of the 1,000 days, but that the output was increased gradually until it reached its current output of 1,000 widgets per day. What is the probability that the machine will fail on the next day? If we assume a danger function based on time, then we have 1,000 past successes and one new trial; and the probability of failure on the new trial is less than or equal to 1/1,001. But if we assume a danger function based on the number of widgets produced, then we have 50 past successes of the production of 1,000 widgets and one new trial of the production of 1,000 widgets; and the probability of failure on the new trial is less than or equal to 1/51. How would we choose between these two danger functions? We would have to identify an objective, uniform measure of a successful trial. To do so, we would have to ask: What factors would cause the machine to fail? For example, would they be factors related to possible faulty raw materials used in the production of each widget? Or would they be factors such as possible leaks or corrosion from the air, whose likelihood is independent of how many widgets are produced each day?

The same kind of reasoning applies to assessing the future prospects for the human race. The population argument implies that the real risks to mankind are not external risks—such as asteroids—that are thought to be constant per unit time, but rather risks caused by humans themselves, as if each new human born might be the one that kills us all. Thus, this argument assumes that the major risks are things such as nuclear wars or humans making their own environment unsuitable to humans. In my judgment, this assumption is mistaken. Even if there were some manmade catastrophe like a nuclear war or a return to the Dark Ages, it is very unlikely that every human being would die; what would probably happen is a return to population levels of earlier times, times which mankind survived. Regarding the environment, it can be argued that the greatest danger is environmentalism itself, which calls for a return to preindustrial societies and preindustrial population levels, which mankind survived.

Indeed, common sense indicates that higher population levels actually

https://doi.org/10.1086/599273 Published online by Cambridge University Press

reduce the risk of extinction per unit time. After all, species do not make the endangered species list by having their birth rates and life spans increase. Imagine a New York City Department of Health official proclaiming: "The birth rate of rats in the city is doubling every year, and the rats are living longer too; therefore, the problem will soon be over." In the case of human beings, the benefit of greater population is much more pronounced than in the case of other species, because of the accumulation of knowledge and the division of labor. The more humans that have existed and the more that exist at any time, the more mental capacity there is to solve any threat to human survival that does arise.

Some do argue that certain manmade dangers such as nuclear bombs or biological weapons can indeed kill every single human. But such arguments are over and above the Doomsday Argument itself. There is nothing inherent in population growth per se that increases the probability of doom per unit time. For example, the probability of manmade doom during the 18th century was not greater than the probability of manmade doom during the 10th century, when the human population was much lower.

Therefore, the conclusions based on growing population, the most pessimistic aspect of the Doomsday Argument, should be rejected along with the argument itself. But in rejecting these conclusions, we have uncovered a standard by which to judge the validity of the time-based conclusions as well. What is the evidence that each uniform interval of time should be equally risky, as the time-based danger-function argument assumes? Instead of each unit of time being equally risky, why isn't it each *square* unit, or any other power of time, that is equally risky? (Goodman 1994, e.g., raises this objection to Gott's delta t argument.)

The evidence is the fact, known from countless observations of reality, of the constancy of the laws of nature. Moreover, there is no discernible significant trend in the amount or kind of motion on the Earth, or in its surroundings, *per unit of time*. (This condition is little more than that the laws of motion continue to apply to things we call clocks in the same way as they apply to other things.) Therefore, in the absence of any specific evidence of a particularly dangerous set of circumstances, the chance that there will be a calamitous motion is constant per unit of time.

Here is another perspective on the differing premises of the Copernican arguments and the danger-function arguments. The Copernican principle holds that there is nothing probabilistically special about an observer's temporal rank among observers. The danger-function argument holds, correctly, that there is nothing causally significant, and therefore nothing probabilistically special, about a given time period's temporal rank among uniform periods. Let me elaborate.

The current time per se is not a causal factor, because time is not an

https://doi.org/10.1086/599273 Published online by Cambridge University Press

entity. It is the universe's entities, in their current states, that will cause the next actions of those entities, irrespective of how many states have transpired so far and what those prior states were. (For one account of Aristotle's conception that it is entities that cause actions, see Gotthelf [1976] 1987, 208–211.)

In other words, causal chains do not hopscotch or leapfrog across time. They are unbroken across time. If an action of an entity a century ago is significant to what will happen in the next instant, it is because that action of that entity has had a lasting effect on some entity *throughout* the past century. Therefore, while our *evidence* for the future is our knowledge of the past through the present, the entire *cause* of the future is contained in the present.

Therefore, the causal progression of the states of entities from time t to time $t+1$ is independent of the absolute value of t. And that is why past causal progressions are evidence for future ones. (Of course, a full philosophical defense of this position on causality, given the tradition of skepticism since Hume, would require a whole paper—or treatise. Here I rely on the reasonable, benevolent expectation that this treatise eventually will be written by someone.) In particular, if survival at time t has always led to survival at time $t + 1$ for many values of t, with no evidence of any significant change of state, then that is mounting evidence that the same thing will happen at time t = the present. (In other words, the intervals $[t, t + 1]$ are exchangeable for all t.) Any variation in the danger function from one value of t to the next is due to unobserved factors, which might be present for any value of t; the probability that these factors will cause the highest danger at time t = the present is the reciprocal of the number of values of t.

In short, this essay has argued in favor of man's ability to predict the future based on the past, and has dispelled all the reasons for pessimism that were based on the Copernican anthropic principle and the Doomsday Argument. Perhaps a new, more optimistic name is called for: The Longevity Argument.

REFERENCES

Adams, Tom (2007), "Sorting Out the Anti-Doomsday Arguments: A Reply to Sowers", *Mind* 116: 269–273.

Bartha, Paul, and Christopher Hitchcock (1999), "No One Knows the Date or the Hour: An Unorthodox Application of Rev. Bayes's Theorem", *Philosophy of Science* 66 (Proceedings): S339–S353.

Bostrom, Nick (1999), "The Doomsday Argument Is Alive and Kicking", *Mind* 108: 539–550.

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

Bostrom, Nick, and Milan M. Cirkovic (2003), "The Doomsday Argument and the Self-Indication Assumption: Reply to Olum", *The Philosophical Quarterly* 53: 83–91.

https://doi.org/10.1086/599273 Published online by Cambridge University Press

Bradley, Darren, and Branden Fitelson (2003), "Monty Hall, Doomsday, and Confirmation", *Analysis* 63: 23–31.
Caves, Carlton M. (2000), "Predicting Future Duration from Present Age: A Critical Assessment", *Contemporary Physics* 41: 143–153.
Coolen, Frank P. A. (1998), "Low Structure Imprecise Predictive Inference for Bayes' Problem", *Statistics and Probability Letters* 36: 349–357.
——— (2006), "On Probabilistic Safety Assessment in the Case of Zero Failures", *Journal of Risk and Reliability* 220: 105–114.
Dempster, A. P. (1963), "On Direct Probabilities", *Journal of the Royal Statistical Society B* 25: 100–110.
Fisher, R. A. (1939), "Student", *Annals of Eugenics* 9: 1–9.
Goodman, Steven N. (1994), "Future Prospects Discussed", *Nature* 368: 106–107.
Gott, J. Richard III (1993), "Implications of the Copernican Principle for Our Future Prospects", *Nature* 363: 315–319.
——— (1994), "Future Prospects Discussed", *Nature* 368: 108.
Gotthelf, Allan ([1976] 1987), "Aristotle's Conception of Final Causality", *Review of Metaphysics* 30: 226–254. Reprinted in Allan Gotthelf and James G. Lennox (eds.), *Philosophical Issues in Aristotle's Biology*. Cambridge: Cambridge University Press, 204–242.
Hill, Bruce M. (1968), "Posterior Distribution of Percentiles: Bayes' Theorem for Sampling from a Population", *Journal of the American Statistical Association* 63: 677–691.
——— (1988), "De Finetti's Theorem, Induction, and $A_{(n)}$ or Bayesian Nonparametric Predictive Inference", in J. M. Bernardo et al. (eds.), *Bayesian Statistics 3*. Oxford: Oxford University Press: 211–241.
——— (1993), "Parametric Models for A_n: Splitting Processes and Mixtures", *Journal of the Royal Statistical Society B* 55: 423–433.
Jaynes, E. T. (2003), *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.
Jeffreys, Harold (1932), "On the Theory of Errors and Least Squares", *Proceedings of the Royal Society of London Series A* 138: 48–55.
Korb, Kevin B., and Jonathan J. Oliver (1998), "A Refutation of the Doomsday Argument", *Mind* 107: 403–410.
——— (1999), "Comment on Nick Bostrom's 'The Doomsday Argument Is Alive and Kicking'", *Mind* 108: 551–553.
Laplace, Pierre-Simon (1812), *Théorie Analytique des Probabilités*. Paris: Courcier.
——— ([1825] 1995), *Philosophical Essay on Probabilities.* Reprint. Translated by Andrew I. Dale. Originally published as *Essai philosophique sur les probabilités* (Paris: Bachelier). New York: Springer-Verlag.
Leslie, John (1996), *The End of the World: The Science and Ethics of Human Extinction*. London: Routledge.
Monton, Bradley, and Brian Kierland (2006), "How to Predict Future Duration from Present Age", *Philosophical Quarterly* 56: 16–38.
Olum, Ken D. (2002), "The Doomsday Argument and the Number of Possible Observers", *Philosophical Quarterly* 52: 164–184.
Rand, Ayn ([1966] 1990), "Introduction to Objectivist Epistemology, Part I", *The Objectivist* 5(7): 1–11. Reprinted in Harry Binswanger and Leonard Peikoff (eds.), *Introduction to Objectivist Epistemology*, expanded 2nd ed. New York: Meridian, 1–18.
Richmond, Alasdair (2006), "The Doomsday Argument", *Philosophical Books* 47: 129–142.
Russell, Bertrand (1912), *The Problems of Philosophy*. New York: Holt.
Sober, Elliott (2003), "An Empirical Critique of Two Versions of the Doomsday Argument—Gott's Line and Leslie's Wedge", *Synthese* 135: 415–430.
Sowers Jr., George F. (2002), "The Demise of the Doomsday Argument", *Mind* 111: 37–45.

https://doi.org/10.1086/599273 Published online by Cambridge University Press