

The Demise of the Doomsday Argument

GEORGE F. SOWERS JR.

A refutation of the doomsday argument is offered. Through a simple thought experiment analyzed in Bayesian terms the fallacy is shown to be the assumption that a currently living person represents a random sample from the population of all persons who will ever have existed. A more general version of the counter argument is then given. Previous arguments that purport to answer this concern are also addressed. One result is determining criteria for the applicability of time sampling arguments, i.e., under what conditions can a specific instant in time be regarded as a random sample from a time span. Given this new understanding, the incredible consequences of the doomsday and related arguments evaporate.

1. Introduction and Background

The doomsday argument (DA) has generated quite a lot of discussion since it was first posed in the late 1980's. Early published accounts of the argument are due to John Leslie (1990, 1992, 1996) and Richard Gott (1993, 1994). The argument is seemingly simple but the conclusions are quite fantastic. The interested parties seem drawn to one of two camps, those who embrace the fantastic conclusions and extend the argument to reach even more fantastic conclusions, and those who, bolstered by the intuitive wrongness of the whole thing, concoct refutation after refutation only to see them refuted in turn by the other camp. For a recent thrust and parry see Korb and Oliver (1998) and Bostrom (1999).

The basic doomsday argument can be understood as an analogy to a simple statistical problem set in the familiar venue of balls and urns. Consider a situation where you are confronted with two large urns. You are informed that one urn holds 10 balls numbered from 1 to 10, and the other holds 1,000,000 balls numbered from 1 to 1,000,000. Your assignment is to determine which is which by blindly reaching into one of the urns and drawing out a single ball. Suppose the ball obtained is marked with a seven. What can you conclude? An application of Bayes' theorem allows you to compute the posterior probability of the two contrasting hypotheses: (A) this urn contains 10 balls, or (B) this urn contains 1,000,000 balls. Bayes' theorem states that

$$P(A | e) = \frac{P(A)P(e | A)}{P(e)} \text{ and } P(B | e) = \frac{P(B)P(e | B)}{P(e)},$$

where e is the evidence that a seven was drawn from this urn. Assuming that the prior probabilities of A and B are equal, the ratio of the posterior probabilities will be the same as the ratio of the likelihoods, $P(e | A)$ and $P(e | B)$. The likelihoods are easily computed provided you make the assumption that it is just as probable to pull one ball out as any other ball. In practice, this can be achieved by thoroughly mixing the balls before sampling. A sample prepared in just this way is known as a random sample and can be summarized by the condition that all possible outcomes are equi-probable. With that proviso, the ratio of the likelihoods is 1,000,000/10 or hypothesis A is 100,000 times more probable than hypothesis B . In the doomsday literature, this result is referred to as the probability shift, i.e., the odds against B have shifted from the prior odds of 1:1 to the overwhelming posterior odds of 100,000:1.

The doomsday argument attempts to map this simple situation into the more complex arena of human populations. Imagine two contrasting hypotheses. *MANY* is the hypothesis that the human race will continue, hale and healthy, far into the future. Under that scenario, many, many people will and have existed. In contrast, *FEW* is the hypothesis that the human race is doomed to become extinct in the near future. Under that scenario, comparatively few people will ever have existed. Suppose each person is assigned a number equal to their birth rank; i.e., you would be assigned the number n if you were the n^{th} person born. If you can regard yourself as a random sample from the set of all persons who will ever have existed, then based on reasoning analogous to the above, there would be a probability shift in favor of the hypothesis *FEW* of magnitude $N_{\text{MANY}} / N_{\text{FEW}}$, where N is the total number of persons under a given hypothesis.

Based on that reasoning, it would seem to be far more likely that the human race would die out sooner rather than later. In fact, the sooner the extinction the higher the probability! This conclusion in and of itself is not all that remarkable. Many other arguments can be used to reach the same conclusion, e.g. nuclear proliferation, global warming, etc. What's remarkable is that it seems to be entailed by the simple fact of my

or your birth rank. In what follows, I will argue that DA is in fact fallacious, and that the fallacy rests in the assumption that you or I or anyone else when indexed by their birth rank can be regarded as a random sample from the set of all humans who will ever have existed.

2. *The Simplified Argument*

A modest adjustment of the original urn problem serves to illustrate the gist of the fallacy. There are two urns populated with balls as before, but now the balls are not numbered. Suppose you obtain your sample with the following procedure. You are equipped with a stopwatch and a marker. You first choose one of the urns as your subject. It doesn't matter which urn is chosen. You start the stopwatch. Each minute you reach into the urn and withdraw a ball. The first ball withdrawn you mark with the number one and set aside. The second ball you mark with the number two. In general, the n^{th} ball withdrawn you mark with the number n . After an arbitrary amount of time has elapsed, you stop the watch and the experiment. In parallel with the original urn scenario, suppose the last ball withdrawn is marked with a seven. Will there be a probability shift? An examination of the relative likelihoods reveals no.

Given the importance of this point, let's examine the issue in detail. As before, the competing hypotheses are: (A) this urn contains 10 balls, and (B) this urn contains 1,000,000 balls. Take for example the likelihood $P(e|A)$. In this term, e represents the outcome of the trial, i.e., the statement 'the last ball drawn reads seven.' A key item is the epistemic state of the experimenter also known as the background data, which we can denote by H . In this case, the relevant portion of H consists of statements describing the sampling procedure as well as the elapsed time recorded on the stopwatch. Often suppressed in the nomenclature, each term in Bayes' theorem should be regarded as conditional on H , for example, $P(e|A)$ should be read as $P(e|A \& H)$, where $\&$ represents logical conjunction. Now e is entailed by H (given the stopwatch reading and the sampling procedure, the marking on the last ball can be deduced) and A is consistent with e . Therefore, e is entailed by $A \& H$ and $P(e|A) \equiv P(e|A \& H) = 1$. Similarly, the

likelihood $P(e | B)$ is also one since the number seven is also consistent with B . In general, if the evidence e consists of a number that is consistent with a particular scenario A or B , then with this sampling procedure, the probability of the evidence is one given that scenario.

If the number drawn exceeds 10, then e and A are inconsistent, the likelihood $P(e | A)$ is zero, and we can conclude that A is false and B is true. So long as the number drawn is less than 10, however, there is no probability shift and the experiment has provided no information to help us distinguish the urns. This basic result does not depend on the time at which the experiment is performed—any time at all is permitted within the sampling procedure. Neither does it depend on advanced knowledge of the stopwatch reading—the watch could be covered without affecting the overall conclusion. In this case, however, the reasoning is more subtle. The likelihood may not be 1, but it would be equivalent between the two scenarios. This is because the stopwatch, a mechanism independent of the competing hypotheses, determines the outcome of the experiment. Therefore as long as the outcome of the experiment is consistent with both A and B the posterior probabilities of both A and B are the same as the prior probabilities. In symbols, if e and A are independent then $P(e | A) = P(e)$ and by Bayes' Theorem $P(A | e) = P(A)$.

What is crucial is that a correlation has been enforced between the stopwatch and the experimental result, which renders the sampling process patently non-random. The watch enforces an ordering on sampling in that it is only possible to sample a seven after it has become impossible to sample a four and before it becomes possible to sample a ten. The strictures of random sampling, on the other hand, require that all results are equally probable any time a sample is taken.

We can now generalize this example to the case of the doomsday experiment. My claim is that by assigning a rank to each person based on birth order, a time correlation is established in essentially the same way that the stopwatch process established a correlation with the balls. Any instance of the experiment, e.g., when I reflect on the argument and attempt to reach its conclusion, occurs at a time when the outcome is (mostly) determined by the correlation. My epistemic state is as follows. I know my birth rank, and that my birth rank is determined by counting in temporal order of birth from the first human born up to me. For example, in the year 2000 the result will be a number

between 60 and 70 billion—no other result is possible now. Here we can distinguish between two possibilities. We can regard our birth rank as being entailed by the time at which we were born. Then the likelihood is one and we have no probability shift. Alternatively, we can simply regard our birth rank as being determined by the correlation, without knowing the details. As above, since the evidence is determined by something that is independent of which scenario happens to be true the posterior probability is equal to the prior probability. Again, no probability shift occurs.

Another (equivalent) way to look at this is that the *relevant* information in my birth rank boils down to the simple statement that at least 60-something billion people will ever have been born. This can be seen by considering again our modified urn experiment. Suppose you have been asked by your boss to determine the number of balls in one of the urns. You decide that the best approach is simply counting the balls. You begin taking balls out one by one, setting them aside, counting as you go. After a minute or two, your boss returns and asks, “What is your answer?” “Does the urn contain 10 or 1,000,000 balls?” At this point you have only counted seven balls. What can you say? Simply that the urn contains at least seven balls. Analogously, for doomsday you can say only that at least 60 something billion people will ever have been born. That statement is entailed by both *FEW* and *MANY* so long as $N_{FEW} \geq 60\text{-something billion}$. Hence the likelihood is one in both cases, no probability shift can occur and the conclusions of DA are avoided.

3. *The General Argument*

The version of the argument put forth by Richard Gott posed the problem not in Bayesian terms, but as an exercise in confidence intervals. The argument goes like this. Suppose we have a process that continues for some duration in time. Then there is a starting time t_{start} and an ending time t_{end} . If we can regard the current time t_{now} as a random sample from the interval (t_{start}, t_{end}) , then certain conclusions can be drawn. In particular, we can establish confidence intervals for t_{now} in terms of t_{start} and t_{end} which can be used to estimate t_{end} in cases where only t_{start} and t_{now} are known.

Variations on this theme have been used to reach some amazing conclusions. For example, the best estimate of the duration of a given process under these assumptions is that t_{end} will be as far in the future as t_{start} is in the past. This “conclusion” can be used to estimate everything from the duration of your marriage, to your own lifespan, to the persistence of American democracy. That is, if you buy the argument. However, this version of the argument fails for the same reason that the urn version did. Time now cannot be regarded as a random sample from a particular time span that contains it. True randomness is much more difficult to achieve than that. The failure is the same as before. The sampling of instances in time given by the procedure of selecting t_{now} is constrained to follow an ordered sequence enforced by the inexorable flow of time itself. If we think of an instant in time as a sample from a time interval, there is a vast difference between that instant being identified as t_{now} for an observer embedded within time, and that instant being selected by a random algorithm from a pre-recorded interval. The former is non-random, while the latter can be made random by using a suitable random algorithm to pick from within the interval. Thus, criteria can be established that allow us to claim that an instant in time is a random sample from a time interval. First, the interval must “exist” in some sense outside of time. Examples would be a set of radio signals indexed by time, a collection of telemetry tapes from a rocket launch or a videotape of an experiment. Second, an algorithm known to possess the appropriate random properties must be used to select from the available moments in the data.

The claim that t_{now} is not a random sample from a time interval that contains it has been made before and the proponents of DA have developed several counter arguments. I will address two of the most prevalent, the emerald example by John Leslie and the amnesia chamber by Nick Bostrom. The emerald example goes like this:

A firm plan was formed to rear humans in two batches: the first batch to be of three humans of one sex, the second of five thousand of the other sex. The plan called for rearing the first batch in one century. Many centuries later, the five thousand humans of the other sex would be reared. You don't know which centuries the plan specified, but you are aware of being female. You

very reasonable conclude that the large batch was to be female, almost certainly. If adopted by every human in the experiment, the policy of betting that the large batch was of the same sex as oneself would yield only three failures and five thousand successes. (Leslie 1996, pp. 222–23)

In this conclusion I would agree with Leslie, but the situation is entirely different from the doomsday scenario. The state of knowledge of the subject consists of several facts. First, there are two groups of humans separated in time by several centuries. Second, there are three in the first group and five thousand in the second group. Third, the subject is in one of the two groups. There is no information that establishes a correlation between the time the subject is living and one or the other of the two groups. As far as she knows, the first batch is living now and the second will be born centuries later. Or equivalently, the second batch is living now and the first batch is long dead. She has no way to rule out either as a possibility. Clearly, if she knew the precise century for the batches she could determine for certain which batch she was in by which century contained t_{now} . But, based on her state of knowledge, all possible outcomes—that she could be any one of the 5003 persons in the experiment—must be considered equi-probable. In this situation the requirements for a random sample are indeed satisfied.

The amnesia chamber is a variation of the doomsday argument where you are in an isolation chamber and don't know what birth rank you have. You do know that there are two mutually exclusive possibilities for the human race, one of which must be true. We can label them as before by *FEW* and *MANY*. Bostrom reasons as follows:

Suppose you obtain a new piece of evidence: your rank is higher than the number of individuals in *FEW*. That conclusively proves that *MANY* is true. This implies that if you had instead found that you had a rank that was low enough to be compatible with both hypotheses, then that would have increased the probability of *FEW*; because if you thought that the new piece of evidence could lower but never raise the probability of *FEW*, then you would be inconsistent, as is easily shown by a standard Dutch Book argument, or more simply by the following little calculation. (Bostrom, 1996)

He then calculates the ratio of the posterior probabilities of *MANY* and *FEW*, to wit

$$\frac{P(MANY | e)}{P(FEW | e)} = \frac{P(e | MANY)P(MANY)}{P(e | FEW)P(FEW)} = \frac{P(e | MANY)P(MANY)}{P(FEW)} < \frac{P(MANY)}{P(FEW)},$$

where use has been made of the “fact” that $P(e | FEW) = 1$ and that $P(e | MANY) < 1$ in the example as posed. Those “facts”, however, depend on the very thing the demonstration purports to show, i.e., that you are a random sample from the set of all persons ever to have existed. As I have shown above, given the actual way in which sampling in the doomsday scenario occurs, the likelihood $P(e | MANY) = 1$ so no probability shift takes place.

Another way to see this point is to suppose that the evidence e consists of the statement “my rank is n .” In the doomsday situation this statement is equivalent to the statement “there are at least n persons who will ever have existed.” In this later form, it is easy to see that

$$P(e | FEW) = \begin{cases} 1, n \leq N_{FEW} \\ 0, n > N_{FEW} \end{cases} \text{ and } P(e | MANY) = \begin{cases} 1, n \leq N_{MANY} \\ 0, n > N_{MANY} \end{cases}.$$

The evidence in the amnesia example is just the conjunction of the two statements “my rank is n ” and “ $n \leq N_{FEW}$ ” which leads again to the conclusion that both likelihoods are unity.

4. Summary and Conclusions

The doomsday argument has been shown to be fallacious due the incorrect assumption that you are a random sample from the set of all humans ever to have existed. Rejection of the doomsday argument leaves us in a much more comfortable position than the alternative. There seems something inherently wrong with accepting the idea that a datum

which is entailed by all possible futures (the number of persons who have existed in the past) can be used to distinguish between those futures. Other apparent anomalies are expunged, like the problem of the early human who might have reasoned in the doomsday manner to conclude that the chance of a human with your or my birth rank was essentially nil.

There are a few object lessons here as well. In none of the literature is there any rigorous justification of the assumptions inherent in doomsday. In the scientific community, the burden of proof is customarily on those who make outrageous claims and not on those who doubt them. An argument that seems wrong often is. However, in the age of Quantum Mechanics, we often embrace a fantastic conclusion simply because it is fantastic and shocking. Our sensibilities have been numbed. But the world is not so topsy-turvy that we can reason a la doomsday.

Lockheed Martin Astronautics
P.O. Box 179
Denver, Colorado, 80201 USA
Email: gfsowers@msn.com

GEORGE F. SOWERS JR.

REFERENCES

- Bostrom, N. 1998: "Investigations into the Doomsday Argument", Preprint at <http://www.anthropic-principle.com/preprints.html>.
- 1999: "The Doomsday Argument is Alive and Kicking", *Mind*, 108, pp.539–550.
- Gott III, R. J. 1993: "Implications of the Copernican Principle for our Future Prospects". *Nature*, vol. 363, 27 May, pp.315–319.
- 1994: "Future Prospects Discussed". *Nature*, vol. 368, 10 March, p.108.
- Korb, K. B. and Oliver, J. J. 1998: "A Refutation of the Doomsday Argument". *Mind*, 107, pp.403–410.
- Leslie, J. 1990 (edt.) *Physical Cosmology and Philosophy*. Macmillan Publishing Company.
- 1992. "Doomsday Revisited". *Phil. Quat.* 42 (166) pp.85–87.
- 1996. *The End of the World: The Ethics and Science of Human Extinction*. Routledge. London.