

Critiquing the Doomsday Argument

by [Robin Hanson](#) 27/08/1998

A thought-provoking argument suggests we should expect the extinction of intelligent life on Earth soon. In the end, however, the argument is unpersuasive.

Brandon Carter, H.B. Nielsen [12], John Leslie [8], J. Richard Gott [5], Nick Bostrom [1], and others have elaborated a creative argument suggesting that "doom" is more likely than we otherwise imagine.

The basic idea is that, all else equal, you should not assume that you are especially unusual on any one axis. If you have just taken up a new fashion, you shouldn't assume you are one of the first; that fashion is just as likely to be on the way out as it is to be on the way in. If you think up a idea that a dozen people are likely to think of as well, you shouldn't assume you are the first.

A similar intuition is applied to the case of finding yourself in an exponentially growing population that will suddenly end someday. Since most of the members will appear just before the end, you should infer that that end probably isn't more than a few doubling times away from now. If you get involved in a pyramid investment scheme, it will likely collapse soon and leave you at a loss. And since humanity has been growing exponentially or faster for millennia, humanity may well end before it can double a few more times. Hence the name, "doomsday argument," (hereafter known as DA).

Of course the end of humanity need not imply doom; perhaps our descendants will just be so different that they are no longer "human." DA proponents often argue, however, that since a valid "reference class" is something like all intelligent creatures, we should conclude that any descendants aren't likely to be intelligent.

Do we face doom soon? Most people's initial reaction is that the argument has to be wrong somehow. Yet most of the knee-jerk counter-arguments they offer are easily rebutted. There have, however, been some thoughtful publications on this topic, mostly critical [2,3,4,6,7,11], and mostly focusing on the same criticism:

All else is not equal; we have good reasons for thinking we are not randomly selected humans from all who will ever live.

You should include everything you know when doing inference, and you usually know things that imply you are not random. If you edit a fashion magazine, you have reason to think you will hear of fashions on their way in. If you never even read fashion magazines, you have reason to think you

will hear of fashions on their way out. Similarly, standard calculations about doom suggest we are earlier than random humans.

For example, assume that human population grows exponentially, that doom happens when the population reaches 10^d , and that our "prior" (before information) expectations about d is that it is uniformly distributed between 0 and 20. In this case if we interpret our information that today's population is about 10^{10} as just telling us that d is at least ten, our "posterior" (after information) expectations should be that d is uniformly distributed between 10 and 20. This implies a median future growth factor of 10^5 before doom, which is far from "doom soon."

But for an exponentially growing population which ends suddenly, don't most members show up within a few doubling times of the end? Yes, but most members also show up in scenarios where the population has nearly reached its maximum possible size. In the example above, most people show up when the population has reached over 10^{19} . If you show up before that point, you already know you are special. Anyone who shows up when the population is well below the largest size that one's prior gives substantial weight to knows they must be unusually early. This is why people who have done these sort of calculations haven't been terribly worried about doom soon.

In response to such standard analyses, DA advocates invite us to imagine that we had amnesia, and have forgotten our place in history. If we then learned that that the population now were small, that *must* increase our posterior beliefs in an early doom. This follows directly from the fact that had we learned instead that the current population was very large, we could have excluded early doom scenarios. So if the above prior on d , uniform within 0 to 20, was reasonable for someone in the amnesia situation, then upon learning our place in history, that the population is now 10^{10} , we must expect doom relatively soon.

A defender of the standard analysis can, however, reply that knowing that you are alive with amnesia tells you that you are in an unusual and informative situation; it is far from being "prior" to information. You may be more likely to develop amnesia in response to a traumatic situation, for example. And even if everyone had the same random chance of developing amnesia, the mere fact that you exist suggests a larger population. After all, if doom had happened before you were born, you wouldn't be around to consider these questions. So the mere fact that you exist would seem to tell you a lot.

For example, imagine that you found yourself hung over in a hotel room, and couldn't remember who you were, other than that you are a musician on tour. You wonder: am I a one-hit-wonder, or do I have a lasting music career? If on average only one quarter of musicians have a lasting career, but musicians with lasting careers spend three times as much time in hotel rooms on tour, then you should estimate a fifty-fifty chance you have a

lasting career. This is because only half of the total musician hotel room-days are filled with one-hit-wonders. Amnesia implies optimism. Similarly, if you can't recall how old you are, you should expect to be older than the average person. Why? Because older people have more hotel-days in their lifetime. Similarly, if you can't recall who you are in humanity, you should become more optimistic about humanity's chances. In the standard approach, priors are defined without considering who, if anyone, will live. In this case, learning that you are alive with amnesia must make you expect both that doom is near, and that doom will happen very late. In our example, beliefs change from being uniform on d to being exponential in d , with most weight near the upper limit of 20. If you then learned, to your great surprise, that the population is now only 10^{10} , your beliefs regarding d would then give much more weight to lower values of d ; it is true that you would expect an earlier doom. But where you would end up is with d being uniformly distributed across 10 to 20, just as if you had never had amnesia.

DA advocates seem to be saying, in contrast, that the amnesia situation is the natural one to be defining "priors" for, with counterfactual situations where you might not have existed being irrelevant distractions. So the question seems to come down to whether, when defining a prior, you should assume that you would have existed at some point in human history, with amnesia about your particular situation. And if one chooses amnesia, there seems the further question of which type of amnesia to assume. Knowing that you lived in a city on Earth, for example, is very different from not being able to exclude either living in space as a computer mind, or as a hunter-gatherer on the savannah.

DA advocates say many other things in defense of their position. But I find it hard to make sense of many of their qualitative arguments, and I find it frustrating that these arguments have rarely been fully formalized using our standard formal approach to modelling inference (at least when everything is finite). This approach is:

1. Choose a space of possible states.
2. Assign a prior probability distribution over states.
3. For each agent in each state, say what other states they can exclude based on their observations. The set of states an agent cannot exclude is their *information set* (and the set of such sets forms a state partition).
4. Posterior probabilities for each agent are just priors renormalized to their information sets.

Admittedly, DA advocates have written down various conditional probability expressions. But they do not seem to have described the state space they have in mind in enough detail for me to judge whether I think their priors reasonable. Reading between the lines, however, DA advocates do seem to be claiming that our standard approach to defining state spaces

and/or priors is biased against doomsday, and so we should revise the practice of decision theory in an important way.

In the above example we chose the state description to be just d , the population exponent at doom. Instead, DA advocates seem to argue, you should also include a description of which human you turned out to be. If you assume you were equally likely to be any one of the actual humans there will actually ever be, this increases the chances of doom soon (though not by a lot in many cases [6,7]).

To explore this, let us consider an even simpler example. Assume there are four spatial positions, a,b,c,d , five points in times, $1,2,3,4,5$, and that each space/time combination can hold one of these: D (ead rock), M (onkey), H (uman), P (osthuman). Each "universe" then describes which of D,M,H,P occupy each space/time slot. Here are four universes: $*,\&,\#,@$.

*	1	2	3	4	5	&	1	2	3	4	5	#	1	2	3	4	5	@	1	2	3	4	5	
D D D D D						a D D D D D						a D D D D D						a D D D D P						a
D D D D D						b D D D D D						b D D D D D						b D D D D P						b
M M M M						c D M M M M						c D M M H D						c D M M H P						c
M M M						d D M H M M						d D M H H D						d D M H H P						d

These are universes where humans never evolve ($*$), evolve but quickly die ($\&$), displace the biosphere and then die ($\#$), or evolve into spacefaring post humans before dying ($@$).

In our example, the usual state description would just specify which was the true universe among $*,\&,\#,@$, and a uniform prior would give equal probability to these four possibilities. If an agent at $d3$ could only observe that $d3$ is occupied by a H , and not a M , she could exclude only the universe $*$. That makes her information set $\{\&,\#,@\}$, and her full partition $\{*\}\{\&,\#,@\}$. Assuming equal prior probabilities for all four universes, she'd then assign only a $1/3$ probability to "doom soon," i.e., being in universe $\&$.

Instead of this usual approach, DA advocates seem to suggest that you should extend your state description to include a description of which member of the "reference class" you turned out to be. Assuming the reference class is humans and post-humans, and using space-time coordinates to denote members, then the universe $\&$ has one associated state $\&d3$, the universe $\#$ has three associated states $\{\#d3,\&d4,\&c4\}$, and the universe $@$ has seven associated states $\{@d3,@d4,@c4,@d5,@c5,@b5,@a5\}$.

The prior that DA advocates seem to prefer is to divide up the prior one might assign to a universe among its associated states. So $P(\&d3) = 1/4$, but $P(\#d3) = 1/(4*3)$ and $P(@d3) = 1/(4*7)$. If my current information were "I'm a H at time 3," that would mean that the true state is one of $\{\&d3,\#d3,@d3\}$, and my posterior probability of "doom soon" ($\&$) is $1/(1+ 1/3 + 1/7) = 72\%$. Thus it seems that DA advocates *can* formalize their intuition that our standard analyses tend to misstate the state space and/or prior, and that

analysis using an extended state space and matching prior can make doom more likely.

Of course we always knew you could make doom more likely if we chose different priors. But how reasonable are these priors? I see these problems with this approach:

1. It is not clear how many states to associate with universes, such as $*$, which have no members. Yet we need to know this to do anthropic reasoning about what our existence tells us about our universe.
2. This DA prior doesn't suggest doom if there are many aliens who count as in the "reference class," and who are insensitive to our doom. If we could have been aliens instead of human, then the fact that we are human suggests that humans are relatively numerous. (Bostrom discusses this at length [1].)
3. There seems to be no satisfactory principle for choosing the reference class of "creatures like us," even though this choice can make a big difference. Changing the reference class in our example from " H or P " to "just H " lowers the chance of doom soon (i.e., universe $\&$) given "I'm a H at 3" from 72% to 60%, and changing the class to " M or H or P " lowers it to 32%.
4. People who want to think of themselves as "educated" tend to define this as everyone with their level of education or higher. Then they can proudly compare the favourable average properties of their "educated" class relative to the "uneducated" class. A similar potential for bias arises when humans define "creatures like us" as creatures almost as intelligent as we are or better.
5. It seems hard to rationalize this state space and prior outside a religious image where souls wait for God to choose their bodies.

This last objection may sound trite, but I think it may be the key. The universe doesn't know or care whether we are intelligent or conscious, and I think we risk a hopeless conceptual muddle if we try to describe the state of the universe directly in terms of abstract features humans now care about. If we are going to extend our state descriptions to say where we sit in the universe (and it's not clear to me that we should) it seems best to construct a state space based on the relevant physical states involved, to use priors based on natural physical distributions over such states, and only then to notice features of interest to humans.

Toward this end, let us think of a human as a certain set of atoms arranged at a certain time in a certain way. Those atoms could be arranged to make a human, or a monkey, or a rock. In this sense, "I could have been a rock." And let us express the DA idea that "I" could have been you by saying that "I" could have been "at" your atoms instead of mine, arranged they way you are.

Formally, let us extend the state description in our simple example to include not only which of $*, \&, \#, @$ is the true universe, but also which

space-time slot I turn out to occupy, allowing *any* valid space-time slot as possible. So, for example, I hope the state is @c4, where I am human in a universe without doom soon, and I'm glad it is not #d5, where "I" would be a rock.

If we choose a uniform prior over all 80 states so defined, then if my current information was "I'm a *H* at time 3," that would mean that the state is one of {&d3,#d3,@d3}, and my posterior would assign equal probability to universes &,#,@. Since this is only a 1/3 chance of "doom soon," we here get back the result we had at first, using universes at states. Thus choosing states and priors in a more physics-oriented way seems to eliminate the doom-enhancing effects of extending the state space to allow us to imagine that I might have been you.

This alternative approach to extending states does have some problems. It seems to suggest that a non zero prior probability of a universe with an infinity of humans implies probability one that we find ourselves in an infinite universe. And it seems difficult to use it when universes have varying numbers of space-time slots. If these difficulties cannot be overcome, however, I would rather go back to the standard approach to defining states.

It is interesting that doomsday argument proponents seem to challenge our usual way of doing inference, by preferring an extended state space where we explicitly model the idea that "I could have been you." However, if we try to do this in a physics-oriented way, avoiding describing states directly in abstract features of interest to humans but not the universe, we get seem to get the same chance of doom as if we hadn't extended states at all.

Humanity may in fact face doom soon, and we have many reasons to be concerned about this. But I do not think the doomsday argument is one of them.

- [1] Nick Bostrom, "[Investigations into the Doomsday argument](#)" Technical Report, www.hedweb.com/nickb/doomsday/investigations.doc 1998.
- [2] D. Dieks, "Doomsday - Or: the Dangers of Statistics" *Philosophical Quarterly* 42:78-84, 1992.
- [3] William Eckhardt, "A Shooting Room-View of Doomsday" *Journal of Philosophy* 59:244-259, 1997.
- [4] William Eckhardt, "Probability Theory and the Doomsday Argument" *Mind* 102(407):483-488, 1993.
- [5] J. Richard Gott, "Implications of the Copernican principle for our future prospects", *Nature*, 363:315-319, 27 May, 1993.
- [6] Tomas Kopf, Pavel Krtous, Don Page, "[Too Soon For Doom Gloom?](#)" Technical Report, xxx.lanl.gov/abs/gr-qc/9407002, 1994.
- [7] Kevin Korb, Jonathan Oliver, "A refutation of the doomsday argument" *Mind* 107(426):403-410, April 1998.
- [8] John Leslie, *The End of the World: The Science and Ethics of Human Extinction*, Routledge, London, 1996.
- [9] John Leslie, "Doom and Probabilities" *Mind* 102(407):489-491, 1993.
- [10] John Leslie, "Doomsday Revisited" *Philosophical Quarterly* 42:85-89, 1992.
- [11] Jonathan Oliver, Kevin Korb, "[A Bayesian Analysis of the Doomsday Argument](#)", Technical Report, www.cs.monash.edu.au/~jono/TechReports/analysis2.ps Jan. 1998.
- [12] H.B. Nielsen, "Random Dynamics and Relations between the Number of Fermion Generations and the Fine Structure Constants", *Acta Physica Polonica*, B 20(5):427-68, 1989.