

THE DOOMSDAY ARGUMENT AND THE SELF-INDICATION ASSUMPTION: REPLY TO OLUM

By Nick Bostrom and Milan M. Ćirković

In a recent paper in this journal, Ken Olum attempts to refute the Doomsday argument by appealing to the self-indication assumption (SIA), the idea that your very existence gives you reason to think that there are many observers. In contrast to earlier refutation attempts that use this strategy, Olum confronts and try to counter some of the objections that have been made against SIA. We argue that his defense of SIA is unsuccessful. This does not, however, mean that one has to accept the Doomsday argument (or the other counterintuitive results that flow from related thought experiments). A developed theory of observation selection effects shows why the Doomsday argument is inconclusive and how one can consistently reject both it and SIA.

I. THE RELATION BETWEEN THE DOOMSDAY ARGUMENT AND SIA

The Doomsday argument purports to show that we have systematically underestimated the probability that humankind will become extinct relatively soon. Originated by Brandon Carter and developed at length by John Leslie¹, the Doomsday argument argues that we have neglected to fully take into account the indexical information residing in the fact about *when* in the history of the human species we exist. In a nutshell: Your birth

¹ Originally in J. Leslie, "Risking the World's End," *Bulletin of the Canadian Nuclear Society*, May: 10-15. (1989). He has a more comprehensive treatment in *The End of the World: The Science and Ethics of Human Extinction* (London: Routledge, 1996). Carter didn't publish on the Doomsday argument. A different version of the Doomsday argument, which we shall not discuss here, was independently discovered by Richard Gott, "Implications of the Copernican principle for our future prospects," *Nature* 363 (27 May): 315-9 (1993).

rank (i.e. your position in the sequence of all humans) is roughly 70 billion. That you should have such a low birth rank is less surprising, and more probable, if the total number of humans that will ever have existed is, say, 200 billion rather than, say, 200 trillion. (By ‘probability’ we here mean rational subjective credence.) Given these unequal conditional probabilities, one can derive from Bayes’ theorem that the probability of impending doom goes up after conditionalizing on your birth rank. That is, after realizing the full evidential import of you having a relatively low birth rank, such as 70 billion, you should increase your probability estimate of hypotheses according to which there will be relatively few extra humans (such as 200 billion in total) at the expense of hypotheses according to which there will be very many more humans (such as a total of 200 trillion).

A key assumption in the Doomsday argument is that you should reason as if you were in some sense a random sample from the class of all observers that will ever have existed. This is encapsulated in the self-sampling assumption²:

(SSA) One should reason as if one were a random sample from the set of all observers in one’s reference class.

We assume for the moment that one’s reference class consists of all observers, although, as it turns out, this supposition is far from innocent.

SSA should be read as a methodological prescription specifying certain types of conditional credences of the form $P(\text{I am such and such an observer} \mid \text{The non-indexical properties of the world are such and such})$. The phrase “as if one were a random sample” is simply shorthand for these recommendations. There is no intimation of any physical

² Originally in N. Bostrom, “Investigations into the Doomsday argument,” *Unpublished manuscript*. <http://www.anthropic-principles.com/preprints/inv/investigations.html> (1997); see also “The Doomsday Argument is Alive and Kicking,” *Mind* 108(431): 539-50 (1999); “Observer-relative chances in anthropic reasoning?” *Erkenntnis* 52: 93-108 (2000); “The Doomsday argument, Adam & Eve, UN⁺⁺, and Quantum Joe,” *Synthese* 127(3): 359-87 (2001); and especially *Anthropic Bias: Observation Selection Effects in Science and Philosophy* (New York: Routledge, 2002).

randomization mechanism – some kind of stochastic time-traveling stork? – responsible for distributing observers in the world.

SSA is what generates the conditional probabilities that, when fed into the Doomsday argument and combined with empirical information about our birth ranks, lead to the probability shift in favor of doom soon. It would take us too far afield to examine what grounds there are for accepting SSA. Detailed arguments for why it (or something much like it) should be applied in a wide range of cases can be found in the literature.³

Olum accepts SSA.⁴ How then does he hope to avoid the conclusion of the Doomsday argument? By appealing to another principle, the self-*indication* assumption (SIA) that says that the sheer fact that you exist should lead you to adjust your prior probability assignment in favor of hypotheses according to which there are lots of observers:⁵

(SIA) Given the fact that you exist, you should (other things equal) favor hypotheses according to which many observers exist over hypotheses on which few observers exist.

If we explicate SIA as saying that the probability boost that a hypothesis gets is proportional to how many observers it says exists, i.e.

$$P(N_{tot} = n | I \text{ exist}) \propto P(N_{tot} = n) \cdot n,$$

(where “ $N_{tot} = n$ ” is the proposition that there are n observers, and $P(N_{tot} = n)$ is the prior probability of this proposition), then it is easy to show that we get an exact cancellation of the Doomsday argument-like probability shift in favor of hypotheses that postulate few observers that comes from applying SSA to our birth ranks. This move was

³ E. g. Bostrom (2002).

⁴ K. Olum, “The Doomsday Argument and the Number of Possible Observers,” *Philosophical Quarterly* 52(207): 164-84 (2002). All page references refer to this paper.

⁵ Bostrom (2002).

first made, albeit not very transparently, by Dennis Dieks.⁶ The exact cancellation was first shown by Kopf et. al and seems to have been independently discovered by Paul Bartha and Chris Hitchcock.⁷ While these authors were satisfied with noting that SIA cancels the apparent Doomsday-effect of SSA, Olum moves the debate a step forward by trying to address some of the objections against SIA that defenders of the Doomsday argument have presented.

II. OLUM'S DEFENSE OF SIA

SIA may seem quite dubious as a methodological prescription or a purported principle of rationality. Why should reflecting on the fact that you exist rationally compel you to redistribute your credence in favor of hypotheses that say that there are many observers at the expense of those that claim that there are few? This probability shift, it should be stressed, is meant to be *a priori*; it is not attributed to empirical considerations such as that it must have taken many generations for evolution to lead to a complex life form like yourself, or that we've discovered that the cosmos is very big and probably contains vast numbers of Earth-like planets. The support for "fecund" hypotheses (the support being proportional to the degree of the number of observers postulated) comes from the sole fact that you exist.

Prima facie, SIA may seem implausible, and our view is that SIA is no less implausible *ultimo facie*. Probably the most positive thing that can be said on its behalf is that it is one way of getting rid of the counterintuitive effects of the Doomsday argument and related thought experiments, including the *Adam & Eve* experiments of Bostrom⁸ to

⁶ D. Dieks, "Doomsday - Or: the Dangers of Statistics," *Philosophical Quarterly* 42(166): 78-84 (1992); see also "The Doomsday Argument," in preparation (available at <http://philsci-archive.pitt.edu/documents/disk0/00/00/02/47/index.html>).

⁷ T. Kopf et al., "Too soon for doom gloom," *Preprint*, <http://xxx.lanl.gov/abs/gr-qc/9407002> (1994); P. Bartha and C. Hitchcock, "No One Knows the Date or the Hour: An Unorthodox Application of Rev. Bayes's Theorem," *Philosophy of Science (Proceedings)* 66: S329-S53 (1999).

⁸ Bostrom (2001).

which Olum alludes in his paper. Yet this palliative relief comes at a cost. Consider the *Presumptuous Philosopher* gedanken:⁹

It is the year 2100 and physicists have narrowed down the search for a theory of everything to only two remaining plausible candidate theories, T_1 and T_2 (using considerations from super-duper symmetry). According to T_1 the world is very, very big but finite and there are a total of a trillion trillion observers in the cosmos. According to T_2 , the world is very, very, *very* big but finite and there are a trillion trillion trillion observers. The super-duper symmetry considerations are indifferent as between these two theories. Physicists are preparing a simple experiment that will falsify one of the theories. Enter the presumptuous philosopher: “Hey guys, it is completely unnecessary for you to do the experiment, because I can already show to you that T_2 is about a trillion times more likely to be true than T_1 !” (whereupon the philosopher runs the argument that appeals to SIA).

Whereas we regard this as something close to a *reductio*, Olum bites the bullet and acknowledges that his position commits him to the view that this philosopher would be right, provided that the prior probabilities for the two theories (based on the super-duper symmetry) were roughly equal.

Olum tries to take the sting out of this consequence by offering us an analogy, meant to demonstrate that we can justifiably reject the thought experiment’s assumption that T_1 and T_2 would have roughly equal priors. Suppose that a stranger comes up to you in the street with the claim that if you give him one dollar, he will give you ten dollars in return tomorrow. Presumably you will decline his offer,

which shows (if you are maximizing your expectations) that you think the chance he will come through as he says is less than 10%. On the other hand, it would be strange to say that the chance that he will come through is less than one in a million, since sometimes people making statements like this are honest. ...

⁹ Bostrom (2002).

Nevertheless if the payoff is raised to ten million dollars, you still will not give the dollar, which shows that now you think the chance for a payback is in fact less than 1 in 10 million. In order not to have a proposed payback large enough to deprive you of money, you must think that the likelihood of getting paid decreases at least inversely with the proposed payback.

Applying this to cosmology, it is possible that one should think that a theory involving a very large universe is unlikely in proportion to the size of the universe it proposes. (p. 183)

The analogy seems tenuous; we doubt that many will find much comfort in it. At any rate, what Olum seems to be suggesting is that there are independent reasons (although he doesn't specify what those may be) for thinking that the universe is not very big, and that these reasons would balance out the SIA-induced probability shift that the presumptuous philosopher is advocating.

These alleged reasons, whatever they are, couldn't stem from the Doomsday argument, since the presumptuous philosopher is lacking crucial empirical information on which the Doomsday argument relies, viz. what his own sequential position is in the population whose size he is trying to predict. For we can suppose that the philosopher and his physicist interlocutors are clueless about their birth ranks in the class of all observers in the cosmos. In this respect, they would be in the same situation as you and I, who are, of course, likewise ignorant about how many observers came into existence before we did somewhere in the cosmos. (We could additionally make the assumption, without destroying the coherence of the thought experiment, that the philosopher and the physicists don't know their sequential position in the human species. The reason why this assumption is not made in the original thought experiment is that it is not needed. Information about one's human birth rank would not suffice, even if the basic form of Doomsday argument were correct, for deriving a prediction about the total number of

observers in the cosmos; for in order to do that, one would need to know one's *cosmic* birth rank, one's sequential position among *all* observers, not just the human ones.¹⁰⁾

It is hard to see what other kind of justification there could be for thinking a theory a priori unlikely in proportion to the size of the universe it proposes. Maybe one could attempt to base a case on some variant of Occam's razor. We could formulate a new simplicity principle stating that one must not postulate more space or matter than is minimally necessary to account for known phenomena, and appeal to this as the reason for Olum's new prior probability function (which we can dub the "OLUM-prior").

This proposal is problematic for several reasons. First, Occam's razor and various simplicity principles used in science are not so much injunctions against postulating excess matter (or space) as against excess theory. Simplicity and parsimony usually means economy of independent assumptions, avoiding proliferation of unrelated explanatory mechanisms, keeping the number of free parameters as small as possible, not introducing too many ad hoc clauses to account for recalcitrant findings, keeping the basic ontology neat and slim, and things of this sort. It is not generally taken to imply an aversion to thinking that the world is very big or that there are vast numbers of electrons, photons etc. out there.¹¹

Second, if we look at the case of cosmological theory, which is the scientific field that the presumptuous philosopher hopes to influence, we see that spatial or material parsimony is emphatically not part of established best practice in this discipline. All cosmologists agree that the cosmos is enormously large; a majority probably thinks it is infinite and that our universe is just one in a vast ensemble of physically real universes (in particular in the currently popular inflationary scenarios). Among those best placed to judge these things, there is no sign of any preference for smaller universes and one detects no epistemic compunction about hypothesizing that the world is very, very big.

¹⁰ Bradley Monton, in a paper forthcoming in this journal, argues that the Doomsday argument is applicable in the absence of knowledge about one's birth rank. We disagree but lack the space here for a detailed analysis.

¹¹ For instance, this is generally not used as an argument against many-worlds versions of quantum mechanics, to which we shall return shortly.

Third, the reference to “super-duper symmetry” in the thought experiment was meant to suggest a situation where relatively clear-cut theoretical considerations existed for assigning T_1 and T_2 roughly similar prior probabilities. Unless the coherence of this possibility is disputed, we can use the hypothetical case presented in the gedanken as a test bed for the plausibility of SIA, whatever our actual circumstances may be in regard to the OLUM-prior.

But suppose, all this notwithstanding, that our a priori credence distributions ought to be skewed against hypotheses that say that the world is very big, and that this relation is one of proportionality: a universe twice as big gets, other things equal, half as much prior probability. Now, if we adopt this OLUM-prior, we can counter the presumptuous philosopher’s argument, not by negating SIA, but by pointing out that the SIA merely serves (when combined with the fact that you exist) to compensate for the OLUM-prior’s favoring of smaller worlds. T_1 and T_2 end up with equal probabilities, as the cosmologists in the gedanken thought all along. (First we were asked to buy SIA as an insurance against the Doomsday argument; now we are asked to buy the OLUM-prior to insure us against SIA.)

It may look as though the only effect of the OLUM-prior is to cancel out the effect of SIA. This is not quite the case. The reason is that SIA speaks about *population size* while the OLUM-prior speaks about *spatial size*. In *Presumptuous Philosopher*, this distinction made no difference because there we supposed that we had background knowledge to the effect that the density of the observer-population was the same according to both theories; T_2 said that the cosmos was both bigger and had more observers than did T_1 . We can drive a wedge between SIA and the OLUM-prior, however, by considering a variation of the thought experiment where the two theories disagree about the population density.

To this end, consider *Presumptuous Philosopher II*:

The year is now 2200, and more careful study in the intervening century has revealed that the super-duper symmetry considerations actually do not favor T_1 and T_2 , but instead T_3 and T_4 , and they are neutral as between these two new theories. T_3 and T_4 instantiate theoretical virtues such as simplicity, fit with

evidence, etc. to similar degrees, and contemporary cosmologists therefore assign them roughly equal credences. One difference between T_3 and T_4 is that the density of observers is a trillion times greater if T_4 is true than if T_3 is true. This difference, however, is not directly observable, for both theories agree that the density of civilizations is so low that the probability that there should be another within our horizon is negligible. The theories also agree about the size of the cosmos, and, moreover, that the size of the cosmos is so exceedingly vast that whichever of the theories is true there was virtually no probability that it should contain less than many quintillions of civilizations. There is also another more subtle difference between the predictions that the theories make, a difference that is testable in an easy experiment that cosmologists are just about to perform. Again the presumptuous philosopher, who thanks to a breakthrough in life-extension research is still alive and well, makes an appearance and argues, on grounds of SIA, that the probability of T_3 being correct is one in a trillion; wherefore it is completely unnecessary to do the experiment.

Luckily, the physicists do not abort the experiment but instead offer the philosopher a bet on the outcome, agreeing to pay him one thousand dollars if the test comes out in favor of T_4 in return for ten thousand dollars if it favors T_3 . The philosopher gladly accepts. As it happens, the physicists win the bet and get ten thousand dollars. As there is a one-in-a-million chance that the experiment has yielded a misleading result, a second experiment is proposed to verify the first. Despite the setback, the philosopher's SIA-based confidence in T_4 is hardly perturbed; he still assigns a probability of merely one in a million to T_3 , so he accepts a repeat bet with the physicists. The presumptuous philosopher is making a fool of himself.

The OLUM-prior doesn't forestall this unpalatable consequence because T_3 and T_4 agree about the size of the cosmos. What they centrally disagree about is the number of observers, and it is this difference that, when conjoined with SIA, gives rise to the implausible conclusion that the philosopher is quixotically sticking by.

One way out of this predicament would be by modifying the OLUM-prior to one (call it the OLUM*-prior) that levies on hypotheses a probability tax that is proportional

to the number of observers they entail, rather than to the implied size of the universe. Starting with the OLUM*-prior, and then taking account the fact that you exist as SIA says you should, the two effects – the first arguing for few observers and the second for many observers – cancel each other out. In light of the combined illumination from the OLUM*-prior and SIA, people considering T_3 and T_4 would assign them similar degrees of credence, a bottom line that seems reasonable.

The problem with this solution (apart from the fact that it appeals to an assumption, the OLUM*-prior, that seems as lacking in justification as the original OLUM-prior) is that it proves too much. The OLUM*-prior “tames” the unruly SIA-induced probability shift simply by killing it. The probability distribution we get after taking both the OLUM*-prior and SIA into account is the same as we had before hearing of either. We have been offered two unsupported new principles whose net effect is nil. This brings us right back to where we started, for with no net epistemic subsidy of hypotheses that favor many observers, there is nothing to counterbalance the pessimistic probability shift in favor of few observers argued for by the Doomsday argument and in related paradoxes, which are therefore left standing tall.

III. OBSERVATION SELECTION THEORY AND QUANTUM MECHANICS

Readers of Olum’s paper could be forgiven for coming away with the impression that we accept all the strange consequences of an unrestricted use of SSA.¹² This is not the case. We agree with Olum that these consequences are very hard to accept. But rather than taking this as a justification for SIA, which, as we’ve argued above, comes with its own set of equally unacceptable consequences, we see it as a reason for replacing SSA with some other principle that does a better job at guiding the use of indexical information.

Such a principle needs to be part of a general theory of observation selection effects, a theory that explains how to reason when indexical information is linked to non-indexical hypotheses (and vice versa) and how to avoid anthropic biases in the course of such reasoning. The theory must cater to legitimate scientific needs as well as being paradox-free in philosophical thought experiments. One of us has recently presented such

¹² One of us has elaborated on these elsewhere (Bostrom 2001).

a theory and placed it in a Bayesian framework.¹³ It shows that by taking *more* indexical information into account than SSA does (SSA considers only information about which observer you are, but you also have information about e.g. *which temporal part of this observer* you are at the current moment) it is possible to relativize your reference class so that it may contain different observers at different times, depending partly on your epistemic situation on the occasion. SSA, therefore, describes the correct way of assigning probabilities only in certain special cases. This enables us to avoid the original paradoxes without invoking SIA. In particular, the Doomsday argument is shown to be inconclusive in that it depends on particular assumptions about the part of one's subjective prior probability distribution that has to do with indexical information – assumptions that one is free to reject, and indeed, arguably, ought to reject in light of their strongly counterintuitive consequences.

This theory of observation selection effects has applications for both philosophy and several scientific areas including cosmology, evolution theory, thermodynamics, traffic analysis, game theory problems involving imperfect recall, and quantum physics. As a brief illustration, we can consider the application to quantum physics that Olum refers to in his paper. In this application (for which we're indebted to Don N. Page¹⁴), the theory concurs with what we would get if we simply used SSA with the universal reference class (the one containing all observers), so Olum's remarks on this point are pertinent to the view we actually espouse. The idea is that we can in principle drive an empirical wedge between single-history versions and many-worlds versions of quantum theory by considering cases where different numbers of observers come to exist depending on the outcome of some early quantum measurement-like interaction (perhaps some random factor in the symmetry-breaking that determined the strengths of the force constants).

¹³ Bostrom (2002).

¹⁴ D. N. Page, "Can Quantum Cosmology Give Observational Consequences of Many-Worlds Quantum Theory?" in *General Relativity and Relativistic Astrophysics, Eighth Canadian Conference, Montreal, Quebec*, eds. C. P. Burgess and R. C. Myers, 225-32 (Melville, New York: American Institute of Physics, 1999); also personal communications.

The point can be made most simply by considering a quantum cosmological toy model:

World 1: Observers; measure or probability 10^{-30}

World 2: No observers; measure or probability $1-10^{-30}$

Given these measures, the single-history version predicts with overwhelming probability ($P = 1 - 10^{-30}$) that World 2 would be the (only) realized world. (How we could figure out the measure of “worlds”, i.e. of whole branches of the universal wave function, is a separate problem that need not concern us here; we can think of the proposed test as being not of versions of quantum physics itself but of versions of quantum physics taken in conjunction with some particular quantum cosmological theory that specifies these measures). If we exist, and consequently World 1 has been realized, this gives us strong reasons for rejecting the single-history theory, given this particular toy model. By contrast, on the many-worlds version, both World 1 and World 2 exist, and since World 2 has no observers, what is predicted (by SSA) is that we should observe World 1, notwithstanding its very low measure. In this example, if the choice is between the single-history and the many-worlds versions, we should therefore accept the latter.

Here’s another toy model:

World A: 10^{10} observers; measure or probability $1-10^{-30}$

World B: 10^{50} observers; measure or probability 10^{-30}

In this model, finding that we are in World B does not logically refute the single-history version, but it does make it extremely improbable. For the single-history version gives a conditional probability of 10^{-30} to us observing World B. The many-worlds version, on the other hand, gives a conditional probability of approximately 1 to us observing World B.¹⁵ Provided, then, that our subjective prior probabilities for the single-history and the

¹⁵
$$P = \frac{10^{50} \cdot 10^{-30}}{10^{50} \cdot 10^{-30} + 10^{10} \cdot (1 - 10^{-30})} \approx 1$$

many-worlds versions are in the same (very big) ballpark, we should in this case again accept the latter. (The opposite would hold, of course, if we found that we are living in World A.)

Olum charges that “the fact that [this] treatment yields ambiguous predictions argues against its use in evaluating theories of quantum cosmology” (p. 183). In our view, however, keeping single-history and many-worlds version of quantum theory empirically inseparable is not an adequacy condition on theories of observation selection effects. On the contrary, it would be quite interesting if we could have a methodology that makes the above kind of consideration intelligible and that may enable us to tease these very different conceptions of the world apart observationally. (There are several other recent suggestions¹⁶ for how one could make such a discrimination, which are neatly complemented by the present considerations.)

To sum up: Olum and we agree that an unrestricted use of SSA (with the universal reference class and in the absence of SIA) has a number of counterintuitive consequences. SIA cancels these consequences. SIA, however, as the *Presumptuous Philosopher* gedanken reveals, has counterintuitive consequences of its own that seem just as bad as the ones it was meant to defeat. Olum suggests that we can fix this problem by adopting what we have dubbed the OLUM-prior. We have argued that the OLUM-prior is unjustified and that it doesn’t fix the problem anyway (as revealed by *Presumptuous Philosopher II*). What we called the OLUM*-prior would fix the problem, but it is equally unjustified, and, crucially, adopting it has the effect of canceling SIA, bringing back us back to square one. SIA is a blind alley.

Threatened by paradoxes on both sides, whether we accept or reject SIA, it may appear as if we have a crisis. But in fact, what we have is a philosophical opportunity – we have some powerful theory-constraints that can be used to evaluate reasoning about information that has an indexical component. If we develop a theory of observation selection effects that heed these constraints along with many other desiderata and criteria derived from both philosophical and scientific applications of anthropic reasoning, we

¹⁶ R. Plaga, “Proposal for an experimental test of the many-worlds interpretation of quantum mechanics,” *Found. Phys.* 27, 559 (1997).

can hope to get serious theoretical leverage – methodological techniques and results that can turn over boulders of ignorance and reveal something new about the world.