Alternatives to the Self-Indication Assumption Are Doomed

Abstract: The self-indication assumption (SIA) claims that given that one exists, one should think that the universe has many people, for a universe that has more people is more likely to contain any particular person. SIA is attractive to many because it diffuses the infamous doomsday argument, and avoids the problems of its main rival, the self-sampling assumption (SSA), which instructs one to reason as if they're randomly selected from the people in their reference class. Here, I will go further than the traditional argument for SIA from its ability to avoid the doomsday argument: I will show that any view other than SIA must accept the soundness of the doomsday argument and certain even stranger conclusions. Additionally, I'll argue that the main objection to SIA, the presumptuous philosopher case, is unconvincing. Together then, these considerations give one a strong reason to adopt SIA.[1]

1 Introduction

The doomsday argument (DA) claims that one gets strong evidence, from their existence, that their civilization won't have many more observers. The basic idea is the following: suppose the human race lasted billions of years, containing quadrillions of people. Well, then it would be quite odd that I'm so early on. What are the odds that I'd be one of the first few billion out of quadrillions? Very low.

Such an argument has had its fair share of defenders (e.g. Bostrom, 2013; Leslie, 2002) since its formulation by Carter (1983). Fortunately, the DA is only salvageable if the self-indication assumption (SIA) is false (or, what Bostrom, 2013, p.65 refers to as SSA&SIA). SIA states "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," (Bostrom, 2013, p.66). It's been demonstrated elsewhere (see Kopf, Krtous, & Page 1994) that the SIA precisely cancels out the doomsday argument for the following reason: if one theory predicts the existence of 1,000 people and the other of 100 people, and I'm in the

[1] This work is a draft that has not undergone peer review. The version of record of this article is published in Synthese Volume 204, article number 23 https://link.springer.com/article/10.1007/s11229-024-04686-w

first 100 people, while the first hypothesis makes it ten times as likely that I'd exist, it makes is one-tenth as likely that I'd be within the first 100 people, so the probabilities exactly cancel out.

Merely resolving one puzzle isn't necessarily enough to make a view plausible. Some have argued (e.g. Ćirković, 2001) that though the self-indication assumption may diffuse the doomsday argument, there is little else to be said for it, and certain improbable results follow from it, meaning that it's not worth accepting. Whether this is so is controversial (for an alternative view, see, for example, Olum (2002) and Cushman (2019)). If, however, it could be demonstrated that the DA could only be resolved by adopting SIA, that would be quite a strong reason to adopt SIA. Here, I'll attempt to show both that, and that SIA is the only way to avoid certain other, even stranger conclusions.

The plan of the paper is as follows: first I'll present three implausible results that resemble the doomsday argument. Then I'll argue that SIA is the only way to avoid these results. After this, I'll argue that these same sorts of results arise for any view that instructs one to half in the sleeping beauty problem. Finally, I'll address the main objection to SIA and rebut the charge that it implies the same sort of implausible, doomsday-style reasoning.

2 Presenting a few paradoxes

Some are inclined simply to accept the DA, seemingly undeterred by the surprising claim that we could get strong evidence that there won't be too many future humans from the armchair. Such proponents think that the fact that we find ourselves so early does give us strong reason to think humanity will end soon. There's nothing illegitimate about figuring out things from the armchair in general, but there is something suspicious about the notion that one's current existence gives them a reason to think that the future won't have many observers. Whether this is a kosher kind of anthropic reasoning is contested. The DA is one of the most debated areas in anthropics: whether we should expect to go extinct soon is important. Additionally, as I'll discuss later, accepting SIA has certain important implications for the ultimate view one should have of fundamental reality. Here, I'll argue that the DA should be rejected because it's so implausible. To see this, consider the following analogous case:

Future Graham's Number: there is some machine that will ensure a future of Graham's number years, containing more than Graham's number observers unless the next 100 hands in some particular poker game all contain royal flushes.

One can fill in the details of the case however they wish. Perhaps there is a very powerful God who, in his omnipotence, has made a machine that will protect a civilization from destruction until they birth Graham's number babies, but only if the next 100 hands in poker don't contain royal flushes. Or perhaps we could imagine that physics works differently, and that the civilization will naturally last for Graham's number years unless destroyed by the machine, which will destroy it if and only if 100 hands in poker all contain royal flushes. While the scenario may be bizarre, there's nothing contradictory or ill-formed about it, and the point is primarily conceptual.

Graham's number is a ridiculously large number, large enough to swamp any number that can be easily reached by exponentiation. In such a scenario, it seems one has very strong reason to expect the future to contain Graham's number people. After all, for the future not to have Graham's number people, something unbelievably improbable must occur, namely, a particular poker game must have 100 consecutive royal flushes. This is ridiculously unlikely. However, defenders of the DA must reject this verdict, thinking that probably the next 100 hands will, in fact, contain royal flushes. For if the future has over Graham's number people, then the odds that some observer should give to being so early—in the first 110 billion out of Graham's number people—are even lower than the odds of a game of poker having 100 consecutive royal flushes. This is a pill too difficult, in my judgment, to swallow.[2]

This case is not the only one that bears out the oddness of the DA. A related case is Serpent's Advice, from Bostrom (2001, p.366):

Eve and Adam, the first two humans, knew that if they gratified their flesh, Eve might bear a child, and if she did, they would be expelled from Eden and would go on to spawn billions of

[2] One might doubt the usefulness of such farfetched thought experiments, but exploring this in any detail is outside the scope of this paper.

> progeny that would cover the Earth with misery. One day a serpent approached the couple and spoke thus: "Pssst! If you embrace each other, then either Eve will have a child or she won't. If she has a child then you will have been among the first two out of billions of people. Your conditional probability of having such early positions in the human species given this hypothesis is extremely small. If, one the other hand, Eve doesn't become pregnant then the conditional probability, given this, of you being among the first two humans is equal to one. By Bayes's theorem, the risk that she will have a child is less than one in a billion. Go forth, indulge, and worry not about the consequences!"

A third case, even more surprising and deeply counterintuitive, also derives from Bostrom (2001, p.367):

> Assume as before that Adam and Eve were once the only people and that they know for certain that if they have a child they will be driven out of Eden and will have billions of descendants. But this time they have a foolproof way of generating a child, perhaps using advanced in vitro fertilization. Adam is tired of getting up every morning to go hunting. Together with Eve, he devises the following scheme: They form the firm intention that unless a wounded deer limps by their cave, they will have a child. Adam can then put his feet up and rationally expect with near certainty that a wounded dear – an easy target for his spear – will soon stroll by.

Surely, we cannot just accept these wildly implausible conclusions! After forming the firm intention to procreate unless a deer drops dead at Adam's feet, Adam would be irrational to expect that, in fact, such an animal would fall dead. Some disagree, however. Bostrom himself (2001, p.373-374) argues that in such a case, it is justified for Adam to expect the deer to fall dead at his feet, but is careful to note that this would not involve anomalous causation. The viability of such a view is contested: see, for instance, Ćirković (2004). Bostrom claims that his result is acceptable because "it is in the nature of probabilistic reasoning that some people using it, if they are in unusual circumstances, will be misled," (p.373).

Bostrom is, of course, correct that it's not a decisive objection to a theory of probability that it implies that some people will sometimes be misled. However, a theory of probability should provide people with reasonable expectations. If, for instance, some theory implied that one who is feeling lucky should assign a 90% probability to their getting heads in a coin toss, this would be a decisive objection to the theory. This would not be because it implies that they'd be misled, but instead, because it implies that their credence in some proposition should be different from what it actually should be.

Similarly, the objection to the DA is not that it implies counterintuitively that Adam and Eve might be wrong. It's that it's plainly obvious that they should not be very confident that a deer will drop dead at their feet. Even for the first two people, ensuring that pregnancy would result in lots of future descendants is not a safe form of birth control. Similarly, in Future Graham's Number, one shouldn't be almost totally certain that the poker game will have 100 consecutive royal flushes.

It's hard to stress just how counterintuitive DA-style reasoning is. On such an account, if in the actual world, one could guarantee that were they to get pregnant, civilization would last a very long time, such a method would be a very safe form of birth control. But this is clearly crazy!

A more promising and subtle alternative comes from Ćirković (2004). Just as to avoid the bilking paradox in time travel, according to which one could paradoxically go back in time and kill their grandfather, thus leading to them being unable to be born and assassinate their grandfather, one should think that they can't go back and kill their grandfather (for if they did so, they would not be around in the first place), so too can one avoid these paradoxes by saying that they are extremely unlikely to arise. In Serpent's Advice, for instance, perhaps if SIA is wrong, Adam should expect not to have sex in the first place, because he is confident that if he has sex, something improbable will have to happen (namely, Eve not getting pregnant). Similarly, in Lazy Adam, he should expect that even if the deer doesn't drop dead, he is unlikely to have many offspring.

On such views, the situation for the various characters in our thought experiment might be akin to a person who has received, with near certainty, the information that the future won't have many people. Thus, they shouldn't expect to get a multitude of royal flushes in Future Graham's Number, but should

instead expect something else to come up which makes it so that the future will not contain Graham's number people.

First of all, even if one accepts this, we could imagine much more limited cases where the most likely way that the future won't have many people is for some improbable chancy event to happen. For instance, imagine a nuclear device is created that will detonate and destroy the world only if a random number generator, which produces a number between 1 and 50, comes up 47. Imagine that this world is likely to have many people and face few threats from other sources, so the most likely way it would go extinct soon is for the random number generator to come up 47. On such views, one should expect the number to come up 47. But this is very strange! The oddness comes from the notion that one's credence in a future event should differ from its objective probability when the various outcomes of the event will be different only in the future.

Second of all, even though these cases are extremely strange, surely they are conceivable! We can imagine agents with such detailed knowledge of the world, perhaps because their physics is very simple or they received extremely powerful divine revelation, that they can be extremely confident in these predictions. We can even imagine that these agents are justified and yet mistaken, perhaps because they are deluded by a God or some other being to not even consider the possibility of the setup described failing to occur. In such cases, where agents think the set-up is as described as it is in Future Graham's Number, for instance, even though the set-up is strange, I submit that they should not confidently predict that they'll get ten consecutive royal flushes. Not only are they mistaken that the scenario occurred, they are wrong about how to respond to the scenario if it occurred.

Third, even if one thinks that a scenario like this could not actually arise, we can still entertain scenarios where people have credences that they couldn't rationally have. For example, we should accept that one who is certain in a proposition should be willing to bet on it at 1:1 odds, even if we think no agent could ever rationally be justified in being certain in a belief. I submit, therefore, that even if the scenario described produces impossible claims of justification, this is not a mark against it.

Fourth, even if the takeaway of these scenarios is that they are unlikely to arise, that is a very strange result. In the Lazy Adam case, for instance, Ćirković's reasoning leads to the conclusion that after forming the firm intention to procreate, Adam and Eve should both expect not to procreate. Yet that is very strange: how could your expectations of what future act you're likely to engage in depend on its future consequences? It's hard to believe that one should expect themself not to have sex simply because if they had sex, they'd have many offspring.

3 Every view other than SIA must bite these bullets

The last section argued that one should not accept the DA or any of the similar verdicts in the various cases described. Accepting that a deer is, in fact, likely to drop dead at Adam's feet is absurd. But each of these follow quite straightforwardly from basic probabilistic reasoning if SIA is false. Thus, the last section described some troubling results, and this one will argue that one must accept those results if they reject SIA. As a result, this serves as a strong argument for SIA, showing that it's the only way to avoid certain mistaken results.

Let's first consider the DA. If SIA is false, then a universe that contains many people is not one I should think myself likelier to be in, all else equal, than one that lasts a shorter time and contains a much smaller number of people, given that views other than SIA must reject the claim that a universe with more people is more likely to have me (after all, that is the distinct claim made by SIA, and thus something that alternatives to SIA must reject). Let's call the first universe W1 and the second W2. Assume that W1 has 100 quadrillion people, W2 has 100 billion people only, and I'm in the first 100 billion people. Given that I could have been born at any time (an assumption which I'll discuss more later), on W2, the odds that I'd be one of the first 100 billion people are 100%, while on W1 the odds are 1 in 1 million (because 100 quadrillion people can be divided into 1 million intervals of 100 billion people, and I could have been born in any of those intervals.) Thus, my early existence is very strong evidence for the existence of W2 over W1.

To put an intuitive gloss on this, if a world with more people is no more likely to contain me, and a world with fewer people entails that I'd be early, while a world with more people does not, then my existence at some early time is very strong evidence the future won't have very many people. To consider an analogy, suppose that there are two hypotheses that make it equally likely that I'd exist, but one of them makes it more likely that I'd be born in a yellow room. Upon discovering that I am in a yellow room, I get evidence for the hypothesis that makes it more likely that I'd be in a yellow room. Being born early is relevantly analogous to being born in a yellow room.

Notably, SIA agrees that upon finding out that I'm the first person, I get very strong evidence that there won't be very many future people. But crucially, SIA gives reason to think that, prior to finding that out, I should think that the hypothesis that the future will have many people is much more likely than the hypothesis that it won't have many people, so after updating on both facts, I should end up indifferent between the two hypotheses. My prior in W1 starts out 1 million times higher than my prior in W2, all else equal, and so after receiving evidence that favors W2 by a factor of 1 million, I end up indifferent between the two possibilities. Thus, SIA doesn't imply any aberrant predictions of future events.

Crucially, only SIA can avoid these troubling paradoxes. This is because there's a straightforward update in favor of theories on which there are fewer people from finding out one is one of the first few people. If the universe has fewer total people, it's likelier that one would be one of the first few people than if there are many people. Only if one thinks that more people existing is inherently more likely to explain one's existence do they have the tools to cancel out this update. But SIA is simply the view according to which there's a straightforward update in favor of theories on which there are more people, from the fact that one exists.

The same basic point about why alternatives to SIA must accept the DA applies to the other two cases. The Serpant's Advice and Lazy Adam cases are just other instances of the same basic way of reasoning. Just as I should be surprised to find myself so early if a universe with more people is no more likely to contain me, Adam and Eve should be surprised to find themselves so early unless a bigger

universe is more likely to contain them, and thus can be confident that there won't be many people, and consequently, that Eve won't get pregnant.

It might be worried that this assumes something controversial about reference classes or origin essentialism. In the argument, I have claimed that I could have been born at some other time. But perhaps, if origin essentialism is true, I could not have been born at some other time, for the circumstances resulting in my birth are a requirement for my existence. Similarly, the self-sampling assumption instructs one to reason as if they were randomly selected from the class of all possible observers in their reference class (see Bostrom, 2013, for a defense of it). Perhaps the beings in the future are not in my reference class, and thus I shouldn't reason as if I could have been one of them.

Note here that the suggestion that I'm addressing is the idea that future people, even if they resemble us, would not be in the same reference class. Perhaps one's reference class only includes past observers. If one doesn't hold a view like this, but just thinks that the future people will be so different from us—perhaps radically different superhumans—then they should still accept the conclusion of the DA, that the future will likely not have too many people that are like you. On such a view, however, this is not necessarily because civilization will die off soon but instead because future people will be so different as not to be in one's reference class. The view that avoids the DA, therefore, isn't merely one that claims future people will be very different, but instead one that claims that future people are categorically barred from being in one's reference class.

Beginning with the origin essentialism point, this wrongly conflates epistemic possibility with metaphysical possibility. It's true that if origin essentialism is true then it would be metaphysically impossible for me to be born at any time other than the time that I was actually born. But it's still true as an epistemic point that the odds I'd be born so early are low. Even if whenever I'm born it will be necessary that I'm born then, the prior probability that it will be necessary that I'd be born at an early point is low.

To see this, imagine that a necessitarian (one who thinks everything that happens is necessary) got ten consecutive royal flushes in poker. You may think that they're probably cheating. You should still

think that even though on their view, their getting lucky is necessary. Even if necessitarianism is true, the probability that it would be necessary that they'd get ten royal flushes is very low (in an epistemic sense). Similarly, even if whenever I happen to exist is necessary for me to exist, the odds that I'd be so early are low.

Modifying the reference class while adopting SSA might seem to resolve the puzzle. SSA says one should reason as if they're a random member of their reference class, where the reference class is some group of beings roughly similar to oneself. Perhaps, for instance, I should reason as if I'm a random human or a random intelligent conscious agent (for a discussion of reference classes see Bostrom, 2013). Thus, if one thinks one should reason as if they're a random member of their reference class, and their reference class doesn't include future humans, then they have no reason to accept the DA. Such a view does seem strangely parochial, however: what's so special about current people?

Restricting the reference class while adopting SSA does solve the puzzle, but it only does this by ignoring the straightforward probabilistic reasoning. If one is previously indifferent between two hypotheses, but one of them entails some fact about the world and the other does not, then that gives them evidence for the theory that entails the fact. So if one has no reason to accept that worlds with more people are more likely, but worlds with more people don't entail that they're early, while worlds with fewer people do, then their earliness favors worlds with fewer people. But from this, the DA and associated puzzles follow.

To illustrate the point, and show that neither modifying the reference class nor accepting origin essentialism is enough to escape the puzzle, suppose that Eve has not been created yet—all there is is Adam, yet he is very ignorant. He doesn't know who he is, that he's the first human, what year it is, or much else. Crucially, he rejects the self-indication assumption. Adam is considering two possibilities: first, that the first human is such as to, at some point (possibly in the past, possibly in the future, for he doesn't know what year it is), have enough descendants to make the total number of people 10 billion, and second that the original human is such as to have no descendants (I say is such as to have to denote that he doesn't know if this humans procreation has occurred in the past or in the future). Before

considering the fact of his existence, he thinks each possibility has a 50% probability. Given that he has no special evidence that he's the first person, he reasons in the following way: "if the first person is such as to have enough offspring to make there be 10 billion people, then it's very unlikely that I'd be the first person, for I could either be the first person or any of the other people. In contrast, if the first person is such as to have no descendants, then I'd be guaranteed to be the first person. So if I ever discover that I am, in fact, the first person, then I will get very strong evidence for the second possibility—that the first person shall never have offspring."

Suddenly, a voice out of the heavens booms "you are Adam, the first human." Upon finding this out, Adam gains very strong evidence for the hypothesis that the first human won't have many descendants, and because he knows he is the first human, he has very strong evidence that *he* will not have many descendants. Thus, even if God tells him that unless 20 coins all come up heads, God will fashion from his ribs a wife, with whom he will make billions of offspring, he should be confident that the 20 coins will all come up heads. Yet this is quite implausible.

From here, we can reconstruct each of the cases mentioned in section 2. For instance, now Adam can be confident that Eve won't get pregnant if he knows that were she to get pregnant, she'd have a vast number of offspring, and by setting up a process that would produce many offspring unless a deer drops dead at his feet, he could be quite confident that the deer would drop dead at his feet. Similarly, Adam can be confident that he'll get consecutive royal flushes in a poker game that has not been played yet, if God agrees to make a huge number of people unless Adam gets many royal flushes.

The basic thing that is surprising is that, if one ponders the odds of their existence prior to knowing their birth rank, and then discovers that they're early, on every view other than SIA, they have a reason to think it's very unlikely that the world will have many people. This gives them a bizarre ability to be confident that future events won't turn out in a way that facilitates the creation of many people.

In response to this, one might claim that prior to finding out one is so early they should be indifferent between hypotheses that say the world will have many people and ones that say it has few people, ceteris paribus, and think it's unlikely, conditional on the universe having many people, that

they'd be early, but after finding out that they're early, they should remain indifferent between the hypotheses (for a view like this, see Builes, 2020, ch.2). But this straightforwardly violates probabilistic reasoning; theories on which there are few people entails one is early, but theories on which there are many don't, so finding out that one is early must favor theories one which one is early. In addition, this means that one should expect their credence in the future having many people to go up after discovering their birth rank, which violates the rule of probability that one shouldn't expect their credence to rise after seeing some evidence (for if they did expect to get evidence that raises their credence, then their credence should be higher now).[3] If one doesn't update, based on their being so early, against hypotheses with many people, then they should expect to, on average, grow more confident after finding out whether they're early. If they're late then the world must have many people, so if being early doesn't give evidence that there won't be many people, then upon finding out one's birth rank, one can only grow more confident in a universe containing more future people. This is the probabilistic version of heads I win, tails we're even.

Cirkovic (2001, p.7) claims that worries about alternatives to SIA implying that the consequences of some event determine its probability rely on conflating objective and subjective probability. The objective odds of the deer dropping dead at one's feet are unaffected by the impact that such an event would have on the number of people. However, the odds some observer should give to a deer dropping dead will be affected by anthropic considerations, and thus by the impact that this would have on the number of people.

But this isn't where the intuition comes from. The intuition is that the probability of some future event shouldn't depend on its future consequences. Sometimes one has reason to think that past events

[3] There might seem to be exceptions to this. For instance, you should expect your credence tomorrow in the proposition that you didn't die today to be higher than it is today. However, in such cases, you'll only be able to observe the future chance process if it turns out a certain way. In cases where you'll observe the future outcome whatever it is, one shouldn't predictably update.

occurred some way, because them occurring that way more strongly predicts the existence of various observed phenomena. For instance, it's perfectly reasonable for me to be confident that Alexander the Great existed, even though that has a low prior probability, because it predicts the existence of evidence that I've already observed. But future events can't predict evidence you've already observed, for they are in the future, so one's credence in some future event occurring shouldn't depend on the consequences that it would have.

4 Halfers are doomed!

The considerations presented in previous sections can be mirrored to resolve the sleeping beauty problem. Specifically, they can show quite a significant problem for the halfing view. Seeing as SIA implies thirding, showing that these considerations rule out the halfing view provides support for SIA. In addition, this section shows another significant result of my basic argument: views according to which more person-moments existing are not inherently more likely enable aberrant predictions of chancy events. The sleeping beauty problem was nicely summarized in an early paper by Elga (2002, p.143):

> Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you to back to sleep with a drug that makes you forget that waking. When you are first awakened, to what degree ought you believe that the outcome of the coin toss is Heads?

There are two main answers: a half and a third. Here, I will argue that an argument similar to that employed in the previous section can be employed against the halfing view. To see this, consider a case by Haig (2016), like the one described by Elga, except if the coin comes up tails, the person will be awoken an infinite number of times. Upon waking up, halfers think one should be 50/50 between the options, thirders think that one should be certain that the coin came up tails.

Let's modify the case in the following way. What determines if the person will be awoken an infinite number of times is not the flip of a coin but instead, the results of a poker game to be played at the

end of the first day. If, in the poker game, ten consecutive royal flushes are dealt, then the person won't be awoken up again, while if there are not ten consecutive royal flushes, then the person will be awoken infinitely.

Upon waking up, one should think that, almost certainly, they'll be awoken an infinite number of times, because the odds of ten consecutive royal flushes are quite low. But then suppose that after waking up, one finds it's the first day. Upon learning this, they should be certain that they will not be awoken infinitely. For on the hypothesis that they will be awoken infinitely, the odds it would be the first day now are zero—for it is equally likely to be any of the infinite days. But then this means that they can be confident that the poker game that they will play at the end of the day today will involve them getting ten royal flushes, simply because if it doesn't, then they'll be awoken many times in the future. This is surely wrong! Such reasoning should not justify confident conclusions about fantastically unlikely outcomes in poker games that have not yet been played. The likelihood of some outcome in a poker game shouldn't depend, in this way, on its consequences.

5 Does SIA have the same problem?

This section will thus address arguments for why SIA might imply something like the Doomsday result, as well as address the primary objection to SIA in the literature. The earlier sections provided an argument for SIA, this one addresses the main objections to it. Together, then, they serve to make a powerful case for SIA, showing both that the main objections to it are unconvincing and that only it can avoid certain troubling paradoxes.

It might be worried that SIA implies the same kind of doomsday argument. This has been argued by Grace (2010). Grace claims that, to resolve the Fermi paradox, which is the puzzle of why we don't come across many aliens, we must think that there is something that makes life like us rare. This can either be an early filter or a late filter. An early filter is something that would make the limiting factor on life like us be something that occurred long in the past. For example, an early filter might be something that makes eukaryotic life, life at all, or multicellular life, rare.

A late filter, in contrast, is something that occurs late in the stage of development and limits the development of civilizations. For instance, perhaps it's very difficult to develop an intergalactic civilization because that requires building artificial intelligence that necessarily wipes out a civilization. Grace claims that anthropic considerations, on SIA, give us reason to think that the filter is late because those worlds predict the existence of more observers like us.

Grace correctly identifies that if there are two otherwise identical worlds, but one of them posits an earlier filter, a believer in SIA should think that the one with the later filter is more likely. But this doesn't seem especially counterintuitive—a world with a later filter will have more people, so one should expect to be more likely to exist in a world with a later filter. This is perhaps bad news for us, but is not a reason to reject SIA. Nothing like the Future Graham's Number case or Lazy Adam can be constructed for the view.

However, it's not clear that this means we're likely doomed. First, SIA supports thinking that the universe contains infinite people. But if the universe contains infinite people, then the same number of people will exist whether the filter is early or late. Second, universes with late filters that don't eliminate many civilizations will have more people. Perhaps they can construct vast simulations of the early life. Third, SIA gives one strong reason to think that every possible person exists, for theories where every possible person exists, predict the existence of more people than other theories. But if every possible person is created, SIA makes no predictions about where those people would be located, and thus it doesn't make any prediction about which percent of people would be in our local universe, rather than some other universe. All of these proposals are rather speculative but plausible enough that SIA doesn't lead one to confidently predict doom.

One might worry that this is a major downside of SIA. It predicts that all possible people exist! But surely all possible people don't exist! So SIA must be wrong, claims the critic. In response, however, let me first note that, as I will show later, every view of anthropics will imply certain major shifts in how we should view the world in many cases. Thus, the question is simply whether SIA implies

a particularly surprising or implausible picture of what the world looks like (namely, that it has every possible person).

I think, however, that SIA's picture of reality is not so surprising once one has adopted the mindset of a proponent of SIA. As Descartes famously claimed to deduce from his thinking the existence of God, so too does SIA claim to deduce from your thinking the existence of every possible person.[4] If every possible person exists, then one's existence is guaranteed, while if they do not, then one's existence is very unlikely: what are the odds that *you* would exist out of all the possible people that could? Just as it's more likely that someone will name some randomly selected natural number if they name all of the natural numbers than if they just name some of them, it's more likely that you would exist if all possible people do than if only some do.

It might still be worried that SIA implies some view of reality that we have reason to reject. If there is no plausible picture of reality on which every possible person exists, then this might make SIA untenable. But there are various views of reality with some independent plausibility—adopted by people who don't endorse SIA for reasons having nothing to do with SIA—that might very well predict the existence of every possible person.

For instance, some people—especially Lewis (1986)—have argued for modal realism, according to which every possible world concretely exists. On such a view, independently motivated, every possible person is guaranteed to exist. Similarly, many people believe in the existence of God. Arguably, a good God would create all possible people (for a defense of this, see Adelstein & Wollen, 2024). Perhaps also the view of Tegmark (2008) that every possible mathematical structure is instantiated predicts the existence of every possible person. Thus, while SIA will narrow the range of plausible views of the nature of fundamental reality, there are several different, independently motivated views that naturally predict the existence of every possible person.

[4] Credit to an anonymous reviewer for this insight.

Furthermore, a similar problem will apply to alternatives to SIA. If one adopts SSA, for example, then unless they're a solipsist, they should think that their reference class is very large including, at least, billions of other humans. Given that SSA privileges theories according to which one's reference class is smaller, the fact that fundamental reality has a big reference class is extremely surprising if SSA is right. SIA is, indeed, one of the few views of anthropics on which the anthropically most probable picture of fundamental reality might actually be right!

Perhaps the more troubling case for SIA is the infamous presumptuous philosopher case, helpfully explained by Bostrom (2013, p.124):

> It is the year 2100 and physicists have narrowed down the search for a theory of everything to only two remaining plausible candidate theories, T1 and T2 (using considerations from super-duper symmetry). According to T1 the world is very, very big but finite and there are a total of a trillion trillion observers in the cosmos. According to T2, the world is very, very, very big but finite and there are a trillion trillion trillion observers. The super-duper symmetry considerations are indifferent 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 T2 is about a trillion times more likely to be true than T1!"

It might be argued that the presumptuous philosopher case is a solid reason to reject SIA—one that is the twin of the various paradoxes supporting SIA. The argument for SIA noted that alternatives to SIA imply that we can be very sure that improbable circumstances will eventuate bringing about the total existence of few people. But the presumptuous philosopher notes that SIA implies we can be very sure that improbable events *have* eventuated bringing about many people. For instance, on SIA if one knew that 100 years ago, a poker game was played, and an otherwise finite universe became infinite only if many royal flushes were dealt, then one should think that many royal flushes were dealt.

But crucially, the scenarios are not the same. SIA does imply that one can be confident that some past event brought about the existence of many people. But it does not imply that one can be confident

that a *future* event will turn out some way, simply because that way will bring about the existence of many people (this is noted by Carlsmith, 2022, p.46-48 and Cushman, 2019). One has no reason to think that there will be large numbers of future people on SIA, because, even though the presence of more people makes one's existence more likely, it makes it less likely that one would exist so early, so the two probabilities precisely cancel out.

It might be objected that SIA's claim that one can be confident in the outcome of past events, based on the fact that their turning out one way would result in the existence of more people, is similarly wrong. How can I be so confident that a poker game played a hundred years ago had so many royal flushes, simply because it would mean that more people came to exist? But we often accept that improbable events happened in the past based on strong evidence in the present. For instance, I am confident that Caesar crossed the Rubicon, based on there being various features of the world that are far more likely if he crossed the Rubicon than if he didn't. Proponents of SIA reason in the same way—I can be confident that many royal flushes were dealt in the poker game because if that hadn't happened, I'd be very unlikely to exist.

For this reason, I don't even find the presumptuous philosopher result particularly counterintuitive. It's true that on anthropic grounds, according to SIA, worlds with more people are more likely to contain me, which implies I should have a higher credence in physical theories with more people. But this seems like straightforward updating on the evidence of my existence—worlds with more people are likelier to have me, just as, upon discovering a blueberry, one has some reason to think the world has more blueberries.

This might seem to be mistaken reasoning. Upon seeing a blueberry one gains evidence that the universe has blueberries in a large percentage of places, not that there are many blueberries overall (one gains no evidence, for instance, that the universe is big enough to have many blueberries from seeing a blueberry). But crucially this is because one only looks in a limited area, so the odds of seeing a blueberry are higher if the world is more densely populated with blueberries. In contrast, however, if I thought of a random possible blueberry, before having some method of determining if that blueberry is

actual, and discovered that that particular blueberry was actual, I'd get evidence that there are many blueberries. This is because in this case, wherever the blueberry is, I'd be in a position to observe it. But this is analogous to the case of my existence: wherever I exist, I'll be the first to know.

Perhaps one worries that this is the wrong way to think about one's existence. Your existence is, on such a view, not the kind of thing you should think of as a random event made more likely to occur in a universe with more people. Instead perhaps, as proponents of SSA argue, you should think of your existence as a random event selected from the pool of existences in your reference class. But this is simply to deny SIA. Of course if one thinks about the presumptuous philosopher result from the perspective of some view other than SIA, they won't find it appealing. But my claim is that there is a perfectly reasonable and consistent way to think about it, in accordance with SIA, where the result becomes not merely some unfortunate result to explain away, but one that actively makes sense. If the objection relies entirely on thinking about things in a way that a proponent of SIA would reject, then it is ineffective in the dialectical context.

Much of the unintuitiveness of the presumptuous philosopher case comes from the idea that the philosopher is justified, if SIA is true, in being extremely confident that one theory is correct. But this doesn't follow from SIA. If one is not 100% sure of SIA—which surely they shouldn't be, given the great amount of philosophical disagreement—then they should take empirical evidence seriously (this point is also made by Carlsmith, 2022, p.43).

More can be said in favor of the presumptuous philosopher. In fact, SSA, the main rival of SIA, is similarly presumptuous (Carlsmith, 2022, p.44-45). To see this, consider the following case (not from Carlsmith, but broadly similar to his examples):

> Presumptuous Archeologist: archeologists discover that there was a type of prehistoric humans that was very numerous—numbering in the quadrillions. Over time, they uncovered strong empirical evidence that these beings were exactly like modern humans—certainly similar enough to be part of the same reference class. The archeologists give a talk about their findings. At the end of the talk a philosopher gets up and declares "your data must be wrong, for if there were that

many prehistoric humans then it would be unlikely we'd be so late (for most people in your reference class have already existed). Thus, they can't be in our reference class and therefore the data must be wrong."

Clearly here the philosopher is being irrational. Yet this follows from SSA. SSA gives one reason to think that their reference class is small, and thus one must reject archeological evidence that conflicts with this belief. You might worry that this assumes some controversial notion about there being a symmetry, on SSA, between past and future observers. However, I think this problem will apply to every view that updates on one's existence and rejects SIA. If a universe with more prehistoric humans is no more likely to contain me, but a world with more prehistoric humans makes it less likely that I'd exist so late, then one has strong reason to reject the existence of many prehistoric humans. This can be seen once again in the same way: if I didn't know when I existed, and the existence of more people like me doesn't make my existence more likely, then I would initially assign equal probability to the hypothesis that there are many prehistoric humans like me and that there are few. Upon finding out that I'm not a prehistoric human, however, I'd get extremely strong evidence that there are few prehistoric humans. Thus, just as SIA proponents must disagree with the cosmologists, so too must those who reject SIA disagree with the archeologists!

In addition, attempts to avoid the presumptuous philosopher result have severe problems (for a helpful collection of them, see Olum, 2002).[5] Thus, even if the presumptuous philosopher result is as bad as the results surveyed in this paper, other considerations provide one strong reason to adopt SIA.

An additional argument can be given for the presumptuous philosopher result. The presumptuous philosopher is relevantly analogous to:

Presumptuous Hatcher: Suppose that there are googolplex eggs. There are two hypotheses: first that they all hatch and become people, second that only a few million of them are to hatch and

[5] Olum's argument has been criticized, however, by various people, see, for example Bostrom & Ćirković (2003).

become people. The presumptuous hatcher thinks that their existence confirms hypothesis one quite strongly.

This case is relevantly like the presumptuous philosopher case. Both the presumptuous philosopher and the presumptuous hatcher infer that there were many people that existed. The only difference is that in the Presumptuous Hatcher, both existent and non-existent people are paired with eggs, which only existent people hatch from.

Yet in Presumptuous Hatcher, the reasoning seems perfectly sound. Specifically, upon emerging from one's egg, one can safely reason in the following way: the probability that my egg would hatch is much higher on the first hypothesis than on the second. Given that it did hatch, I have very strong evidence that the first hypothesis is true.

We can give an additional argument for the correctness of the presumptuous hatcher's reasoning. Suppose that one discovered that when the eggs were created, the people who the eggs might have become had a few seconds to ponder the anthropic situation. Thus, each of those people thought "it's very unlikely that my egg will hatch if the second hypothesis is true but it's likely if the first hypothesis is true." If this were so, then the presumptuous hatcher's reasoning would clearly be correct, for then they straightforwardly update on their hatching and thus coming to exist! But surely finding out that prior to birth, one pondered anthropics for a few seconds, does not affect how one should reason about anthropics. Whether SIA is the right way to reason about one's existence surely doesn't depend on whether, at the dawn of creation, all possible people were able to spend a few seconds pondering their anthropic situation. Thus, we can argue as follows:

1) The presumptuous hatcher's reasoning, in the scenario where they pondered anthropics while in the egg, is correct.
2) If the presumptuous hatcher's reasoning, in the scenario where they pondered anthropics while in the egg, is correct, then the ordinary presumptuous hatcher's reasoning is correct.
3) If the ordinary presumptuous hatcher's reasoning is correct then the presumptuous philosopher's reasoning is correct.

4) Therefore, the presumptuous philosopher's reasoning is correct.

6 *Conclusion*

Here, I have argued that considerations related to the doomsday argument give us very strong reason to accept SIA. Contrary to the claims of Cirkovic (2001), the reason to support SIA isn't just that it nicely avoids the doomsday argument. Rather, it's that only it can avoid the doomsday argument. Given the many problems for alternatives to SIA, we have quite strong reason to accept SIA. Those who do not accept SIA must think that we're likely doomed and that one can be confident that they'll get hundreds of royal flushes in poker if they guarantee that were they not to get those royal flushes, they'd have enormous numbers of offspring.

In addition, I've argued that the main objection to the self-indication assumption, the presumptuous philosopher counterexample, does not succeed. The presumptuous philosopher's reasoning is not especially counterintuitive when one takes into account that their existence is likelier in a bigger universe. In addition, the presumptuous philosopher's reasoning is relevantly analogous to other forms of clearly correct reasoning, and views that avoid the presumptuous philosopher result imply their own, equally strange presumptuousness. Thus, while every view of anthropic reasoning will have some costs, the most common objection to the self-indication assumption is weak, and only the self-indication assumption can avoid utterly implausible results.[6]

[6] I'd like to thank Chris Minge and Amos Wollen for their helpful commentary on this argument.

Bibliography

Builes, D. D. A. (2020). *The empirical relevance of metaphysics* (Doctoral dissertation, Massachusetts Institute of Technology). https://dspace.mit.edu/bitstream/handle/1721.1/127150/1191839832-MIT.pdf?sequence=1

Bostrom, N. (2001). The Doomsday Argument Adam & Eve, UN++, and Quantum Joe. *Synthese*, 127, 359-387.

Bostrom, N. (2013). *Anthropic bias: Observation selection effects in science and philosophy*. Routledge.

Bostrom, N., & Ćirković, M. M. (2003). The doomsday argument and the self-indication assumption: reply to Olum. *The Philosophical Quarterly*, 53(210), 83-91.

Carlsmith, J. (2022). A stranger priority? Topics at the outer reaches of effective altruism [PhD thesis]. University of Oxford.

Carter, B. (1983). The anthropic principle and its implications for biological evolution. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *310*(1512), 347-363.

Ćirković, M. (2001). Is Many Likelier than Few? A Critical Assessment of the Self-Indicating Assumption. Unpublished. Retrieved from https://arxiv.org/ftp/gr-qc/papers/0110/0110002.pdf

Ćirković, M. (2004). Agencies, Capacities, and Anthropic Self-Selection, *Philosophical Writings*, 27, 43-62

Cushman, M. (2019). Anthropic Indexical Sampling and Implications for The Doomsday Argument. Unpublished. Retrieved from https://philsci-archive.pitt.edu/16088/1/anthropic.pdf.

Elga, A. (2020). Self-locating belief and the Sleeping Beauty problem. In *Arguing About Knowledge* (pp. 142-145). Routledge.

Grace, K. (2010). SIA doomsday: the filter is ahead. Meteuphoric, available at: *https://meteuphoric.com/2010/03/23/sia-doomsday-the-filter-is-ahead/*.

Haig, D. (2016). Sleeping Beauty in a grain of rice. *Biology & Philosophy*, *31*, 23-37.

Kopf, T., Krtous, P., & Page, D. N. (1994). Too soon for doom gloom?. arXiv preprint gr-qc/9407002.

Leslie, J. (2002). *The End of the World: the science and ethics of human extinction*. Routledge.

Lewis, D. K. (1986). *On the plurality of worlds* (Vol. 322). Oxford: Blackwell.

Olum, K. D. (2002). *The Philosophical Quarterly*, 52(207), 164-184.

[Adelstein, M], [Wollen, A]. (2024). A Presumptuous Proof of Theism. [Unpublished]

Tegmark, M. (2008). The mathematical universe. *Foundations of physics*, *38*(2), 101-150.