

# Average Utilitarianism Implies Solipsistic Egoism

Christian J. Tarsney

To cite this article: Christian J. Tarsney (2023) Average Utilitarianism Implies Solipsistic Egoism, Australasian Journal of Philosophy, 101:1, 140-151, DOI: [10.1080/00048402.2021.1962375](https://doi.org/10.1080/00048402.2021.1962375)

To link to this article: <https://doi.org/10.1080/00048402.2021.1962375>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 26 Sep 2021.



Submit your article to this journal [↗](#)



Article views: 2415



View related articles [↗](#)



View Crossmark data [↗](#)



# Average Utilitarianism Implies Solipsistic Egoism

Christian J. Tarsney

Global Priorities Institute, University of Oxford

## ABSTRACT

Average utilitarianism and several related axiologies, when paired with the standard expectational theory of decision-making under risk and with reasonable empirical credences, can find their practical prescriptions overwhelmingly determined by the minuscule probability that the agent assigns to solipsism—that is, to the hypothesis that there is only one welfare subject in the world, namely, herself. This either (i) constitutes a *reductio* of these axiologies, (ii) suggests that they require bespoke decision theories, or (iii) furnishes an unexpected argument for ethical egoism.

**ARTICLE HISTORY** Received 17 March 2020; Revised 2 July 2021

**KEYWORDS** average utilitarianism; solipsism; egoism; variable value; rank-discounted utilitarianism; population ethics

## 1. Average Utilitarianism

Average utilitarianism (AU) holds that the overall value of a world is equal to the average lifetime welfare of all welfare subjects in that world.<sup>1</sup> Among population axiologies, AU has some notable and distinctive virtues: For instance, it avoids Parfit's [1984] 'Repugnant Conclusion', and would be chosen by selfish agents from behind a particularly natural version of the veil of ignorance. It has plenty of vices to counter-balance these virtues (see Hurka [1982a], among many others), and has therefore never been an especially popular doctrine. But it has attracted its share of advocates over the years, including Hardin [1968], Harsanyi [1977], and Pressman [2015].<sup>2</sup>

In the world that we appear to inhabit, AU has the slightly dispiriting consequence that we can only make a very small difference to the overall value of the world: since the number of welfare subjects is very large, even acts that produce enormous welfare improvements in absolute terms have only minuscule effects on *average* welfare.

To illustrate, there are currently about  $7 \times 10^9$  human beings alive on Earth. There are also many billions of mammals, birds, and fish being raised by humans for meat and other agricultural products. And there are perhaps some  $10^{11}$  mammals living in the wild, along with similar or greater numbers of birds, reptiles, and amphibians,

<sup>1</sup> This, at least, is an especially natural and widely discussed version of average utilitarianism. There are others, which we will come to shortly.

<sup>2</sup> Mill [1863] can also be read as an average utilitarian (see Gustafsson [forthcoming: n2]), although the textual evidence for this reading is not entirely conclusive.

and a significantly larger number of fish—conservatively  $10^{13}$ , and possibly far more.<sup>3</sup> This is despite a significant decline in wild animal populations in recent centuries and millennia as a result of human encroachment.

To determine the *total* number of welfare subjects (by which we can divide total welfare to find an overall average), we must consider past as well as present individuals. (Future individuals count too, of course, but their numbers are much harder to estimate, and might depend on our choices.) Estimates of the number of human beings who have ever lived are on the order of  $10^{11}$  [Kaneda and Haub 2018]. But this number is dwarfed by past populations of non-human animals. In wild animal populations, most individuals die young (with smaller animals being both more numerous and shorter-lived), and so birth and death rates in the wild animal population as a whole are unlikely to be less than 1 per individual per year (corresponding roughly to an average lifespan of 1 year). Being extremely conservative, then, we might suppose that all and only mammals are welfare subjects and that  $10^{11}$  mammals have been alive on Earth at any given time since the K-Pg boundary event (~66 million years ago), with a population birth/death rate of 1 per individual per year. This implies a ‘timeless population’ of at least  $\sim 6.6 \times 10^{18}$  welfare subjects. Being a bit less conservative (although perhaps still objectionably conservative), we might suppose that all and only vertebrates are welfare subjects and that  $10^{13}$  vertebrates have been alive on Earth at any time in the last 500 million years (since shortly after the Cambrian explosion), with a population birth/death rate of 10 per individual per year. This implies a timeless population of at least  $\sim 5 \times 10^{22}$  welfare subjects.

Even by the more conservative estimate, we find that providing one unit of welfare to one individual increases average welfare by at most  $\frac{1}{6.6 \times 10^{18}} \approx 1.5 \times 10^{-19}$  units. By the less conservative estimate, increasing someone’s welfare by one unit will increase average welfare by at most  $\frac{1}{5 \times 10^{22}} = 2 \times 10^{-23}$  units.

## 2. Solipsism

There is, however, one hypothesis according to which the total number of welfare subjects in the world is quite a bit smaller—solipsism. Solipsism is the proposition that only I exist (or, in your case, the proposition that only you exist), and that what appears to be an external world populated with other individuals is in fact just a figment of my (or, respectively, your) imagination. If solipsism is true, then the total number of welfare subjects, the size of the universal population, is 1.

Solipsism is surely improbable. But just how improbable is it? Like average utilitarianism, it has several notable virtues: It is simple and ontologically parsimonious. It is a natural conclusion to draw from various arguments for external world scepticism (of the kind found, for example, in Descartes [1641] and Berkeley [1710]), if one does not simultaneously accept some ‘rescue’ hypothesis like theism. (And it is, at the very least, contentious whether those arguments for external world scepticism have ever been satisfactorily refuted.) Solipsism also provides a powerful answer to the otherwise intractable question ‘Why am I me?’—namely, ‘There isn’t

<sup>3</sup> For useful surveys of evidence on present animal population sizes, see Tomasik [2019] and Bar-On et al. [2018] (especially pp. 61–4 and Table S1 in the supplementary appendix). These numbers come with very large error bars, but are more likely to be underestimates than overestimates, and, in any case, the exact numbers will not matter much for the arguments that follow.

anyone else I could have been!’ And, finally, it is a recurring and enduring idea in the history of philosophical thought, having been entertained (in various forms) by such thinkers as the Buddhist philosopher Ratnakirti (see Kajiyama [1965]), Wittgenstein [1922], and Hare [2009].<sup>4</sup>

For our purposes, it will be necessary to go beyond these general observations and to say something about what *probability* one might reasonably assign to solipsism. A bit more specifically, the interest of the following arguments depends on the claim that one’s credence in solipsism should be *not absurdly small*—not less than, say,  $10^{-9}$  (one in a billion).

Of course, assigning probabilities to philosophical hypotheses is at best a matter of rough guesswork; we do not, for instance, have objective chances to guide us. But we can do better than simply pulling plausible-seeming numbers out of thin air. We can, for instance, consider a more complete and fine-grained space of possibilities over which our probabilities should sum to 1, and aim for reflective equilibrium among the credences that we assign to the various possibilities in that space. Fully carrying out that exercise here would be a tedious experience for the reader. But we can do a first approximation, aiming simply to find a plausible lower bound.

It seems clear that my credence that my common-sense view of the world has ‘got things basically right’, metaphysically speaking, should not be greater than 0.9. That is, given how little we have to go on, the lack of expert consensus in basic metaphysics, and trying to correct for the general human tendency toward overconfidence, I should have at least 0.1 credence that the world is in *some* way fundamentally very different than I take it to be. A good chunk of that credence should go to ‘some possibility that nobody has ever thought of’. But it also seems overconfident, conditional on my ordinary view of the world being wrong, to have credence greater than 0.9 in *that* possibility. This leaves at least 1% of my credence to distribute over known revisionary metaphysical hypotheses. And then, to get some sense of a lower bound on my credence in solipsism, I should ask, first, ‘Are there any other known revisionary hypotheses that are *many orders of magnitude* more probable than solipsism?’ (to which, it seems to me, the answer is ‘no’) and, second, ‘Are there thousands or millions of known revisionary hypotheses that are at least roughly as plausible as solipsism?’ (to which, again, the answer seems to be ‘no’). Taken together, these observations suggest that my credence in solipsism should be at most a few orders of magnitude less than 1% (that is,  $10^{-2}$ ).

All in all, then, while it would strike me as somewhat unreasonable to assign solipsism a probability greater than  $10^{-2}$ , it also seems unreasonable to assign it a probability less than  $10^{-9}$ . To assign any more extreme probability would not display due modesty about our understanding of matters metaphysical.<sup>5</sup>

---

<sup>4</sup> It is hard to find straightforward endorsements of the strongest form of metaphysical solipsism, but this provides little if any evidence against the view. First, true solipsists have little reason to publicize their beliefs. Second, and more to the point, if solipsism is true and the whole external world is a figment of my imagination, there is no obvious reason to expect that I would imagine the particular sub-figments that I mistake for other people to go around proclaiming that the whole world is a figment of *their* imaginations.

<sup>5</sup> In drafting this paper, I became curious what credence people actually *do* assign to solipsism, when prompted. So, in an *extremely* unscientific and non-incentive-compatible survey on social media ( $n = 32$ ), I asked respondents to assign credences to solipsism, defined as ‘the hypothesis that only I exist (or, in your case, the hypothesis that only you exist), and that what appears to be a physical world containing objects, other people, etc. is in fact just a product of/contained within my (or, in your case, your) mind’.

**Table 1.** Solipsistic swamping for average utilitarianism

	Solipsism is false $1 - 10^{-9}$	Solipsism is true $10^{-9}$	EV
Altruistic option	$\frac{10^6}{10^{18}} = 10^{-12}$	0	$\sim 10^{-12}$
Selfish option	$\frac{1}{10^{18}} = 10^{-18}$	1	$\sim 10^{-9}$

3. Solipsistic Egoism

Now, consider an average utilitarian, Ava, who assigns solipsism a subjective probability of  $10^{-9}$ , and must choose between taking one unit of welfare for herself, or providing a thousand other welfare subjects with a thousand welfare units each. And let's suppose she believes that, if solipsism is false and the external world/other minds are real (hereafter, 'realism'), then the total number of welfare subjects in the world is  $10^{18}$ . (For simplicity, I am rounding down our already conservative lower-bound estimate of  $6.6 \times 10^{18}$ , and ignoring the credence that Ava ought to have in larger population sizes, which would only strengthen our conclusions.) And let's assume (without any loss of generality) that, regardless of whether or not solipsism is true, average welfare prior to Ava's intervention is 0. This situation is summarized in Table 1.

Now, suppose that Ava responds to risk in the standard way, by maximizing expected value. Given the facts stipulated above, the expected value of the altruistic option is  $(1 - 10^{-9}) \times 10^{-12} + 0 \times 10^{-9} \approx 10^{-12}$ , while the expected value of the selfish option is  $(1 - 10^{-9}) \times 10^{-18} + 1 \times 10^{-9} \approx 10^{-9}$ . That is, even though the altruistic option *almost certainly* yields a million times more value than the selfish option, the selfish option has a thousand times greater *expected* value, because if solipsism is true, and only Ava exists, then she can have astronomically greater impact on average welfare than she could otherwise hope for. Despite the enormous disparity in stakes, we find that Ava ought to choose the selfish option as long as her credence in solipsism is greater than  $\sim 10^{-12}$ . Conversely, holding fixed her credence in solipsism at  $10^{-9}$ , we find that she should give her own interests a billion times more practical weight than anyone else's: that is, her interests carry a billion times greater weight in expectation.<sup>6,7</sup>

Among those who gave sharp credences, responses spanned the range [0,0.5], with a median of 0.01 and a mean of  $\sim 0.105$ . Arbitrarily excluding the answers that I take to be clearly irrational (those outside the interval (0,0.1]) still gives a median of 0.01 but reduces the mean to  $\sim 0.048$ . On the other hand, including a few participants who gave interval credences, at the lower bound of their intervals, gives a median of  $10^{-6}$  and a mean of  $\sim 0.091$ . Of course, all of these numbers are quite a bit greater than  $10^{-9}$ .

<sup>6</sup> As I discovered while revising this paper, the vulnerability of AU to this sort of 'solipsistic swamping' has been noticed once before, in a short blog post by Caspar Oosterheld [2017]. I am not aware of any other discussions of the phenomenon, or any discussion of its generalization to the other axiologies besides AU discussed below.

<sup>7</sup> We will consider shortly whether *maximize expected value* is the right decision rule for Ava to follow here. But I will take it for granted that ethical theories like AU must, in one way or another, include or be combined with *some* rule for decision-making under risk. This can be done in various ways. For instance, an average utilitarian might adopt a subjective criterion of rightness according to which an option is right iff it maximizes average welfare in expectation, or she might recognize both objective and subjective rightness. But even an average utilitarian who is a univocal objectivist about rightness, holding that an option is right iff it maximizes average welfare *in fact*, can and should provide a substantive theory of decision-making under risk. She might, for instance, hold that an option is *right* iff it maximizes average welfare in fact, but *rational* iff it maximizes average welfare in expectation. Or she might hold that her objective criterion of rightness applies not only to choices among acts but also to choices among decision procedures, and that, at least

#### 4. Generalizations

We assumed that Ava accepts a particular (very natural) version of average utilitarianism, which has been our exclusive focus so far. But, as Thomas Hurka has emphasized, there are many non-equivalent views that can be described as ‘average utilitarian’. He describes [1982a, 1982b] a total of eleven such views, which he names A1–A11. (Ava’s view, which I have called AU and which tells us to maximize average lifetime welfare in the timeless population, is Hurka’s A1.) These theories are not all equally vulnerable to solipsistic swamping. For A2, which tells us to maximize the sum of momentary welfare averages (that is, averaging welfare at each time and then summing across times), the crucial number that determines how much solipsism magnifies one’s efficacy is the size of the *present* population, rather than the *timeless* population. For A7, which tells us to maximize the average lifetime welfare of *present and future* people (ignoring the past), the crucial number is of course the size of the present and future population. So, either of these views somewhat dampens the swamping phenomenon. On several other views (Hurka’s A3, A4, A6, A8, A9, and A11), which involve averaging across times, things depend on how long the agent believes that she will exist if solipsism is true, and how long the universe as a whole will contain welfare subjects if solipsism is false. A5 and A10 evade the solipsistic swamping problem entirely; but, as Hurka points out, these views are independently very implausible.

Solipsistic swamping also threatens other axiologies that try to capture the intuitive attractions of AU in large-population contexts (like avoiding the Repugnant Conclusion). For instance, consider the view that Hurka [1983] calls ‘Variable Value I’ (VV1), according to which the value of a population  $X$  is given by

$$V(X) = \bar{X}f(|X|)$$

where  $\bar{X}$  is the average welfare level in  $X$ ,  $|X|$  is the number of welfare subjects in  $X$ , and  $f$  is a function that is strictly increasing, strictly concave, and has a horizontal asymptote.<sup>8</sup> ‘Variable value’ axiologies are meant to resemble total utilitarianism for small populations and average utilitarianism for large populations, reflecting the intuition that increasing the size of a (happy) population, without changing its average welfare, adds value when the population is small but has diminishing marginal value as population size increases.

How vulnerable is VV1 to solipsistic swamping? Very roughly, the crucial factor is the ratio  $r$  between  $f(1)$  and the horizontal asymptote of  $f$ . If this ratio is much larger than the minimum population size conditional on realism, then VV1 might agree arbitrarily closely with total utilitarianism, and so be safe from solipsistic swamping. If  $r$  is much smaller than that minimum population size, then VV1 will reduce the extreme practical weight that AU gives to solipsism by approximately a factor of  $r$ . To illustrate

---

for an agent who faces many future choices, what will maximize average welfare in fact is to adopt the decision procedure of maximizing average welfare in expectation.

<sup>8</sup> This view is also discussed by Ng [1989], under the name “Theory X”.

the latter case, suppose that

$$f(|X|) = 1 - \frac{1}{1 + 10^{-9}|X|}$$

Here  $r \approx 10^9$ , meaning that the value of a large population with a given average welfare level converges to roughly one billion times the value of a singleton population at that same welfare level.

Now the problem of solipsistic swamping persists, but is much less extreme. Since  $\frac{f(10^{18})}{f(1)} \approx 10^9$ , the relative weight of the solipsistic hypothesis is reduced by a factor of nearly  $10^9$ , and so, using the numbers from our original example (in Table 1), we now find that the altruistic option has greater expected value than the selfish option. But the problem is far from vanquished. Consider a new agent, Valerie, who (i) accepts VV1 with the  $f$  specified above, (ii) assigns solipsism a credence of  $10^{-9}$ , (iii) accepts our slightly-less-conservative estimate of the minimum population size conditional on realism, which, for simplicity, we will round down to  $10^{22}$  (as compared to  $10^{18}$  in the case of Ava), and (iv) must choose between taking one welfare unit for herself

**Table 2.** Solipsistic swamping for Variable Value I

	Solipsism is false $1 - 10^{-9}$	Solipsism is true $10^{-9}$	EV
Altruistic option	$\frac{10^3}{10^{22}} \times f(10^{22}) \approx 10^{-19}$	0	$\sim 10^{-19}$
Selfish option	$\frac{1}{10^{22}} \times f(10^{22}) \approx 10^{-22}$	$1 \times f(1) \approx 10^{-9}$	$\sim 10^{-18}$

or providing a thousand other welfare subjects with one welfare unit each (for a thousand units in total, as compared to a million in the case of Ava). For Valerie, like Ava, selfishness is the order of the day—even though she is nearly certain that the altruistic option will produce far more value (Table 2).

A bit more generally, given a VV1 axiology with the  $f$  specified above,  $10^{-9}$  credence in solipsism, and a minimum population of  $10^{22}$  conditional on realism, Valerie should give her own interests at least 10,000 times as much weight as anyone else's, because of her credence in solipsism.<sup>9</sup>

On the other hand, the *rank-discounted utilitarian* (RDU) axiology defended by Asheim and Zuber [2014] faces a more extreme form of solipsistic swamping than even AU does. On this view, the value of a population  $X$  is given by

$$V(X) = \sum_{r=1}^{|X|} \beta^r w_x(x_r)$$

where the members of  $X$  are indexed in order of increasing welfare by their rank  $r$ ,  $w_x(x_r)$  gives the welfare of the  $r$ th-worst-off individual in  $X$ , and  $\beta \in (0, 1)$  is a constant

<sup>9</sup> Hurka also describes a view that he calls Variable Value II, which applies an increasing and concave transformation  $g$  to average welfare, so that the overall value of a population  $X$  is given by  $g(\bar{X})f(|X|)$ . This view behaves like VV1 for our purposes, with the additional caveat that, if  $g$  is sufficiently concave or the amount by which an agent can improve her own welfare is sufficiently great, then  $g$  will further moderate the difference in stakes between solipsism and realism.

**Table 3.** Solipsistic swamping for RDU ( $\beta = 0.99999$ )

	Solipsism is false. $1 - 10^{-9}$	Solipsism is true. $10^{-9}$	EV
Altruistic option	$< 10^6 \times \beta^{10^9}$	0	$< \sim 10^{-4337}$
Selfish option	$< 1 \times \beta^{10^9}$	$\beta$	$\sim 10^{-9}$

that determines the degree to which worse-off individuals are prioritized over better-off individuals.

This view does not uniformly discount the interests of each individual in large population scenarios, as average utilitarianism does: the worst-off individual, for instance, always gets exactly the same weight, regardless of population size. But because the weight given to the interests of better-off individuals diminishes geometrically with their welfare rank, the interests of *all but the very worst off* can be dramatically discounted in large-population scenarios. For instance, suppose that Ragnar (i) accepts RDU with  $\beta = 0.99999$  ( $\beta$  closer to 1 implies less discounting of the better off), (ii) assigns solipsism a credence of  $10^{-9}$ , (iii) believes that there are at least  $10^{18}$  welfare subjects, conditional on realism, and (iv) must choose between taking one welfare unit for himself or providing a million other welfare subjects with one welfare unit each. Further, suppose Ragnar knows that none of the million individuals he has the chance to benefit are among the  $10^9$  worst-off. (If the total number of welfare subjects is at least  $10^{18}$ , then it is *extremely* unlikely, in any given choice situation, that one is in a position to help any of the  $10^9$  worst-off.) On the other hand, Ragnar recognizes that if solipsism is true, then he is very well positioned to improve the welfare of the very worst-off individual in the whole universe—namely, himself.

Because of the power of geometric discounting, Ragnar will find that there is a truly dramatic disparity between the solipsistic and non-solipsistic stakes (Table 3): if solipsism is true, then he can do at least  $10^{4337}$  times more good (by acting selfishly) than he could do if solipsism is false (by acting altruistically). RDU, then, produces a much stronger form of solipsistic swamping than even AU.<sup>10</sup>

**5. Escape Routes**

So, not just AU but several other *prima facie* plausible population axiologies as well are vulnerable (in different degrees) to solipsistic swamping. This suggests that we must either reject all of these axiologies or embrace *de facto* ethical egoism. But there are various ways in which we might try to avoid that forced choice. I will briefly consider five such escape routes.

First, perhaps we were too generous to solipsism in supposing that it deserves a credence of at least  $10^{-9}$ . As we have seen, however, even *much* smaller credences (say,  $10^{-1000}$ ) are enough to make trouble for RDU. And if the minimum population size conditional on realism is significantly larger than we have so far supposed (for

<sup>10</sup> Of course, the proponent of RDU could always hand-select a  $\beta$  close enough to 1 to avoid solipsistic swamping. But, apart from the *ad hoc*-ery of this tactic, it seems very likely that such a large  $\beta$  would make RDU practically indistinguishable from total utilitarianism. At the very least, RDU must thread a very tight needle to avoid collapsing into egoism, on the one hand, or totalism, on the other. (And, no matter what  $\beta$  we choose, as long as we hold it fixed in the face of new empirical information, we will be in constant danger of collapsing into solipsistic egoism if we come to believe that the total world population is significantly larger than we thought, or into *de facto* total utilitarianism if we come to believe that it is smaller.)



example, if the universe contains many other biospheres with welfare subjects, and perhaps many large interstellar civilizations), then AU and VV1 are also vulnerable to swamping from much smaller credences in solipsism. So, reducing our credence in solipsism (as long as it remains non-zero) is at best an insecure escape route.

The potential for solipsistic swamping to survive even many-order-of-magnitude reductions in the probability of solipsism highlights its resemblance to other problem cases in which expected value calculations are dominated by tiny probabilities of extreme scenarios (for example, Pascal's wager [Pascal 1669], Pascal's mugging [Bostrom 2009], or the St. Petersburg game [Bernoulli 1738]). Merely noting this similarity serves at best to categorize the solipsistic swamping problem, not to solve it. But it also suggests that we might avail ourselves of existing strategies for responding to problems of this general type. For instance, a second possible response is to adopt the perennial suggestion of simply ignoring very small probabilities [Buffon 1777; Smith 2014; Monton 2019]. But, without entering into the large existing debate over this proposal, suffice it to say that it carries quite serious drawbacks and is generally considered unsatisfactory [Hájek, 2014; Isaacs, 2016; Lundgren and Stefánsson *forthcoming*].

The comparison with Pascal's wager also suggests a third response, analogous to the famous 'many gods' objection: perhaps there are other very-small-population scenarios besides solipsism that are more probable and/or higher-stakes, and that favour some practical conclusion other than egoism. For instance (as one reviewer suggested), perhaps the external world, rather than being a figment of *my* imagination (à la solipsism), is an illusion created for me by some more powerful being G (for example, God), who will reward or punish me depending on my choices—perhaps rewarding me if I act altruistically and punishing me if I act selfishly. If we further suppose that G and I are the only welfare subjects in the timeless population, then AU et al will similarly assign outsized importance even to very small credences in this hypothesis.

I am not convinced that this particular hypothesis will reverse or otherwise override the egoistic implications of the solipsistic hypothesis. Consider how it affects the implications of expectational AU, for instance. On the one hand, suppose that G is perfectly morally good. In that case, it seems that I should expect G to 'reward' me by maximizing my welfare whatever I do (since, all else being equal, that will maximize average welfare)—and, at any rate, not to punish me for doing whatever would maximize expected average welfare before accounting for G's rewards and punishments. If G is not all-good, on the other hand, then, as far as I can see, all bets are off. There is no particular reason to think that G is more likely to reward me for one course of action rather than for another (for example, for acting altruistically rather than selfishly, or vice versa). In either case, the hypothesis that only G and I exist seems practically inert—or, rather, reinforces the case for solipsistic egoism by providing another high-stakes (because small-population) scenario in which I have no clear reason to do anything other than to maximize my own welfare.

With that said, I do want to concede the general possibility of other low-probability scenarios involving very small populations whose practical implications contradict and supersede those of solipsism. But it would be remarkable, at least, if the practical implications of these ultra-swamping scenarios turned out to match closely the implications of our ordinary common-sense beliefs about the world. So, while the possibility of 'many gods'-style objections should make us less confident that averagist, variable value, and rank-discounted axiologies support *egoism* in the last analysis, it only reinforces the general point that the practical implications of these views can be

hijacked, and taken in unexpected and counterintuitive directions, by very small credences in very-small-population scenarios.

The ‘G’ hypothesis, however, also serves to motivate a fourth response to solipsistic swamping: perhaps the problem comes from responding to *metaphysical* uncertainty by using a decision rule (expected value maximization) that is appropriate only for responding to *empirical* uncertainty. Metaphysical uncertainty seems to provide especially fertile ground for extreme low-probability hypotheses that, when handled with an expectational decision rule, generate fanatical ‘swamping’ effects. The ‘G’ hypothesis arguably illustrates this point. For another example (suggested by a different reviewer), consider a metaphysical hypothesis that endorses unrestricted mereological composition and allows that welfare subjects can overlap with and include one another, such that what we take to be single welfare subjects are in fact vast colonies of overlapping welfare subjects (for example, you-minus-one-neuron, you-minus-a-different-neuron, ...). On this hypothesis, plausibly, the vast majority of all welfare subjects will be proper parts of the single largest welfare subject, because its greater number of parts permits a vastly greater number of mereological combinations. In this case, it is *total* utilitarianism and kindred theories that are in danger of swamping: even a tiny credence in this hypothesis might imply that it is overwhelmingly important, in expectation, to benefit the largest welfare subject, and thereby its proper parts.

Again, I don’t want to deny that there might be other metaphysical hypotheses besides solipsism that can generate fanaticism problems for particular normative theories. But I doubt that drawing a decision-theoretic distinction between metaphysical and empirical uncertainty is the right way to deal with these problems. First, the line between empirical and metaphysical uncertainty is far from clear. For instance, although solipsism may be a topic for metaphysicians, I find it hard to say precisely what makes it a *metaphysical hypothesis* as opposed to a (deeply revisionary) *empirical* hypothesis about what *concretia* exist and how they interact with and explain one another. Second, if we don’t respond to metaphysical uncertainty like empirical uncertainty, how should we respond to it? There is a clear analogy here with the recent literature on decision-making under *normative* uncertainty, where there has been much debate over whether the same decision rules should apply to normative and empirical uncertainty. The two most widely discussed views that distinguish normative from empirical uncertainty are (i) ‘My Favourite Theory’, according to which an agent should simply act on the single normative theory in which she has greatest credence [Gracely 1996; Gustafsson and Torpman 2014] and (ii) normative externalism, according to which an agent should simply act on the *true* normative theory, regardless of whether she believes that theory or her evidence supports it [Weatherson 2014; Harman 2015; Hedden 2016]. By analogy, we might avoid problems of metaphysical fanaticism by adopting a ‘My Favourite Metaphysics’ view, according to which an agent should simply act on the single metaphysical theory in which she has greatest credence, or ‘metaphysical externalism’, according to which an agent should simply base all her choices on the metaphysical *truths*, regardless of her beliefs or evidence. To my knowledge, nobody has defended either of these views, and they seem fairly *ad hoc* and undermotivated. In particular, the standard motivations for the corresponding views about normative uncertainty (for example, that hedging between rival normative theories requires problematic intertheoretic comparisons [Hedden 2016] or involves ‘moral fetishism’ [Weatherson 2014]) do not transfer to their

metaphysical counterparts.<sup>11</sup> In any event, if the right conclusion to take from the solipsistic swamping problem is that we must draw a basic decision-theoretic distinction between empirical and metaphysical uncertainty, this itself would be an important and unexpected conclusion.

Perhaps most importantly, however, solipsistic swamping is just the limiting case of a more general problem for averagist, variable value, and rank-discounted axiologies, which has no intrinsic connection either with metaphysical uncertainty or with very small probabilities—namely, that, when combined with standard expectational decision rules, these views all seem to over-weight small-population scenarios. For instance, consider an average utilitarian who assigns 1% credence to the hypothesis that the universe will only ever contain  $10^{20}$  welfare subjects, and 99% credence to the more optimistic hypothesis that advanced future civilizations will eventually support  $10^{50}$  welfare subjects or more [Bostrom 2013]. The same absolute welfare improvement matters  $10^{30}$  times more in the former scenario and therefore, discounting for her credence, matters  $10^{28}$  times more in expectation. Thus, even though she is quite confident in the ‘optimistic’ hypothesis, she should base her choices almost entirely on the ‘pessimistic’ hypothesis.<sup>12</sup> More generally, expectational average utilitarians will generally give almost no practical weight to states that imply very large timeless populations, even when those states are very probable. Apart from optimism about the future of humanity, such states might correspond to (i) hypotheses that attribute sentience to more beings—for example, to insects, other invertebrates, or relatively simple artificial intelligences—or (ii) cosmological hypotheses that imply that the universe is very large and hence contains many non-Earth-originating welfare subjects (as well as exobiological hypotheses that imply a higher probability of welfare subjects emerging in a given star system). If we find this general phenomenon of ‘small-population swamping’ counterintuitive, then ignoring either metaphysical uncertainty or small probabilities won’t help, since we cannot assume that small-population scenarios will always be metaphysical in character or deserve *de minimis* probability.

A fifth and final strategy, then, is to attempt a more thoroughgoing abandonment of standard expectational decision theory: perhaps averagist, variable value, and rank-discounted axiologies must be equipped with their own bespoke theories of decision-making under risk that avoid the tyranny of small-population scenarios. This strikes me as the most plausible way for average utilitarians et al. to avoid solipsistic swamping, and the only possible way to avoid the more general phenomenon of small-population swamping. But it is not immediately obvious what these revised decision rules should look like, and, in departing from standard decision theory, they are likely to incur significant theoretical costs.<sup>13</sup> And, in any case, if we conclude

<sup>11</sup> On the other hand, several of the *objections* to My Favourite Theory and normative externalism *do* seem to transfer—e.g. the problem of theory individuation for My Favourite Theory [MacAskill and Ord 2020: 334–5] and the ‘dependence problem’ for externalism [Podgorski 2020].

<sup>12</sup> Of course, this is complicated by the facts that (i) if the optimistic hypothesis is true then agents like us may be able to have much greater impact on *total* welfare, and so perhaps a similar level of impact on average welfare, and (ii) we might be in a position to significantly influence the population size of future civilization.

<sup>13</sup> Here is one example. Teruji Thomas suggests an extension of average utilitarianism that ranks risky prospects by *expected total welfare divided by expected population size* [2016: 150]. This view straightforwardly avoids solipsistic swamping. But it has the very significant downside of violating statewise dominance—that is, preferring options that yield worse outcomes in every possible state of the world. (As proof, consider a choice between a lottery  $L_1$  that yields one individual with welfare 10 in state  $S_1$  or 9 individuals with

that particular views in population ethics cannot safely appeal to the best-developed and most widely accepted theory of decision-making under risk, this itself would also be a notable conclusion.

Unless we can find some clever decision-theoretic escape, then, we are left with a conditional: if certain otherwise plausible axiologies are correct, then the best thing that we can do, *ex ante*, to make the world a better place is to act selfishly (to greater or lesser extents, depending on the axiology). This leaves us, of course, with two further options: reject all of these axiologies, or embrace (*de facto*, impartially motivated) ethical egoism.<sup>14</sup>

## Disclosure Statement

No potential conflict of interest was reported by the author.

## ORCID

Christian J. Tarsney  <http://orcid.org/0000-0002-6324-7145>

## References

- Asheim, G.B. and S. Zuber 2014. Escaping the Repugnant Conclusion: Rank Discounted Utilitarianism with Variable Population, *Theoretical Economics* 9/3: 629–50.
- Bar-On, Y.M., R. Phillips, and R. Milo 2018. The Biomass Distribution on Earth, *Proceedings of the National Academy of Sciences* 115/25: 6506–11.
- Berkeley, G. 1710 (1982). *A Treatise Concerning the Principles of Human Knowledge*, Indianapolis, IN: Hackett Publishing.
- Bernoulli, D. 1738 (1954). Exposition of a New Theory on the Measurement of Risk, *Econometrica: Journal of the Econometric Society* 22/1: 23–36.
- Bostrom, N. 2009. Pascal's Mugging, *Analysis* 69/3: 443–5.
- Bostrom, N. 2013. Existential Risk Prevention as Global Priority, *Global Policy* 4/1: 15–31.
- Buffon, G.-L. L. d. 1777 (2010). *Essai d'Arithmetique Morale* (Essays on Moral Arithmetic), in *The Selten School of Behavioral Economics: A Collection of Essays in Honor of Reinhard Selten*, trans. John D. Hey, Tibor M. Neugebauer, and Carmen M. Pasca, ed. A. Ockenfels and A. Sadrieh, Heidelberg: Springer: 245–82.
- Descartes, R. 1641 (2016). *Meditations on First Philosophy: With Selections from the Objections and Replies*, ed. John Cottingham, Cambridge: Cambridge University Press.

---

welfare 20 in state  $S_2$ , and a lottery  $L_2$  that yields 9 individuals with welfare 11 in  $S_1$  or 1 individual with welfare 21 in  $S_2$ , where  $S_1$  and  $S_2$  are equiprobable.  $L_2$  statewise dominates  $L_1$ , by average utilitarian lights, but the 'expected total welfare divided by expected population size' method of evaluating lotteries gives  $L_1$  a value of 19 and  $L_2$  a value of 12.) A simpler strategy is just to change the shape of the utility function—for example, to hold that average utilitarians should maximize the expectation, not of average welfare, but of some non-linear transformation of average welfare. But this does not seem to help very much: in any region where the transformation is concave, solipsistic swamping will be less extreme when the possible outcomes in the solipsistic state are better than the possible outcomes in the realist state, but correspondingly more extreme when the possible outcomes in the solipsistic state are worse than the possible outcomes in the realist state. Where the transformation is convex, the pattern is reversed. If the possible outcomes, given realism, are located very close to an inflection point where the transformation goes from convex to concave, then solipsistic swamping might be mitigated more generally. But finding ourselves at such an inflection point can only be a matter of luck, and any help that it gives us with the solipsistic swamping problem falls apart if we come to believe that the possible outcomes (average welfare levels), given realism, are significantly better or worse than we previously thought.

<sup>14</sup> For helpful feedback on earlier versions of this paper, I am grateful to Teruji Thomas, Michael Pressman, Dean Spears, audiences at MIT and the 2021 APA Eastern Division Meeting, and two anonymous reviewers for this journal.

- Gracely, E.J. 1996. On the Noncomparability of Judgments Made by Different Ethical Theories, *Metaphilosophy* 27/3: 327–32.
- Gustafsson, J.E. [forthcoming](#). Our Intuitive Grasp of the Repugnant Conclusion, in *The Oxford Handbook of Population Ethics*, ed. G. Arrhenius, K. Bykvist, T. Campbell, and E. Finneron-Burns, Oxford: Oxford University Press.
- Gustafsson, J.E. and O. Torpman 2014. In Defence of My Favourite Theory, *Pacific Philosophical Quarterly* 95/2: 159–74.
- Hájek, A. 2014. Unexpected Expectations, *Mind* 123/490: 533–67.
- Hardin, G. 1968. The Tragedy of the Commons, *Science* 162/3859: 1243–8.
- Hare, C.J. 2009. *On Myself, and Other, Less Important Subjects*, Princeton: Princeton University Press.
- Harman, E. 2015. The Irrelevance of Moral Uncertainty, in *Oxford Studies in Metaethics* 10, ed. R. Shafer-Landau, Oxford: Oxford University Press: 53–79.
- Harsanyi, J.C. 1977. Morality and the Theory of Rational Behavior, *Social Research* 44/4: 623–56.
- Hedden, B. 2016. Does MITE Make Right? On Decision-Making under Normative Uncertainty, in *Oxford Studies in Metaethics* 11, ed. R. Shafer-Landau, Oxford: Oxford University Press: 102–28.
- Hurka, T.M. 1982a. Average Utilitarianisms, *Analysis* 42/2: 65–9.
- Hurka, T.M. 1982b. More Average Utilitarianisms, *Analysis* 42/3: 115–19.
- Hurka, T. 1983. Value and Population Size, *Ethics* 93/3: 496–507.
- Isaacs, Y. 2016. Probabilities Cannot Be Rationally Neglected, *Mind* 125/499: 759–62.
- Kajiyama, Y. 1965. Buddhist Solipsism – A Free Translation of Ratnakirti's Samtanantaradusana, *Journal of Indian and Buddhist Studies (Indogaku Bukkyogaku Kenkyu)* 13/1: 435–20.
- Kaneda, T. and C. Haub 2018. How Many People Have Ever Lived on Earth? *Population Reference Bureau*, URL = <https://www.prb.org/howmanypeoplehaveeverlivedonearth>
- Lundgren, B. and H.O. Stefánsson [forthcoming](#). Against the de Minimis Principle, *Risk Analysis*.
- MacAskill, W. and T. Ord 2020. Why Maximize Expected Choice-Worthiness? *Noûs* 54/2: 327–53.
- Mill, J.S. 1863. *Utilitarianism*, London: Parker, Son, and Bourne.
- Monton, B. 2019. How to Avoid Maximizing Expected Utility, *Philosophers' Imprint* 19/18: 1–25.
- Ng, Y. 1989. What Should We Do About Future Generations? Impossibility of Parfit's Theory X, *Economics and Philosophy* 5/2: 235–53.
- Oesterheld, C. 2017. The Average Utilitarian's Solipsism Wager, at *The Universe from an Intentional Stance* (blog): URL = <https://casparoesterheld.com/2017/03/15/the-average-utilitarians-solipsism-wager>
- Parfit, D. 1984. *Reasons and Persons*, Oxford: Clarendon Press.
- Pascal, B. 1669 (1852). *Pensées*, Paris: Dezobry et E. Magdeleine.
- Podgorski, A. 2020. Normative Uncertainty and the Dependence Problem, *Mind* 129/513: 43–70.
- Pressman, M. 2015. A Defence of Average Utilitarianism, *Utilitas* 27/4: 389–424.
- Smith, N.J.J. 2014. Is Evaluative Compositionality a Requirement of Rationality? *Mind* 123/490: 457–502.
- Thomas, T. 2016. *Topics in Population Ethics*, D.Phil. dissertation, St. Cross College, University of Oxford.
- Tomasik, B. 2019. How Many Wild Animals Are There?, URL = <https://reducing-suffering.org/how-many-wild-animals-are-there>
- Weatherston, B. 2014. Running Risks Morally, *Philosophical Studies* 167/1: 141–63.
- Wittgenstein, L. 1922 (1999). *Tractatus Logico-Philosophicus*, trans. C.K. Ogden. Mineola, NY: Dover.