Saving Fanaticism

Kacper Kowalczyk*

Abstract

Fanaticism is the view that, for every finite good x and every positive probability p, there is a finite good y such that getting y with probability p is better than getting x for sure. I develop a neglected argument for a form of fanaticism limited to rescue scenarios. I explain how my argument is compatible with some forms of small-probability discounting, imprecise probabilism, risk-aversion, and aggregation scepticism. I also try to respond to theoretical problems that fanatical arguments encounter in cases that involve infinitely many possible people. I also suggest why, even if fanaticism is true, we might often be warranted in our intuitive reluctance to accept fanatical conclusions.

It might feel uneasy to sacrifice a great good for the sake of a very small probability of a much greater good. It might even feel fanatical. Take, for example, Bostrom's (2013: 18–19) suggestion that, since humanity could thrive on Earth for another billion years, it might be one hundred times as important to reduce the risk of extinction by one millionth of one percent than to directly save one million human lives.[1] Indeed, 'fanaticism' has recently come to denote the view that, for every finite good x and every positive probability p, there is a finite good y such that getting y with probability p is better than getting x for sure (Wilkinson 2022; Russell 2023).[2]

There are nonetheless compelling arguments for fanaticism. One argument appeals to a spectrum of prospects, seemingly increasing in value, beginning with a prospect which certainly results in a great good and ending with a prospect which results in a much greater good with a very small probability (Beckstead and Thomas 2023). Another argument appeals to principles which seek to ensure that our evaluation does not depend in strange ways on distant space and time (Wilkinson 2022; Beckstead and Thomas 2023).

*Forthcoming in *The Australasian Journal of Philosophy*. Please cite the published version.

[1]Bostrom himself does not necessarily endorse this conclusion; see Bostrom 2009; 2011.

[2]This should be distinguished from Bostrom's original usage, according to which 'fanaticism' denotes the view that, for every finite good x, every infinite good y, and every positive probability p, getting y with probability p is better than getting x for sure (Bostrom 2011, 36–39).

This paper develops an even more compelling argument for a form of fanaticism limited to rescue scenarios, according to which, for every number x and every positive probability p, there is a number y such that it is better to save y people with probability p than to save x people for sure. If this conclusion is correct, it seems that there can be no issue with fanaticism as such but merely with the particular goods to which it might be applied. I explain why my argument should appeal to proponents of many views that otherwise appear hostile to fanatical conclusions. I also suggest why we might often be warranted in our intuitive reluctance to accept such conclusions. Lastly, I try to respond to theoretical problems that fanatical arguments encounter in cases that involve infinitely many possible people (compare Russell 2023).

1 Risky Rescues

Consider

> *Risky Rescue*. A volcano threatens to destroy two islands: a small one and a large one. The small island is less populous than the large one. You happen to be the captain of a rescue boat, the only one in the vicinity. You can head for one island, or the other, but not both. The ocean around the large island is rarely calm enough to be navigable, so it is possible but unlikely that you will reach the large island if you try. But you can certainly reach the small island.

In this case, the fanatical conclusion would be that, regardless of the population of the smaller island or the probability of reaching the larger one, heading for the large island is better, provided that its population is sufficiently vast. Below, I present a wholly general argument for this conclusion; indeed, the argument supports a more specific conclusion that heading for the large island is better as long as more people will then be saved in expectation. But, for the sake of illustration, it is useful to employ some very small numbers, unlikely to prompt accusations of fanaticism. In particular, let's suppose that there is one person on the small island, four people on the large island, and the probability of reaching the large island is one-in-three. This specific case is illustrated in Table 1.

	Calm (1/3)	Stormy (2/3)	Calm (1/3)	Stormy (2/3)
A	1	1	0	0
B	0	0	1	0
C	0	0	1	0
D	0	0	1	0
E	0	0	1	0
	Rescue Small Island		Rescue Large Island	

Table 1: Risky Rescue

In this table, columns correspond to the possible states of the ocean, which have the probabilities shown, while rows specify the fate of each of the five people affected, called 'A' to 'E', with a shaded '1' denoting survival and a '0' denoting death. The argument that it is better to head for the large island in this case begins by considering its 'opaque' version, which differs only in that, conditional on either state of the ocean, every possible distribution of the affected people across the two islands is equally probable as every other. The details of this version can be filled out in multiple different ways. For example, you might be uncertain because you misplaced the address book for the two islands, even though you still have the list of all the inhabitants. Alternatively, you might be uncertain because the individuals are recent shipwreck survivors, with one soon to be carried by unpredictable currents to the small island and the rest to the large one.[3] This version is illustrated in Table 2.

[3]For other contrasting pairs of opaque and non-opaque cases, see Hare 2013; 2016. Note that he does not discuss cases relevant to fanaticism.

	Calm (1/3)					Stormy (2/3)				
	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5
A	1	0	0	0	0	1	0	0	0	0
B	0	1	0	0	0	0	1	0	0	0
C	0	0	1	0	0	0	0	1	0	0
D	0	0	0	1	0	0	0	0	1	0
E	0	0	0	0	1	0	0	0	0	1
	Rescue Small Island									

	Calm (1/3)					Stormy (2/3)				
	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5
A	0	1	1	1	1	0	0	0	0	0
B	1	0	1	1	1	0	0	0	0	0
C	1	1	0	1	1	0	0	0	0	0
D	1	1	1	0	1	0	0	0	0	0
E	1	1	1	1	0	0	0	0	0	0
	Rescue Large Island									

Table 2: Opaque Risky Rescue

In this table, the two original columns have been subdivided into as many sub-columns as there are possible distributions of the people involved across the two islands, each sub-column given an equal share of the old column's probability. So the probability of any given cell is equal to the result of multiplying the probability of its column (either 1/3 or 2/3) by the probability of its sub-column (always 1/5). It is easy to verify that, in this opaque case, heading for the small island means that every affected person has a 3/15 probability of survival, while heading for the large island means that every affected person has a 4/15 probability of survival. So, in this case, heading for the large island gives every affected person a greater probability of being saved from death than heading for the small island. Plausibly, then, it is better to head for the large island in this case. This line of argument seems plausible enough as it stands. But it could also be supported by means of the following, more general principle:

> **The Greater-Chance Principle**. In rescue scenarios, it is better if every affected person has a greater probability of being saved from death.

This principle combines aspects of two other, more familiar principles: the principle of personal good (see Broome 1991, p. 165) and the better-chance condition (see Resnik 1987, p. 92). Here and below, I understand a rescue scenario to be one where every affected person can either live or die, with death equally bad, and life equally good, for everyone affected.

Now, note that the opaque case and the non-opaque case are analogous in the following way: in both, heading for the small island means that one person is saved no matter what, while heading for the large island means that four people are saved if the ocean is calm but zero are saved otherwise. It does not seem to matter who is saved. Plausibly, then, if it is better to head for the large island in the opaque case, it is likewise better to head for the large island in the non-opaque case. This line of argument seems plausible enough as it stands. But it could also be supported by means of three more general principles. The first is

> **Impartiality**. In rescue scenarios, saving some finite number of people from death is equally good as saving an equal number of other people from death.

This is a relatively undemanding principle of impartiality (compare, for example, Fleurbaey and Michel 2003). Here, it implies that, conditionally on each possible state of the world, the outcome of heading for the large island is equally good in the opaque case as in the non-opaque case, while the outcome of heading for the small island is likewise equally good in the opaque case as in the non-opaque case. The second principle is

> **Statewise Indifference**. Two options are equally good, if, conditionally on every possible state of the world, the outcomes of both are equally good.

This kind of principle has often been described as a basic demand of rationality (see, for example, Fleurbaey 2010, p. 655). Here, it implies that heading for the large island is equally good in the opaque case as in the non-opaque case, while heading for the small island is likewise equally good in the opaque case as in the non-opaque case. Lastly, to argue that heading for the large island in the opaque case is better only if it is better in the non-opaque case, we can appeal to

> **Transitivity**. If one option is at least as good as another which in turn is at least as good as a third, then the first option is at least as good as the third.

This principle has often been described as a conceptual truth (see, for example, Broome 2004, pp. 50–63).

So, given the specific numbers selected at the outset – with one person on the small island, four on the large island, and a one-in-three probability of reaching the large island

– it follows that heading for the large island is the better option. But, as I said before, the argument itself is wholly general. To see this, suppose that there are n people on the small island, m people on the large island, and the probability of reaching the large island is p. Thus, $n + m$ people are involved in total. Now, consider an opaque version of this case where, conditional on either state of the ocean, every possible distribution of the affected people across the two islands is equally probable as every other. It can be calculated that, in this opaque case, the probability that a given person is on the small island is equal to $\frac{n}{n+m}$, while the probability that a given person is on the large island is equal to $\frac{m}{n+m}$.[4]

So, in this opaque case, there are three possibilities for every affected person:

- they are on the small island, this has probability $\frac{n}{n+m}$;
- they are on the large island, and the ocean is calm, this has probability $p\frac{m}{n+m}$;
- they are on the large island, and the ocean is stormy, this has probability $(1-p)\frac{m}{n+m}$.

So heading for the small island means that every affected person has a $\frac{n}{n+m}$ probability of being saved from death, while heading for the large island means that every affected person has a $p\frac{m}{n+m}$ probability of being saved from death. So, for every number n and every positive probability p, heading for the large island means a higher probability of being saved for every affected person, as long as m is sufficiently large. The rest of the argument can then proceed as before.

2 Philosophical Payoffs

My argument should appeal to proponents of many views that otherwise appear hostile to fanatical conclusions. First, the argument is compatible with views that allow, or even

[4]This footnote provides the relevant calculations. The number of possible ways to distribute $n + m$ people across the two islands, one of which has population n, is equal to a quantity known as '$n + m$ choose n':

$$\binom{n+m}{n} = \frac{(n+m)(n+m-1)...(n+m-(n-1))}{n(n-1)(n-2)...(1)}$$

If a given person is to be on the small island, there remain $(n + m) - 1$ people to be distributed across the two islands across the two islands, with $(n - 1)$ people still to be distributed on the small island. The number of possible ways to do that is equal to a quantity known as '$(n + m) - 1$ choose $(n - 1)$':

$$\binom{n+m-1}{n-1} = \frac{(n+m-1)(n+m-2)...(n+m-1-((n-1)-1))}{(n-1)(n-2)...(1)}$$

Since every possible distribution is to be equally probable as every other, the probability that a given person is on the small island is equal to the second quantity divided by the first quantity, which simplifies to $\frac{n}{n+m}$.

require, individual people to discount sufficiently small probabilities. To see this, note that the part of the argument which considers what would be best for every affected person does not involve individual trade-offs between the magnitude of a good and its probability. Instead, the choice for every affected person is between a greater or a lesser probability of being saved from death. Even those inclined to discount sufficiently small probabilities would agree that, in this kind of case, it would be better for each person to have a greater probability of being saved from death, no matter how slight that increase might be.[5] My argument is, however, incompatible with views that discount sufficiently small probabilities in determining what is best overall, as opposed to what is best individually. This is because the argument does involve trade-offs between the number of people saved and the probability of saving anyone at all. But, in this case, discounting the unlikely possibility of saving many people seems implausible because it is a possibility that, as we saw, no individual person is plausibly allowed to discount (compare with cases discussed in Kosonen 2021).

Second, my argument is compatible with views that allow, or even require, imprecise probabilities in place of sharp ones. These views typically replace single-valued probability assignments with collections of multiple admissible ones, motivated by the idea that imprecise or incomplete evidence should be matched by imprecise or incomplete opinions. This raises the question of which option is best in a case where the probability of some event is imprecise in that way. A popular answer is that an option is better than another if, and only if, it is better relative to every admissible probability assignment (for discussion see, for example, Mogensen 2021). If this answer is correct, my argument still goes through, as long as the admissible probability of reaching the large island is not arbitrarily close to zero, that is, as long as it is at least ϵ, for some fixed positive ϵ, across all admissible probability assignments. The argument for fanaticism can then be applied to that ϵ in order to find a population size for the large island that would make heading there a better option. If heading for the large island is better at probability ϵ, it is also better at greater probabilities. So, for all admissible probability assignments, heading for the large island is the better option. The argument no longer goes through, however, if the admissible probability of reaching the large island is arbitrarily close to zero. In that case, heading for the large island might, at best, be no worse than heading for the small island (compare the discussion of Pascal's Wager in Hájek 2000).

Third, my argument does not presuppose risk neutrality. To see this, note that it is com-

[5] For example, Monton (2019, pp. 20–21) suggests supplementing his proposal of ignoring sufficiently small probabilities with dominance principles that would have this implication. Smith (2014, p. 499) is committed to denying these kinds of principles but only in cases involving infinitely many possible outcomes. Beckstead and Thomas (2023), on the other hand, describe a version of Monton's proposal that automatically satisfies these principles.

patible with the following, ostensibly risk-averse view:

> **Ex-Post Lexical View**. In cases which involve the same finitely many individuals as well as finitely many possible states of the world, an option is better than another if, and only if,
>
> (1) the lowest possible level of welfare is higher given the first option than given the second,
>
> (2) or, if these are equal, fewer people are expected to be at that level given the first option than given the second,
>
> (3) or, if these are also equal, the second-lowest possible level of welfare is higher given the first option than given the second,
>
> (4) or, if these are also equal, fewer people are expected to be at that level given the first option than given the second, and so on.
>
> Two options are equally good if, and only if, neither is better than the other.[6]

In other words, lower possible levels of welfare are to be given lexical priority, with ties broken by the expected number of people to be found at these levels. The ex-post lexical view is in some ways risk-averse despite implying a form fanaticism limited to rescue scenarios. For example, in cases that only involve one person or that involve no chance of inequality, it implies that it is worth making arbitrarily large sacrifices in expected total welfare for the sake of arbitrarily small increases in total welfare received in worst-case scenarios.

Fourth, the same example also shows that my argument does not presuppose anything as controversial as utilitarianism. The ex-post lexical view is anti-aggregative as well as inequality-averse. For example, it implies that no loss to a worse-off person can be balanced by any aggregate benefit to better-off people. So, while my argument is in

[6]This footnote provides a sketch of the argument for compatibility. First, consider the greater-chance principle. Note that the probability of a person's dying is equal to the expectation of an indicator function which takes value 1 if the person dies and 0 otherwise. Now, the sum of the expectations of these indicator values is equal to the expectation of their sum, which is simply the expected number of deaths. So, if everyone's probability of being saved is greater, the expected number of fatalities is lower. Next, consider impartiality. Note that if, in a rescue scenario, two options save the same number of people from dying, then they must agree with respect to the numbers of people at every possible welfare level. Next, consider statewise indifference. Note that, if two outcomes are equally good under the ex-post lexical view, then they must agree with respect to the numbers of people at every possible welfare level. So, if the outcomes of two options are equally good conditionally on every possible state of the world, the two options must agree with respect to the expected numbers of people at every possible welfare level. Lastly, the argument for transitivity is straightforward but involves multiple possible cases, depending on the welfare level at which the compared options diverge in terms of expected numbers of people.

some ways analogous to more general arguments for utilitarianism, its premises should be acceptable across a wider range of moral views.[7]

Still, it is worth pointing out that my argument only leads to fanaticism understood as an axiological view about what is best, as opposed to a deontic view about what ought to be done. While the argument can be recast in deontic terms, it might not be as compelling, as it would have to rely on a deontic version of transitivity, which is more controversial than the axiological version of transitivity assumed in my argument (compare, for example, Willenken 2012).

I expect that many of us will nonetheless remain uneasy about fanaticism, even if limited to rescue scenarios and understood in axiological terms. Wilkinson suggests that our uneasiness is untrustworthy because we tend to irrationally conflate very small probabilities (Wilkinson 2022, pp. 451–452). But I want to sketch a more charitable account of why our uneasiness is trustworthy, at least in some cases, without this constituting strong evidence against the truth of fanaticism in general.

Imagine, for example, that a reliable expert – not an eloquent mugger – tells you that by making a small donation you can save 100 quadrillion lives with a probability of 1 in 80 quadrillion. The trouble is that even trustworthy sources of information come with a margin of error. For example, if a thermometer shows that it is 20.0°C outside, you can only be certain that the true temperature is between 20.1°C to 19.9°C, corresponding to a margin of error of 0.1°C. Similarly, if the expert says that the probability is 1 in 80 quadrillion, you can only be certain that the true probability is within some margin of that, perhaps between 1 in 400 quadrillion and 9 in 400 quadrillion, which would make for a very thin margin of error of 1 in 100 quadrillion.[8]

Now, in many cases where we are comparing the certainty of a great good with a very small probability of a much greater good, even small variations within the margin of error correspond to large variations in expected effects. In the present scenario, you can only be certain that the expected number of lives saved by your donation is between 0.25 and 2.25. This would mean that, regardless of the truth of the matter, you should be quite uncertain that the option of saving 100 quadrillion lives with probability 1 in 80

[7] My argument is analogous to arguments for utilitarianism inspired by Harsanyi 1955 and developed by Broome 1991, Fleurbaey 2009, McCarthy et al. 2020, among others. But my premises are in many ways weaker and narrower in scope; for example, they do not include the controversial Bernoulli Hypothesis, to the effect that it is best for every person to maximize expected welfare, which is needed for a Harsanyi-style argument for utilitarianism; see, for example, Broome 1991, pp. 142–148. Similarly, while my argument is analogous to an argument for the Repugnant Conclusion discussed in Nebel 2019, that argument relies on the intrapersonal version of the Repugnant Conclusion, which is a controversial claim not needed for my argument.

[8] This would mean having second-order uncertainty about how much first-order uncertainty is appropriate in light of the available evidence (see, e.g., Williamson 2008). This is not the same as having an imprecise probability assignment (see Carr 2020).

quadrillion is better than the option of certainly saving one life. So, even if fanaticism is true, we should expect rational people to be quite uncertain about many specific fanatical conclusions. Our uneasiness about specific fanatical conclusions is therefore not necessarily evidence against the truth of fanaticism, but merely a reflection of our imperfect reliability at estimating probabilities (compare Monton 2019, p. 14; Gustafsson 2022, n. 24).

3 Infinity Problems

There are more serious problems for my fanatical argument. To see what they are, consider

> **Infinite Risky Rescue**. An infinite row of people is in harm's way. There are two ways of helping them. The first has a 1/3 probability of saving the first person in the row and nobody else, a 2/9 probability of saving the first two people in the row and nobody else, a 4/27 probability of saving the first four people in the row and nobody else, and so on. The other has a 1/3 probability of saving the second person in the row and nobody else, a 2/9 probability of saving the next two people in the row and nobody else, a 4/27 probability of saving the next four people in the row and nobody else, and so on.

This case is illustrated in Table 3.

	1/3	2/9	4/27	8/81	...	1/3	2/9	4/27	8/81	...
First one	1	1	1	1	...	0	0	0	0	...
Next one	0	1	1	1	...	1	0	0	0	...
Next two	0	0	1	1	...	0	1	0	0	...
Next four	0	0	0	1	...	0	0	1	0	...
Next eight	0	0	0	0	...	0	0	0	1	...
...	...					...				
	Help Initial People					Help Other People				

Table 3: Infinite Risky Rescue

Which way of helping is best? On the one hand, helping the initial people gives every affected person a greater probability of being saved from death. Plausibly, then, it is better to help in the first way rather than the second.[9] On the other hand, the same numbers of people are saved with the same probabilities either way. Both options save one person with probability 1/3, two people with probability 2/9, four people with probability 4/27, and so on. It does not seem to matter who is saved. Plausibly, then, both ways of helping are equally good. But now we have reasoned ourselves into a contradiction. This is puzzling.[10]

This also means that the premises of my earlier fanatical argument become mutually contradictory once infinitely many possible people have some positive probability of existence. So my earlier argument for fanaticism cannot be sound after all. This fits a familiar pattern whereby principles that are mutually consistent and supportive of compelling moral views become mutually contradictory when infinite numbers of people are involved (compare, for example, Van Liedekerke 1995).

A natural solution is to restrict one or more of the offending principles, for example, the greater-chance principle or statewise indifference. The former might be restricted to cases where at most finitely many people have a positive probability of existence, while

[9]This footnote provides the relevant calculations. As for helping the initial people (represented by the table on the left), the probability of saving the individuals corresponding to the first row is equal to the sum of the geometric series $\frac{1}{3}, \frac{2}{9}, \frac{4}{27}, \frac{8}{81}, \ldots$, initial term $\frac{1}{3}$ and common ratio $\frac{2}{3}$, which is 1. For the second row, it is equal to the sum of the series $\frac{2}{9}, \frac{4}{27}, \frac{8}{81}, \frac{16}{243}, \ldots$, with initial term $\frac{2}{9}$ and common ratio $\frac{2}{3}$, which is $\frac{2}{3}$. For the third row, it is equal to the sum of the series $\frac{4}{27}, \frac{8}{81}, \frac{16}{243}, \frac{32}{729}, \ldots$ with initial term $\frac{4}{27}$ and common ratio $\frac{2}{3}$, which is $\frac{4}{9}$. And so on, with the initial term, and thereby the sum, reduced by $\frac{2}{3}$ with each descending row. So, for the n-th row, the probability is equal to $\left(\frac{2}{3}\right)^{n-1}$. As for helping the other people (represented by the table on the right), the probability of saving the individuals corresponding to the first row is equal to 0. For the second row, it is equal to $\frac{1}{3}$. For the third row, it is equal to $\frac{2}{9}$. And so on. So, for the n-th row, except for the first, the probability is equal to $\frac{1}{3}\left(\frac{2}{3}\right)^{n-2}$. This is less than for the corresponding row of the other table. So, helping the initial people gives every affected person a greater probability of being saved than helping the other people.

[10]Similar puzzles have recently been independently developed by Goodsell 2021, Russell 2022, and Wilkinson 2023. It is also worth noting that a similar puzzle arises even if population is certain to be finite, as long as infinitely many possible people have a positive probability of existence. To see this, consider the following modification of the table in the main text, where '—' denotes nonexistence.

	1/3	2/9	4/27	8/81	...	1/3	2/9	4/27	8/81	...
First one	1	1	1	1	...	0	0	0	0	...
Next one	0	1	1	1	...	1	0	0	0	...
Next two	–	0	1	1	...	–	1	0	0	...
Next four	–	–	0	1	...	–	–	1	0	...
Next eight	–	–	–	0	...	–	–	–	1	...
...	...					...				
	Help Initial People					Help Other People				

Here, choosing the first option over the second means that every possible person has a greater probability of a longer life while having the same probability of having any life at all. Hence, this modified case involves no trade-offs between a person's probability of existence and the quality of their existence. Plausibly, then, the first option should come out better than the second. But, note that, conditionally on each possible state of the world, the number of people at any given welfare level is finite and fixed, regardless of the option chosen.

the latter might be restricted to cases where at most finitely many possible states of the world have a positive probability of being actual. The trouble is that these restricted principles seem arbitrary, casting doubt on the foundations of fanaticism itself. As Russell puts it when discussing similar issues for other arguments for fanaticism, 'the modifications and truncations seem like they might be telling us that there is something wrong about the underlying ideas.'[11]

I want to sketch a promising response to this worry. The response is to posit indeterminacy in our concept of betterness. The idea is that, since there is no determinate fact of the matter as to which of the offending principles should be restricted, there is no need to make an arbitrary-seeming choice between them; yet fanaticism is determinately true as it comes out true regardless of how this indeterminacy is resolved. This kind of appeal to indeterminacy has precedents in discussions of philosophy of physics, personal identity, and measurement of welfare (see, respectively, Field 1973, Johnston 1989, Greaves 2017). But I think that a more illuminating analogy comes from the history of the mathematics of the infinite.

Take, for example, the following argument that the collection of prime numbers between 6 and 10 – to wit {7} – is less numerous than the collection of prime numbers between 1 and 5 – to wit, {2,3,5}. First: {7} is equally numerous as {2}, since these two collections can be put into one-one correspondence. Next: {2} is less numerous than {2,3,5}, since the first is a proper sub-collection of the second. Finally: {7} must, by transitivity of the less-or-equally numerous' relation, be less numerous than {2,3,5}. Now, Galileo famously realized that the premises of this argument turn out to be mutually contradictory when applied to infinite collections. To use his example, the collection of squares of natural numbers, {1,4,9,...}, can be put into one-one correspondence with the collection of natural numbers themselves, {1,2,3,...}, yet the first is a proper sub-collection of the second (Galilei 1914, pp. 31–33). So, in the case of infinite collections, a conflict arises between the 'Cantorian' principle that the existence of a one-one correspondence is sufficient for equinumerosity, and the 'Aristotelian' principle that being a proper sub-collection is sufficient for being less numerous. This is puzzling since both principles remain intuitively plausible and apparently trustworthy as applied to finite collections.

A recent account of Galileo's puzzle posits that there is no determinate fact of the matter as to which of these principles is to be restricted. This is because there exist multiple, equally eligible ways to extend our intuitive concept of numerosity to infinite collections and all of them keep the intuitively plausible principles when restricted to finite

[11]See Russell 2023, discussing arguments for fanaticism presented in Wilkinson 2022 and Beckstead and Thomas 2023.

collections, but differ as to which one is dropped in the case of infinite collections.[12] This account does not force us to make an arbitrary-seeming choice as to which of the conflicting principles should be restricted (compare Parker 2008, Mancosu 2009).

The account also fits with a broader Lewisian account of the meaning of theoretical terms, according to which a theoretical term refers to the maximally eligible, maximally extensive satisfier of the relevant theoretical role. Eligible items are supposed to contrast with gerrymandered, grue-like items; a theoretical role is specified by a suitably abstract statement of the relevant theory. The reference of a term is indeterminate if there exist multiple, maximally eligible, maximally extensive satisfiers of the relevant theoretical role. A claim involving that term can nonetheless be said to be determinately true if it comes out true regardless of how the indeterminacy is resolved. Talk of indeterminacy is not redundant, as it can be cashed out in terms of eligibility and theoretical role. It is compatible both with classical and non-classical logic and semantics (compare Lewis 1999, but also Field 1973; Sider 2001).

I want to suggest that this kind of account might be applicable to many puzzles of infinite ethics.[13] In the context of the present puzzle, it would have to be true that there exist relations which satisfy as much as possible of the theoretical role of betterness – presumably including impartiality, transitivity, as well as restrictions of the greater-chance principle and statewise indifference – but which differ over these last two principles construed unrestrictedly. To use a simple example, these relations might be a relation of 'ex-ante utilitarian' betterness which compares options by totals of differences in expected individual welfare and a relation of 'ex-post utilitarian' betterness which compares options by expected totals of differences in individual welfare.[14] It would also have to be true that these relations are maximally eligible for the theoretical role of betterness, with neither being more gerrymandered than the other. In this way, the need to restrict the premises of my fanatical argument need not be telling us that there is something wrong about the underlying ideas. Instead, it could be telling us that there exist multiple, maximally eligible ways to extend our intuitive concept of betterness to cases that involve infinitely many possible people, all of which agree on fanaticism.

While this paper is not intended to be a full defence of this account, I do want to address one important objection, according to which the kind of indeterminacy in question is unacceptable in ethics, even if it is acceptable in mathematics. The worry is that, on

[12]Both the Cantorian and the Aristotelian concepts of numerosity are coherent, although the Cantorian concept is more orthodox; see, e.g., Benci and Nasso 2003 and Benci et al. 2006.

[13]Much of infinite ethics seems to consist in a tug of war between ethical intuitions directly analogous to the 'Cantorian' and the 'Aristotelian' principles; see, e.g., Hamkins and Montero 2000 versus Vallentyne 1995.

[14]Ex-ante utilitarian views are discussed by Bostrom 2011, Arntzenius 2014, and Meacham 2020; ex-post utilitarian views are discussed by Wilkinson 2023.

this account, many disputes in infinite ethics will turn out to be merely verbal. If a question about betterness turns out to lack a determinate answer on grounds that there exist multiple, maximally eligible, maximally extensive satisfiers of the theoretical role of betterness, then it seems that any further dispute involving this question should count as merely verbal, lacking substance. This is because the answer can reveal neither which relation is especially eligible, nor which relation is especially suited to play the theoretical role of betterness; it can at best reveal a fact about the reference of certain words. This might be implausible insofar as ethical disputes are expected to never be merely verbal.[15]

I sympathize with this worry. In response, I want to point out that the account is compatible with the substantiveness of ethical questions in many cases as well as with the presence of other marks of ethical objectivity in all cases. First, as far as this paper is concerned, questions about betterness in cases that involve at most finitely many possible people will never turn out to be merely verbal. Moreover, it might always be a mind-independent matter whether the different relations which might play the theoretical role of betterness hold or not.

4 Conclusion

This paper has developed a neglected kind of argument for a form of fanaticism limited to rescue scenarios. This argument should be acceptable across a wide range of moral views, including some forms of probability discounting, imprecise probabilism, risk-aversion, and aggregation scepticism. Any remaining intuitive reluctance about specific fanatical conclusions need not count as strong evidence against fanaticism itself but might instead reflect our imperfect reliability at estimating probabilities. Moreover, while cases involving infinitely many possible people reveal incompatibility between some of the ideas underlying fanaticism, it is plausible to see this as a reflection of indeterminacy in our concept of betterness rather than as a problem for fanaticism. This suggests that fanaticism might, after all, be a surprising truth about our ethical predicament.

References

Frank Arntzenius. Utilitarianism, Decision Theory and Eternity. *Philosophical Perspectives*, 28(1):31–58, 2014. doi: 10.1111/phpe.12036.

[15] For an account of nonsubstantiveness, see Sider 2011, pp. 44–66; see also similar accounts of merely verbal disputes in Sidelle 2007 and Chalmers 2011.

Nick Beckstead and Teruji Thomas. A Paradox for Tiny Probabilities and Enormous Values. *Noûs*, 2023. doi: https://doi.org/10.1111/nous.12462.

Vieri Benci and Mauro Di Nasso. Numerosities of Labelled Sets: a New Way of Counting. *Advances in Mathematics*, 173(1):50–67, 2003.

Vieri Benci, Mauro Di Nasso, and Marco Forti. An Aristotelian Notion of Size. *Annals of Pure and Applied Logic*, 143(1):43–53, 2006. doi: 10.1016/j.apal.2006.01.008.

Nick Bostrom. Pascal's Mugging. *Analysis*, 69(3):443–445, 2009. doi: 10.1093/analys/anp062.

Nick Bostrom. Infinite Ethics. *Analysis and Metaphysics*, 10:9–59, 2011.

John Broome. *Weighing Goods: Equality, Uncertainty and Time*. Basil Blackwell, Oxford, 1991.

John Broome. *Weighing Lives*. Oxford University Press, Oxford, 2004.

Jennifer Rose Carr. Imprecise Evidence without Imprecise Credences. *Philosophical Studies*, 177(9):2735–2758, 2020. doi: 10.1007/s11098-019-01336-7.

David Chalmers. Verbal Disputes. *The Philosophical Review*, 120(4):515–566, 2011. doi: 10.1215/00318108-1334478.

Hartry Field. Theory Change and the Indeterminacy of Reference. *The Journal of Philosophy*, 70(14):462–481, 1973. doi: 10.2307/2025110.

Marc Fleurbaey. Two Variants of Harsanyi's Aggregation Theorem. *Economics Letters*, 105(3):300–302, 2009. doi: 10.1016/j.econlet.2009.08.023.

Marc Fleurbaey. Assessing Risky Social Situations. *Journal of Political Economy*, 118(4): 649–680, 2010. doi: 10.1086/656513.

Marc Fleurbaey and Philippe Michel. Intertemporal Equity and the Extension of the Ramsey Criterion. *Journal of Mathematical Economics*, 39(7):777–802, 2003. doi: 10.1016/S0304-4068(03)00054-5.

Galileo Galilei. *Dialogues Concerning Two New Sciences*. MacMillan, New York, 1914. Originally published 1638. Translated by Henry Crew and Alfonso de Salvio.

Zachary Goodsell. A St Petersburg Paradox for Risky Welfare Aggregation. *Analysis*, 81 (3):420–426, 2021. doi: 10.1093/analys/anaa079.

Hilary Greaves. A Reconsideration of the Harsanyi–Sen–Weymark Debate on Utilitarianism. *Utilitas*, 29(2):175–213, 2017. doi: 10.1017/S0953820816000169.

Johan E. Gustafsson. Our Intuitive Grasp of the Repugnant Conclusion. In Gustaf Arrhenius, Krister Bykvist, Tim Campbell, and Elizabeth Finneron-Burns, editors, *The Oxford Handbook of Population Ethics*, pages 371–389. Oxford University Press, 2022.

Joel David Hamkins and Barbara Montero. With Infinite Utility, More Needn't be Better. *Australasian Journal of Philosophy*, 78(2):231–240, 2000. doi: 10.1080/00048400012349511.

Caspar Hare. *The Limits of Kindness*. Oxford University Press, 2013.

Caspar Hare. Should We Wish Well to All? *Philosophical Review*, 125(4):451–472, 2016. doi: 10.1215/00318108-3624764.

John C. Harsanyi. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy*, 63(4):309–321, 1955. doi: 10.1086/257678.

Alan Hájek. Objecting Vaguely to Pascal's Wager. *Philosophical Studies*, 98(1):1–16, 2000. doi: 10.1023/a:1018329005240.

Mark Johnston. Fission and the Facts. *Philosophical Perspectives*, 3:369–397, 1989. doi: 10.2307/2214274.

Petra Kosonen. Discounting Small Probabilities Solves the Intrapersonal Addition Paradox. *Ethics*, 132(1):204–217, 2021. doi: 10.1086/715276.

David Lewis. Reduction of Mind. In *Papers in Metaphysics and Epistemology*, volume 2, pages 291–324. Cambridge University Press, Cambridge, 1999. Originally published in 1994.

Paolo Mancosu. Measuring the Size of Infinite Collections of Natural Numbers: Was Cantor's Theory of Infinite Number Inevitable? *Review of Symbolic Logic*, 2(4): 612–646, 2009. doi: 10.1017/S1755020309990128.

David McCarthy, Kalle Mikkola, and Teruji Thomas. Utilitarianism with and without Expected Utility. *Journal of Mathematical Economics*, 87:77–113, 2020. doi: 10.1016/j.jmateco.2020.01.001.

Christopher J. G. Meacham. Too Much of a Good Thing: Decision-Making in Cases with Infinitely many Utility Contributions. *Synthese*, 2020. doi: 10.1007/s11229-019-02522-0.

Andreas L. Mogensen. Maximal Cluelessness. *Philosophical Quarterly*, 71(1):141–162, 2021. doi: 10.1093/pq/pqaa021.

Bradley Monton. How to Avoid Maximizing Expected Utility. *Philosophers' Imprint*, 19, 2019. URL `http://hdl.handle.net/2027/spo.3521354.0019.018`.

Jacob M. Nebel. An Intrapersonal Addition Paradox. *Ethics*, 129(2):309–343, 2019. doi: 10.1086/700085.

Matthew W. Parker. Philosophical Method and Galileo's Paradox of Infinity. In Bart Van Kerkhove, editor, *New Perspectives on Mathematical Practices: Essays in Philosophy and History of Mathematics*. World Scientfic, 2008.

Michael D. Resnik. *Choices: an Introduction to Decision Theory*. University of Minnesota Press, Minneapolis, 1987.

Jeffrey Sanford Russell. Problems for Impartiality, 2022. URL `https://globalprioritiesinstitute.org/parfit-memorial-lecture-jeffrey-sanford-russell-problems-for-impartiality/`. The Derek Parfit Memorial Lecture, University of Oxford.

Jeffrey Sanford Russell. On Two Arguments for Fanaticism. *Noûs*, 2023. doi: 10.1111/nous.12461.

Alan Sidelle. The Method of Verbal Dispute. *Philosophical Topics*, 35(1):83–113, 2007. doi: 10.5840/philtopics2007351/25.

Theodore Sider. Criteria of Personal Identity and the Limits of Conceptual Analysis. *Noûs*, 35:189–209, 2001. doi: 10.1111/0029-4624.35.s15.10.

Theodore Sider. *Writing the Book of the World*. Oxford University Press, Oxford, 2011.

Nicholas J. J. Smith. Is Evaluative Compositionality a Requirement of Rationality? *Mind*, 123(490):457–502, 2014. doi: 10.1093/mind/fzu072.

Peter Vallentyne. Infinite Utility: Anonymity and Person-Centredness. *Australasian Journal of Philosophy*, 73(3):413–420, 1995. doi: 10.1080/00048409512346761.

Luc Van Liedekerke. Should Utilitarians Be Cautious About an Infinite Future? *Australasian Journal of Philosophy*, 73(3):405–407, 1995. doi: 10.1080/00048409512346741.

Hayden Wilkinson. In Defense of Fanaticism. *Ethics*, 132(2):445–477, 2022. doi: 10.1080/00048402.2021.2013265.

Hayden Wilkinson. Infinite Aggregation and Risk. *Australasian Journal of Philosophy*, 101(2):340–359, 2023. doi: 10.1080/00048402.2021.2013265.

Tim Willenken. Deontic Cycling and the Structure of Commonsense Morality. *Ethics*, 122(3):545–561, 2012. doi: 10.1086/664750.

Timothy Williamson. Why Epistemology Cannot be Operationalized. In Quentin Smith, editor, *Epistemology: New Essays*. Oxford University Press, 2008. doi: 10.1093/acprof:oso/9780199264933.003.0012.