Expected Value Fanaticism

Petra Kosonen

Guest essays represent only the views of the author(s).

Table of Contents

Tiny probabilities of great value

When we make decisions, we often do not know which choices result in which outcomes: our choices involve risk. In decision theory, [Expected Utility Maximization](#) is the standard way to respond to choices involving risk. An option's expected value (or utility) is obtained by weighing the potential values (or utilities) of its outcomes by their probabilities and summing these up. Expected Utility Maximization then tells rational agents to choose the option with the highest expected value (or utility).

Consider the following example. Charity A saves the life of one person with certainty, while Charity B gives a 1% chance of saving 1000 people. [1] Which charity should you donate to? Charity B will probably not save anyone, with a 99% probability. However, in expectation, it saves 10 lives, while Charity A only saves one life. Expected Utility Maximization therefore directs us to choose Charity B.

Expected Utility Maximization can give counterintuitive recommendations in cases that involve tiny probabilities of great value. Instead of choosing Charity A, Expected Utility Maximization would also direct us to choose Charity C, which gives a 0.1% chance of saving 10,000 people, as Charity C also saves 10 lives in expectation. We can keep lowering the probability of saving any lives at all as long as we compensate for this by increasing the number of lives potentially saved—Expected Utility Maximization would still advise us to choose these options over Charity A. [2] For example, Expected Utility Maximization would advise us to choose Charity Z, which gives probability 10^{-26} of saving octillion (10^{27}) lives, over Charity A. Choosing Charity Z means we would be letting someone die for a minuscule probability of saving (a great number of) lives. Many find this implication deeply counterintuitive.

Consider another case (*Devil at Your Deathbed*).[3] You are down to your last year of life when the devil shows up with an offer: he will give you ten happy years instead, with a 0.999 probability (otherwise immediate death). You accept the offer. Then he comes back offering a hundred years of happy life— 10 times as long—with a 0.999^2 probability—just 0.1% lower. You accept again. After some 50,000 such trades, you end up with a $0.999^{50,000}$ probability of $10^{50,000}$ years of happy life. With a chance of success less than one in 10^{21}, you predictably die soon thereafter—even though you could have chosen, for example, ten billion years of happy life with an over 0.99 probability of success.

This is an example of what is called *fanaticism* in decision theory. Informally, fanaticism is the idea that tiny probabilities of great value can matter a great deal in practical decision-making. More formally, it states that

> **Fanaticism**. For any non-zero probability p, and for any finite amount of utility u, there is some sufficiently large utility U such that probability p of U (and otherwise nothing) is better than certainty of u.[4]

Pick any (finitely good) outcome—call it Utopia. Call *Mega Utopia* an outcome that is even better than Utopia. Fanaticism then says that no matter how good Utopia is, it would be better to get a tiny chance of Mega Utopia than certainty of Utopia, no matter how small the chance of Mega Utopia is (provided that Mega Utopia is sufficiently better than Utopia).

Arguments for fanaticism

The continuum argument

Given how counterintuitive fanaticism is, one might wonder why anyone would accept this view. But there are some strong arguments in its favour.[5] In fact, we have already seen one argument for fanaticism: Devil at Your Deathbed.

This case shows that two plausible principles, *Anti-Timidity* and *Transitivity,* imply fanaticism. [6]

> **Anti-Timidity.** Any tiny decrease in the probability of a good outcome can be compensated for by increasing the payoff sufficiently.
>
> **Transitivity**. If *A* is better than *B*, which is better than *C*, … , which is better than *Z*, then *A* is better than *Z*.

Anti-Timidity seems plausible; however, repeated applications of Anti-Timidity together with Transitivity lead to fanaticism. In Devil at Your Deathbed, each of the devil's offers appeared more appealing than the previous one: the potential payoff increased substantially, whereas the probability of getting it decreased only ever so slightly. Anti-Timidity therefore tells us to accept each new offer. However, by accepting each of these offers, one ends up exchanging a guaranteed one year of life for a minuscule chance of receiving a great outcome. This is an example of fanaticism. Thus, to avoid fanaticism, one must either reject Anti-Timidity or Transitivity—but both principles seem intuitively appealing.

More is Better, Simple Separability and Transitivity imply fanaticism

Fanaticism also follows from the following three plausible principles: Transitivity, *More is Better* and *Simple Separability*. [7]

> **More is Better**. It is better to have a much higher chance of many more happy lives (and otherwise nothing) than a smaller chance of fewer happy lives (and otherwise nothing).
>
> **Simple Separability**. Uncertainty over what happens in distant places is not relevant to what we ought to do when we cannot affect what happens in those distant places (at least when our available prospects only involve a finite number of possible outcomes).

More is Better seems obvious: it is better to get a higher probability of more good things than a smaller probability of fewer good things. Simple Separability is plausible too. It would be very odd if our uncertainty over what happens on Kepler-186f some 500 light-years away could affect which prospects we should choose even when we have no way of affecting what takes place on that exoplanet.

However, these very plausible principles lead to fanaticism. To see how, consider the following prospects (see Table 1):

> **More vs. Less**. There are three states the world could be in. Let the probability of state 1 be p, the probability of state 2 be q and the probability of state 3 be $1-p-q$. Also, let p be a much greater probability than q.
>
> *More.* Gives a great number of happy lives N in state 1 (and nothing in states 2 and 3).
>
> *Less.* Gives a small number of happy lives n in state 2 (and nothing in states 1 and 3).

Table 1

	State 1	State 2	State 3
Probability	p	q (much smaller than p)	$1 - p - q$
More	N	0	0
Less	0	n	0

Next, suppose that you here on Earth have a choice between More and Less, and you think that the inhabitants of Kepler-186f—call them keplerians—face the following prospect (see Table 2):

Kepler. Gives a small number of happy lives n in state 1 (and nothing in states 2 and 3).

Table 2

	State 1	State 2	State 3
Probability	p	q (much smaller than p)	$1 - p - q$
Kepler	n	0	0

Given that keplerians face prospect Kepler, your choice is effectively between a combination of More and Kepler versus a combination of Less and Kepler (let's denote these by More & Kepler and Less & Kepler, respectively):

Table 3

	State 1	State 2	State 3
Probability	p	q (much smaller than p)	$1 - p - q$
More & Kepler	$N + n$	0	0
Less & Kepler	n	n	0

Now, recall that the principle More is Better says that it is better to have a much higher chance of many more happy lives than a smaller chance of fewer happy lives. It follows from More is Better that More is indeed better than Less: More gives a great number of lives in a higher-probability state 1 and Less gives a small number of lives in a lower-probability state 2 (see Table 1).

Next, recall that Simple Separability says that uncertainty over what happens in distant places (like exoplanets) is not relevant to what we ought to do when we cannot affect what happens in those distant places (at least when our

available prospects only involve a finite number of possible outcomes). It follows from this principle that the fact that keplerians face prospect Kepler cannot change our ranking of More and Less: given that More is better than Less (by the principle More is Better), it follows by Simple Separability that More & Kepler must be better than Less & Kepler.

However, as can be seen from Table 3, More & Kepler gives a slightly lower probability p of a much larger number of happy lives $N+n$. Therefore, More is Better and Simple Separability lead to Anti-Timidity: any tiny decrease (from $p+q$ to p) in the probability of a good outcome (happy lives) can be compensated for by increasing the payoff sufficiently (many more happy lives). In the previous section, we saw how Anti-Timidity together with Transitivity leads fanaticism. Therefore, More is Better, Simple Separability and Transitivity also lead to fanaticism. To avoid fanaticism, one must reject one of these principles.

Stochastic Dominance and Simple Separability imply fanaticism

A related argument for fanaticism is that it follows from Simple Separability and another compelling principle: *Stochastic Dominance*.[8]

> Stochastic Dominance. If, for all possible outcomes, prospect A is at least as likely as prospect B to yield an outcome that is at least this good, then A is at least as good as B. If, in addition, there is some outcome for which A is more likely than B to yield an outcome at least this good, then A is better than B.

Suppose, for example, that A gives a 50% chance of saving one life and a 50% chance of saving three lives, while B gives a 50% chance of saving one life and a 50% chance of saving two lives. Both A and B certainly save at least one life, and they both give a 50% chance of saving at least two lives. However, in addition, A gives a 50% chance of saving at least three lives, while B gives this a 0% chance. A is therefore better than B by Stochastic Dominance: A *stochastically dominates* B. Similarly, suppose that C gives a 40% chance of

saving one life and a 60% chance of saving two lives, while D gives a 50% chance of saving one life and a 50% chance of saving two lives. Both C and D certainly save at least one life, but C gives a 60% chance of saving at least two lives, while D only gives a 50% chance of saving at least two lives. Therefore, C stochastically dominates D.

To understand how Simple Separability and Stochastic Dominance lead to fanaticism, consider the following prospects (see Table 4):

> ### Safe vs. Risky
>
> *Safe*. Guarantees one happy life.
>
> *Risky*. Gives a non-zero probability p of $n + 1$ happy lives (a great outcome).

Let p be very tiny. Then, this case is fanatical, so what you ought to choose between Safe and Risky is linked to whether or not fanaticism is true: if Risky is better than Safe, then fanaticism is true.

While you face the choice between Safe and Risky, this time the keplerians face the following prospect (see Table 4):

> *Kepler**. Gives a probability p of nothing, probability q of one happy life, probability q of two happy lives, probability q of three happy lives, … , probability q of n happy lives, where probability q is smaller than probability p.

Table 4

Probability	p (larger than q)	q	q	q	…	q
Safe	1	1	1	1	…	1
Risky	$n + 1$	0	0	0	…	0
Kepler*	0	1	2	3	…	n

Given that keplerians face prospect Kepler*, your choice (in terms of the overall value of the Universe) is effectively between Safe & Kepler* versus Risky & Kepler*:

Table 5

Probability	p (larger than q)	q	q	...	q
Safe & *Kepler**	1	2	3	...	n + 1
Risky & *Kepler**	n + 1	1	2	...	n

Note that probability p is greater than probability q. We can therefore split the second column of Table 5 into two columns, one giving probability $p-q$ and the other probability q:

Table 6

Probability	p-q	q	q	q	...	q
Safe & *Kepler**	1	1	2	3	...	n + 1
Risky & *Kepler**	n + 1	n + 1	1	2	...	n

Lastly, we can reorder the outcomes of Risky & Kepler* that are associated with probability q by shifting each of them to the column on its left side (see Table 7). The leftmost outcome under probability q—namely, $n + 1$—is then moved to the rightmost column (where n appears in Table 6).

Table 7

Probability	p-q	q	q	q	...	q
Safe & *Kepler**	1	1	2	3	...	n + 1

Probability	$p-q$	q	q	q	...	q
Risky & *Kepler**	$n+1$	1	2	3	...	$n+1$

After re-arranging the columns, we can now see clearly from Table 7 that the only difference between Safe & Kepler* and Risky & Kepler* is that the former gives probability $p-q$ of getting one happy life, whereas the latter gives the same probability of getting $n + 1$ happy lives. Because obtaining $n + 1$ happy lives is better than obtaining just one, Risky & Kepler* stochastically dominates Safe & Kepler*: for all possible outcomes, Risky & Kepler* is at least as likely as Safe & Kepler* to yield an outcome that is at least this good, and there is one outcome (namely, $n + 1$ happy lives) for which Risky & Kepler* is more likely to yield an outcome at least this good. Therefore, by Stochastic Dominance, Risky & Kepler* is better than Safe & Kepler*.

Finally, given that Risky & Kepler* is better than Safe & Kepler*, Simple Separability implies that Risky is better than Safe—so fanaticism is true. This is because Simple Separability says that Risky is better than Safe if and only if Risky & Kepler* is better than Safe & Kepler*. So, by Simple Separability, Risky & Kepler* cannot be better than Safe & Kepler* unless Risky is also better than Safe.

To conclude, Simple Separability and Stochastic Dominance lead to fanaticism.[9] Anyone wishing to reject fanaticism must reject either Simple Separability or Stochastic Dominance.

Stochastic Dominance, Negative Reflection and Background Independence imply fanaticism

Next, we will see that Stochastic Dominance, together with two other plausible principles, implies fanaticism:[10]

> **Negative Reflection**. If prospect A is not better than prospect B conditional on any possible answer to some question Q, then A cannot be better than B unconditionally.

> **Background Independence**. Some prospect A is better than another prospect B if and only if the combined prospect A plus some far away outcome O is also better than the combined prospect B plus the same far away outcome O.

Suppose that you are somewhere tropical and warm. The weather forecast says that either there will be heavy rain or burning sunshine. You have an umbrella that can protect against both. If it rains, you prefer having the umbrella. And if there is burning sunshine, you are neutral about having the umbrella and not having it. It would be odd if you nonetheless prefer not bringing the umbrella: no matter what happens, you will be happy or neutral about having brought it! Negative Reflection tells us that you should indeed bring the umbrella.

Background Independence, in turn, is similar to (Simple) Separability: the difference between these two principles is that Separability tells us that far-away *uncertain prospects* do not matter to what we ought to do, while Background Independence says that far-away *final outcomes* do not matter to what we ought to do. According to Separability, you can ignore your uncertainty over what might happen on Kepler-186f when you make decisions here on Earth; according to Background Independence, you can ignore your knowledge of the final outcomes on this distant planet.

In the previous section, we saw how Risky & Kepler* stochastically dominates Safe & Kepler*. By Stochastic Dominance, Risky & Kepler* is therefore the better of these two.[11] Negative Reflection then tells us that Risky & Kepler* must have some possible outcome such that the combined prospect Risky plus a far away outcome O is better than the combined prospect Safe plus the same far away outcome O (let's denote these by Risky & Kepler-O and Safe & Kepler-O, respectively). It would be odd if you would choose Risky because Risky &

Kepler* is better than Safe & Kepler* in expectation, but if you learn what happens on Kepler-186f—no matter what ultimately happens—you would no longer choose Risky over Safe. That would mean that as long as you have uncertainty about what happens on Kepler-186f, you would choose Risky. But the moment this uncertainty is resolved, and you learn what exactly is happening on this planet, Risky is no longer the better option. Surely learning more about the lives of keplerians should not change what you ought to do on Earth!

Now, given that Risky & Kepler-O is better than Safe & Kepler-O (by Stochastic Dominance and Negative Reflection), Background Independence implies that Risky is better than Safe: this principle says that Risky is better than Safe if and only if Risky & Kepler-O is better than Safe & Kepler-O. Given that Risky is better than Safe when the final outcome on Kepler-186f is O (no matter what O is), Risky must be overall better than Safe. What the final outcome on some faraway planet is should not change our ranking of Risky and Safe.

Finally, recall that the choice between Risky and Safe is fanatical; thus, if Risky is better than Safe, fanaticism is true. We just learned that Risky is better than Safe by Stochastic Dominance, Negative Reflection and Background Independence. Therefore, these three principles imply fanaticism.

To summarize, this section has explored how various plausible principles lead to fanaticism. These are:

1. Anti-Timidity and Transitivity.

2. More is Better and Simple Separability.

3. Stochastic Dominance and Simple Separability.

4. Stochastic Dominance, Negative Reflection and Background Independence.

The next section will challenge some of these arguments.

Contra arguments for fanaticism

You may be tempted to conclude that fanaticism is true based on these arguments: it follows from very plausible principles! How could we conclude that even the tiniest decrease in probability cannot be compensated for by massively increasing the payoff? Or that uncertainty over or knowledge of what happens on Kepler-186f could influence what we ought to do here on Earth? Not so fast. Many of the very plausible principles discussed in the previous sections are, in fact, mutually inconsistent, or they imply that fanaticism is false.

Stochastic Dominance and Separability are jointly inconsistent

Recall how Simple Separability and Stochastic Dominance imply fanaticism. However, Stochastic Dominance and a generalization of Simple Separability are, in fact, jointly inconsistent.[12] This inconsistency calls into question the argument for fanaticism that relies on Stochastic Dominance and Simple Separability.

Simple Separability only applies to prospects that have a finite number of possible outcomes. In contrast, Separability applies to all prospects, including ones that have an infinite number of possible outcomes. Otherwise, these principles say the same thing:

> **Separability**. Uncertainty over what happens in distant places is not relevant to what we ought to do when we cannot affect what happens in those distant places.

To understand why Stochastic Dominance and Separability are jointly inconsistent, let's examine the following variations of the *St. Petersburg game* (see Table 8):

> **St. Petersburg Games**. A fair coin is flipped until a heads appears.
>
> *St. Petersburg*. Gives 2^n happy lives, where n is the number of coin flips.

> *St. Petersburg⁻*. Gives $2^n - 1$ happy lives, where n is the number of coin flips.

Table 8

No. of flips	1	2	3	…
Probability	1/2	1/4	1/8	…
St. Petersburg	2	4	8	…
St. Petersburg⁻	1	3	7	…

Both versions of this game have infinite expected value (or utility):

> St. Petersburg: $1/2 \times 2 + 1/4 \times 4 + 1/8 \times 8 \ldots = 1 + 1 + 1 \ldots = \infty$

> St. Petersburg⁻: $1/2 \times 1 + 1/4 \times 3 + 1/8 \times 7 \ldots = 1/2 + 3/4 + 7/8 \ldots = \infty$

St. Petersburg⁻ has the same probability distribution as the standard St. Petersburg game, yet the outcomes are worse. It is clear that St. Petersburg is preferable to St. Petersburg⁻; indeed, this follows from Stochastic Dominance.

Next, by Separability, combining two instances of St. Petersburg (one here on Earth and another on Kepler-186f) is better than combining two instances of St. Petersburg⁻ (one here on Earth and another on Kepler-186f).[13]

However, it is possible to set up the games in such a way that St. Petersburg & St. Petersburg is stochastically equivalent to St. Petersburg⁻ & St. Petersburg⁻: they give the same probabilities of the same outcomes. Concretely, in this setup, the outcomes of the two St. Petersburg games depend on flipping a dime, whereas the outcomes of the two St. Petersburg⁻ games depend on flipping the same dime and a penny (see Table 9).

> **Correlated St. Petersburg Games**: A dime is flipped repeatedly until it lands heads. A penny is flipped once. Both coins are fair.
>
> *Earth and Kepler St. Petersburg*. Each game gives 2^n happy lives, where n is the number of flips with the dime.
>
> *Earth St. Petersburg⁻*. If the penny comes up heads, the game gives one happy life. Otherwise, it gives twice as many happy lives as St. Petersburg, minus one.
>
> *Kepler St. Petersburg⁻*. If the penny comes up tails, the game gives one happy life. Otherwise, it gives twice as many happy lives as St. Petersburg, minus one.

Table 9

Outcome	H, 1	H,2	H,3	...	T,1	T,2	T,3	...
Probability	1/4	1/8	1/16	...	1/4	1/8	1/16	...
Earth St. Petersburg	2	4	8	...	2	4	8	...
Kepler St. Petersburg	2	4	8	...	2	4	8	...
Earth St. Petersburg⁻	1	1	1	...	3	7	15	...
Kepler St. Petersburg⁻	3	7	15	...	1	1	1	...

'H' and 'T' represent the result of flipping the penny, and '1', '2', ... indicate how many times the dime is flipped.

Notice that both Earth St. Petersburg and Kepler St. Petersburg yield the same probabilities for the same outcomes as the St. Petersburg game in Table 8: probability 0.5 of two happy lives, probability 0.25 of four happy lives, and so

on. Likewise, both Earth St. Petersburg⁻ and Kepler St. Petersburg⁻ yield the same probabilities for the same outcomes as the St. Petersburg⁻ game in Table 8: probability 0.5 of one happy life, probability 0.25 of three happy lives, and so on. As usual, let the prospect before "&" indicate what happens on Earth and the prospect after "&" indicate what happens on Kepler-186f. It follows that the combined prospect St. Petersburg & St. Petersburg should be better than the combined prospect St. Petersburg⁻ & St. Petersburg⁻ (by Stochastic Dominance and Separability). To get the values of the outcomes of the combined prospect St. Petersburg & St. Petersburg, we simply add the number of happy lives given by Earth St. Petersburg and Kepler St. Petersburg in each column; similarly, to get the values of the outcomes of the combined prospect St. Petersburg⁻ & St. Petersburg⁻, we add the number of happy lives given by Earth St. Petersburg⁻ and Kepler St. Petersburg⁻ in each column (see Table 10).

But, as seen from Table 10, St. Petersburg & St. Petersburg gives the same probabilities of the same outcomes as St. Petersburg⁻ & St. Petersburg⁻. In each possible state of the world, both prospects give the same number of happy lives. Consequently, St. Petersburg & St. Petersburg is stochastically equivalent to St. Petersburg⁻ & St. Petersburg⁻. By Stochastic Dominance, each is at least as good as the other, so they are equally good.

Table 10

Outcome	H, 1	H,2	H,3	...	T,1	T,2	T,3	...
Probability	1/4	1/8	1/16	...	1/4	1/8	1/16	...
St. Petersburg & St. Petersburg	4	8	16	...	4	8	16	...
St. Petersburg⁻ & St. Petersburg⁻	4	8	16	...	4	8	16	...

Let's go through the argument again. By Stochastic Dominance, St. Petersburg is better than St. Petersburg⁻. By Separability, St. Petersburg & St. Petersburg must then be better than St. Petersburg⁻ & St. Petersburg⁻. However, St. Petersburg & St. Petersburg is stochastically equivalent to St. Petersburg⁻ & St. Petersburg⁻, which implies that they should be equally good. Yet they cannot be equally good when St. Petersburg & St. Petersburg is also better than St. Petersburg⁻ & St. Petersburg⁻. This contradiction shows that Stochastic Dominance and Separability cannot both hold.[14]

Recall that Stochastic Dominance and Simple Separability together entail fanaticism. However, we have just seen that Stochastic Dominance and a generalization of Simple Separability are inconsistent together. Yet whatever reasoning supports Simple Separability should also justify its generalization.[15] Unless there is a special reason to endorse Simple Separability that does not also apply to its generalized form, one has no grounds for accepting Simple Separability if one rejects Separability. Moreover, Stochastic Dominance shows that Separability is wrong. Therefore, without such a unique justification for Simple Separability, we can reject the argument for fanaticism based on Stochastic Dominance and Simple Separability.

However, instead of using Simple Separability to argue for fanaticism, one may use the following Separability principle instead:

> **Separability for Independent Prospects**. Uncertainty over what happens in distant places is not relevant to what we ought to do when we cannot affect what happens in those distant places and our available prospects are probabilistically independent of the prospects those distant places face.

This principle, together with Stochastic Dominance, also leads to fanaticism.[16] Furthermore, this principle is not inconsistent with Stochastic Dominance, as the St. Petersburg prospects discussed in this section are not probabilistically independent of each other: what happens with one of them correlates with

what happens with the others. Consequently, this version of Separability does not apply. So, if we can find a justification for this version of Separability that does not also apply to generalized Separability, we have an argument for fanaticism. One such justification could be that we get to keep Stochastic Dominance and at least some form of Separability without inconsistencies – these are plausible principles after all.[17]

Stochastic Dominance and Negative Reflection imply fanaticism is false

Next, recall how Stochastic Dominance, Negative Reflection and Background Independence imply fanaticism. However, if fanaticism holds, then Stochastic Dominance and Negative Reflection turn out to be jointly inconsistent.[18] Consequently, the argument for fanaticism using these principles as premises cannot be sound.

To understand why Stochastic Dominance and Negative Reflection are inconsistent if fanaticism is true, let's consider the following variations of the St. Petersburg game:

> **St. Petersburg Games**. A fair coin is flipped until a heads appears.
>
> *St. Petersburg*. Gives 2^n happy lives, where n is the number of coin flips.
>
> *St. Petersburg+*. Gives $2^n + 1$ happy lives, where n is the number of coin flips.

Table 11

No. of flips	1	2	3	...
Probability	1/2	1/4	1/8	...
St. Petersburg	2	4	8	...

No. of flips	1	2	3	...
St. Petersburg+	3	5	9	...

The outcomes of St. Petersburg[+] are better than those of St. Petersburg. Stochastic Dominance thus tells us that St. Petersburg[+] is better than St. Petersburg. But, if fanaticism is true, then none of St. Petersburg[+]'s finite *outcomes* are as good as the St. Petersburg *prospect*. This is because both St. Petersburg and St. Petersburg[+] are better than any finite payoff.

Now, recall that Negative Reflection says the following: if prospect *A* is not better than prospect *B* conditional on any possible answer to some question *Q*, then *A* cannot be better than *B* unconditionally. Consider the question: "What is the outcome of St. Petersburg[+]?" By Negative Reflection, it follows that St. Petersburg[+] cannot be better than St. Petersburg once you consider all possible ways St. Petersburg[+] could turn out. No matter the outcome of St. Petersburg[+], this outcome will be worse than the prospect St. Petersburg. Yet, by Stochastic Dominance, St. Petersburg[+] is better than St. Petersburg. Thus, if fanaticism is true, Stochastic Dominance or Negative Reflection must be false; they are inconsistent. Therefore, Stochastic Dominance and Negative Reflection together imply that fanaticism is false.[19]

Fanaticism is vulnerable to money pumps

Suppose fanaticism holds on to Stochastic Dominance while abandoning Negative Reflection. Then it is dynamically inconsistent and susceptible to money pumps: agents following this theory will pay to keep something that they could keep for free.[20] For example, imagine you start with St. Petersburg[+]. After you learn the outcome of St. Petersburg[+], you can pay $100 to exchange whatever outcome you received for the prospect St. Petersburg. You would accept this trade because any finite outcome of St. Petersburg[+] is worse than

having the St. Petersburg prospect. However, before you find out how St. Petersburg$^+$ turned out, you can instead pay \$50 to stick with St. Petersburg$^+$ no matter what, in which case you will receive no future offers. You know that if you do not pay the \$50, you will end up with St. Petersburg and be \$100 poorer overall. By contrast, if you pay the \$50, you keep St. Petersburg$^+$ and you will have paid only \$50. Stochastic Dominance therefore tells you to pay the \$50 to avoid paying \$100 later. But in so doing, you have become a money pump: you have paid to keep something (St. Petersburg$^+$) that you could have kept for free by simply declining all offers. Thus, fanaticism combined with Stochastic Dominance is vulnerable to money pumps.

To summarize, this section has discussed arguments that undermine fanaticism. First, it showed that Stochastic Dominance and (generalized) Separability are jointly inconsistent. Next, it showed that Stochastic Dominance and Negative Reflection together imply that fanaticism is false. Finally, it showed that fanaticism can be exploited by money pumps (given Stochastic Dominance). It is up for debate whether we can draw conclusions about the justifiability of these principles in finite cases from infinite cases, such as those discussed in this section.

Alternatives to fanaticism

Finally, something has to be said about alternatives to fanaticism. Note that these alternatives face serious problems, such as violations of Separability and vulnerability to money pumps. Indeed, they have to violate plausible principles, given that fanaticism can be derived from such principles.

Bounded Utilities

One alternative to fanaticism is that utilities are bounded above and below. Boundedness of utilities means that

> **Boundedness.** There are some highest and some lowest possible values (or utilities) that an outcome can have.

Boundedness excludes the possibilities of unbounded or infinitely good and bad outcomes. Standard axiomatizations of Expected Utility Maximization typically impose the requirement that utilities be bounded.[21] As there is an upper (lower) cap on how good (bad) an outcome can be, Boundedness prevents high-value (low-value) outcomes from dominating the expected utility calculations; thus, it avoids fanaticism.[22] However, note that if the upper (lower) cap is very high (low), then Boundedness can still lead to verdicts that seem intuitively fanatical.

Discounting Small Probabilities

Another alternative to fanaticism is discounting small probabilities:[23]

> **Discounting Small Probabilities**. Very small probabilities should be discounted down to zero.

The simplest version of this view says that you should conditionalize on very-small-probability outcomes not happening, and then maximize expected utility. As tiny probabilities have been discounted down to zero, tiny probabilities of great value (or utility) will contribute nothing to expected value (or utility); thus, discounting small probabilities avoids fanaticism.[24]

Knowledge-Based Discounting

Another alternative to fanaticism is Knowledge-Based Discounting, which says that

> **Knowledge-Based Discounting**. First conditionalize on your knowledge and then maximize expected utility.

Knowledge-Based Discounting avoids fanaticism if you *know* that the very-small-probability outcomes will not occur. In that case, you should conditionalize on such outcomes not happening, and tiny probabilities of great value (or utility) will contribute nothing to expected value (or utility).[25]

Stochastic Dominance as Sufficient

Another alternative to fanaticism is using Stochastic Dominance as a sufficient principle of rationality.[26]

> **Stochastic Dominance as Sufficient**. You are instrumentally rational if and only if your preferences satisfy Stochastic Dominance.

Here's a remarkable result: if some prospect A has higher expected utility than another prospect B, A will stochastically dominate B if there is enough background uncertainty—even if A would not stochastically dominate B without this background uncertainty. However, if these prospects involve tiny probabilities of great value (or utility), background uncertainty generates stochastic dominance less readily. Thus, using Stochastic Dominance as a sufficient principle of rationality can justify the intuition that we are often permitted to reject fanatical prospects. However, Stochastic Dominance can sometimes still require us to choose fanatically.

Conclusion

We have seen how fanaticism follows from various plausible principles. However, then we saw how many of these principles are mutually inconsistent or, in fact, imply that fanaticism is false. The debate between supporters and critics of fanaticism remains inconclusive, as there are compelling arguments for and against this view. But, as Russell (2024, p. 5) writes, "Whatever the truth of the matter, the ethics of huge numbers is deeply weird and full of surprises."

About the Author

Petra Kosonen is a Postdoctoral Fellow at the Population Wellbeing Initiative, University of Texas at Austin. Her research is mostly on ethics and decision theory. She earned a DPhil in Philosophy from Oxford University. The central theme of her thesis was how we should treat tiny probabilities of vast value.

How to Cite This Page

Resources and Further Reading

- Beckstead, N. and Thomas, T. (2024). A paradox for tiny probabilities and enormous values. *Noûs*, *58*(2), 431–455.

- Carlsmith, J. and Balasubramanian, V. (2024). Expected Utility Maximization. https://www.utilitarianism.net/guest-essays/expected-utility-maximization/

- Cibinel, P. (2023). A dilemma for Nicolausian discounting. *Analysis 83*(4), 662–672.

- Fishburn, P. C. (1970). *Utility Theory for Decision Making*. Wiley, New York.

- Goodsell, Z. (2021). A St Petersburg paradox for risky welfare aggregation. *Analysis 81*(3), 420–426.

- Hájek, A. (2014). Unexpected expectations, *Mind 123*(490), 533–567.

- Hammond, P. J. (1998). Objective expected utility: A consequentialist perspective. In S. Barberà, P. J. Hammond and C. Seidl (Eds.), *Handbook of Utility Theory Volume 1: Principles* (pp. 143–211). Kluwer, Dordrecht.

- Hong, F. (2024). Know Your Way Out of St. Petersburg: An Exploration of "Knowledge-First" Decision Theory. *Erkenntnis 89*(6), 2473–2492.

- Isaacs, Y. (2016). Probabilities cannot be rationally neglected, *Mind 125*(499), 759–762.

- Kosonen, P. (2024). Probability Discounting and Money Pumps. *Philosophy and Phenomenological Research 109*(2), pp. 593-611.

- Kosonen, P. (2022). *Tiny probabilities of vast value*. [Doctoral dissertation, University of Oxford]. Oxford University Research Archive. https://ora.ox.ac.uk/objects/uuid:822703dc-56ba-4717-98b4-663d251e8acb

- Kreps, D. M. (1988). *Notes on the Theory of Choice*. Westview Press, Boulder.

- Lundgren, B. and Stefánsson, H. O. (2020). Against the De Minimis principle. *Risk Analysis 40*(5), 908–914.

- Monton, B. (2019). How to avoid maximizing expected utility. *Philosophers' Imprint 19*(18), 1–24.

- Russell, J. S. (2024). On two arguments for fanaticism. *Noûs, 58*(3), 565–595.

- Russell, J. S. and Isaacs, Y. (2021). Infinite prospects. *Philosophy and Phenomenological Research, 103*(1), 178–198.

- Smith, N. J. J. (2014). Is evaluative compositionality a requirement of rationality?. *Mind 123*(490), 457–502.

- Tarsney, C. (2024). Against anti-fanaticism. *Philosophy and Phenomenological Research*, 1–20.

- Tarsney, C. (2020). Exceeding expectations: Stochastic dominance as a general decision theory. Global Priorities Institute Working Paper No. 3–2020. *https://globalprioritiesinstitute.org/christian-tarsney-exceeding-expectations-stochastic-dominance-as-a-general-decision-theory/*

- Wilkinson, H. (2022). In defence of fanaticism. *Ethics 132*(2), 445–477.

- Wilkinson, H. (2024). Egyptology and fanaticism. Philosophical Studies 181(8), 1903-1923.

1. This case is from Carlsmith and Balasubramanian (2024). ↵

2. Expected utility theory with a bounded utility function can avoid this implication (more on bounded utilities later). ↵

3. Beckstead and Thomas (2024, pp. 1–2). ↵

4. Wilkinson (2022, p. 449). ↵

5. The presentation of the arguments in this article is somewhat similar to Kosonen (2022, pp. 1–59). ↵

6. See Beckstead and Thomas (2024, §1) for the original argument. ↵

7. This argument is also from Beckstead and Thomas (2024, §3.2). Its presentation follows Russell (2024, §2, theorem 1). ↵

8. See Wilkinson (2022, §VI A) for the original version of the argument. The presentation follows closely Russell (2024, pp. 23–25, theorem 3). Also see Tarsney (2020), Goodsell (2021) and Beckstead and Thomas (2024). ↵

9. This way of presenting the argument assumes that the number but not the location of happy lives is what matters. See Russell (2024) for the complete argument. ↵

10. This argument is originally from Wilkinson (2022). The presentation follows closely Russell (2024). ↵

11. Wilkinson (2022, §VI) uses a different case to show that a fanatical prospect stochastically dominates a non-fanatical prospect given a certain kind of background uncertainty. However, for simplicity, I will use the case discussed here. See also Tarsney (2020) on this topic. I wish to thank Kacper Kowalczyk for suggesting the use of the simpler case. ↵

12. See Russell (2024) for the original argument. ↵

13. Here's the explanation: St. Petersburg is better than St. Petersburg^{-}, so, by Separability, the combined prospect St. Petersburg & St. Petersburg is better than St. Petersburg^{-} & St. Petersburg. The fact that Kepler-186f faces St. Petersburg should not alter the ranking between St. Petersburg and St. Petersburg^{-} here on Earth.

   Next, as St. Petersburg is better than St. Petersburg^{-}, St. Petersburg^{-} & St. Petersburg is better than St. Petersburg^{-} & St. Petersburg^{-}. The fact that we on Earth face St. Petersburg^{-} should not alter the ranking between St. Petersburg and St. Petersburg^{-} on Kepler-186f.

   Now we have that St. Petersburg & St. Petersburg is better than St. Petersburg^{-} & St. Petersburg, and that St. Petersburg^{-} & St. Petersburg is better than St. Petersburg^{-} & St. Petersburg^{-}. It follows by transitivity that St. Petersburg & St. Petersburg is better than St. Petersburg^{-} & St. Petersburg^{-}. ↵

14. As before, this way of presenting the argument assumes that the number, but not the location, of happy lives is what matters. See Russell (2024) for the complete argument. ↵

15. Russell (2024, p. 10). ↵

16. See Wilkinson (2022) and Wilkinson (2024) for the argument. ↵

17. Wilkinson (2024). See Wilkinson (2024) for other possible justifications of Separability for Independent Prospects. ↵

18. This argument is from Russell (2024, §3) and Russell and Isaacs (2021). Also see Beckstead and Thomas (2020, §4). ↵

19. Stochastic Dominance and Negative Reflection together entail that fanaticism is false. Consequently, they cannot be used together to defend fanaticism. However, a principle similar to Negative Reflection—together

with Stochastic Dominance and Background Independence—implies that either fanaticism or negative fanaticism is true. See Russell (2024, pp. 29−31). ↵

20. See Russell and Isaacs (2021, p. 4 n. 5) for the original argument. Russell and Isaacs (2021) show that fanaticism violates Countable Independence —a principle similar to Negative Reflection. ↵

21. See, for example, Kreps (1988, pp. 63−64), Fishburn (1970, pp. 194, 206−207), Hammond (1998, pp. 186−191). ↵

22. On bounded utilities, see Beckstead and Thomas (2024), Kosonen (2022, pp. 32−38 and Ch. 1) and Tarsney (2024). ↵

23. This is called 'Nicolausian discounting' or 'probability discounting'. See Monton (2019) and Kosonen (2022). ↵

24. For defences of discounting small probabilities, see Monton (2019) and Smith (2014). For criticism of discounting small probabilities, see Hájek (2014), Isaacs (2016), Lundgren and Stefánsson (2020), Kosonen (2022, pp. 137−239), Cibinel (2023), Beckstead and Thomas (2024), Kosonen (2024) and Tarsney (2024). ↵

25. On Knowledge-Based Discounting, see Kosonen (2022, pp. 48−51) and Hong (2024). ↵

26. This view is from Tarsney (2020). ↵