Egyptology and Fanaticism*

Hayden Wilkinson

Forthcoming in *Philosophical Studies.* When available, please cite published version.

Abstract

Various decision theories share a troubling implication. They imply that, for any finite amount of value, it would be better to wager it all for a vanishingly small probability of some greater value. Counterintuitive as it might be, this *fanaticism* has seemingly compelling independent arguments in its favour. In this paper, I consider perhaps the most *prima facie* compelling such argument: an *Egyptology argument* (an analogue of the Egyptology argument from population ethics). I show that, despite recent objections from Russell (2023) and Goodsell (2021), the argument's premises can be justified and defended, and the argument itself remains compelling.

Keywords: *fanaticism; separability; Egyptology objection; risk aversion; expected value theory; unbounded utility.*

*For their extremely helpful feedback on this paper, I am grateful to Jacob Barrett, Zachary Goodsell, Ina Jäntgen, Jakob Lohmar, Jeffrey S. Russell, Teru Thomas, and two anonymous reviewers, as well as audiences at the University of Texas at Austin, the University of Stirling, and the Pacific Division Meeting of the American Philosophical Association.

1 Fanaticism

Consider a small probability, perhaps one in 1,000.[1] Which is better: to save the life of one person for sure; or to have a probability of one in 1,000 of saving some very large number of lives, many more than 1,000? Or consider an even smaller probability—perhaps one in one million. Which is better: to save that one life for sure; or to have a probability of one in one million of saving some vast number of lives? At some point, as the probability of success gets closer and closer to zero, it may seem *fanatical* to claim that the latter option is better, *even if* arbitrarily many lives would be saved if it succeeded.

Nonetheless, fanatical verdicts follow from various widely accepted theories of instrumental (moral) betterness. For instance, *expected* (moral) *value theory* says that one option is better than another if it has a greater probability-weighted sum of (moral) value[2]—a greater *expected value*. Combine this with any theory of moral betterness that attributes equal value to each additional life saved, and the expected value of the low-probability option can always be greater—saving one life for sure won't be as good as saving N lives with some tiny probability ε, no matter how tiny ε is, so long as N is great enough.

Beyond expected value theory as well, other theories can lead to the same verdict. For instance, take expected *utility* theory: whereby options are compared according to their probability-weighted sum of *utility*, where utility can be any increasing[3] function of moral value. One key difference from expected value theory is that we could let additional units of moral value count for less and less utility and expected utility theory could then exhibit risk aversion. But it would still uphold the fanatical verdict above—as long as the chosen utility function is unbounded (with respect to the number of lives saved), expected utility theory will agree that saving some vast number N of lives with probability ε is better than saving one life for sure, no matter how tiny ε is, so long as N is great enough. Likewise, Buchak's (2013) *risk-weighted* expected utility theory—whereby each option's expected utility is further transformed to account for its riskiness, using a 'risk' function—will say the same for many possible risk functions.[4]

In general, the implication of such theories that they imply such fanatical verdicts can be char-

[1]Throughout, I will remain agnostic on exactly how probability is interpreted. Are the probabilities I speak of the agent's subjective degrees of belief, or the probabilities that an idealised agent with the same evidence would assign, or the objective physical chances of particular outcomes? For my purposes, it does not matter.

[2]I will assume that, whatever the true account of moral betterness over outcomes, it admits some privileged cardinal representation of moral value. Theories such as (the various forms of) totalism, averageism, prioritarianism, and egalitarianism certainly do. Where I speak of outcomes with *finite* or *arbitrarily large* values below, I mean finite value or arbitrarily large value on that privileged cardinal representation.

[3]Strictly speaking, a utility function that sometimes decreases or remains level might be considered compatible with expected utility theory. But such a utility function would lead to violations of Stochastic Dominance (see Section 2), and so would be implausible.

[4]Specifically, it will say so for any risk function that is both increasing and continuous, as Buchak (2013, 61) suggests any rational risk attitude must be.

acterised as *Fanaticism.*

> *Fanaticism*: For any finite probability $\varepsilon > 0$, no matter how small, and for any finite value v, there is some large enough finite V such that an option described by O_{risky} is better than O_{safe}.[5]
>
> O_{risky}: an outcome with value V with probability ε; an outcome with value 0 otherwise
>
> O_{safe}: an outcome with value v with probability 1

For my purposes here, we can define an option as a function from states of the world to outcomes. Each such state of the world has some probability of obtaining (together summing to 1), and each such probability is the same no matter which option we choose. Given this, we can represent a pair of options such as O_{risky} and O_{safe} as below, in a decision matrix. Here, the rows correspond to different options, while the columns correspond to different states of the world. At the top of each column is the probability of the relevant state obtaining, and in each cell is the value of the outcome that results from that option in that state.

	ε	$1-\varepsilon$
O_{risky}	V	0
O_{safe}	v	v

By intuition, it may seem deeply implausible that O_{risky} could be better than O_{safe} *no matter how tiny* ε might be (so long as V is large enough)—it may seem deeply implausible that Fanaticism is true. But there are compelling arguments in favour of it, particularly when we consider decisions characterised by the *moral* value of outcomes (as will be my focus here).

Perhaps the single most persuasive such argument is the *Egyptology argument* for Fanaticism, a version of which I present in the next section. But this form of argument is controversial—it has been subject to troubling objections (specifically, from Russell, 2023). The aim of this paper is to evaluate the Egyptology argument and the leading objections to it. As it turns out, such objections can be avoided. Egyptology arguments can, with some care, provide a compelling case for Fanaticism.

[5]I assume that our theory of moral betterness assigns at least cardinal values to outcomes (see note 2 above). But if that theory assigns *only* cardinal values to outcomes (as against values on a ratio or unit scale), then Fanaticism can be interpreted as requiring O_{risky} to be better than O_{safe} *for all* triples of value that 0, v, and V might pick out. (Note that this would make Fanaticism a very strong principle, equivalent to what Evershed, n.d. calls *Super Fanaticism.*) I will assume throughout only that value has a cardinal structure, but what I say will be compatible with it having a richer structure too.

2 Egyptology arguments

When comparing options in terms of moral betterness, the following seems overwhelmingly plausible: if I face a decision here and now in the 21st Century, and the available options differ *only* in the events that occur in the future, the comparison of those options must not depend on distant past events that are unaffected. (By *event*, I mean anything that occurs at a specific time and place in spacetime.) The comparison must not depend, for instance, on unaffected events in ancient Egypt. Likewise, it seems overwhelmingly plausible that the comparison must not depend on other events that are unaffected by and unrelated to your choice, such as events in sufficiently distant galaxies.

Such claims may seem innocuous, but have formed the basis of a well-known objection to some theories of moral value. Such theories include (standard, welfarist) *averageism*: the view that an outcome is valued in proportion to the average lifetime welfare of each person within it. By averageism, when you face a decision of whether or not to bring additional people into existence, to determine whether doing so makes the world better or worse you must ask yourself whether it will increase or decrease the average welfare of everyone who has ever existed. And this depends on how good the lives of the ancient Egyptians were. Even though your decision may not affect the events of ancient Egypt in the slightest, which option is better can still depend on what happened there.[6] But this is absurd, as ethicists have long noted (McMahan, 1981: p. 115; Parfit, 1984: p. 420). It seems obvious that matters of Egyptology are irrelevant to present-day decision-making. Likewise, it seems obvious that any events causally isolated from and unrelated to our actions are also irrelevant, be they in the distant past or in distant galaxies.

We might codify the intuitions behind this objection in the form of *Separability for Outcomes*. Note that, by an *outcome*, I mean a comprehensive history of how the world turns out (independent of how else it *might* have turned out). Such a history includes every morally relevant event that would then ever occur. And, here and below, the symbol $\succcurlyeq$ denotes the relation of 'at least as good as' over both outcomes and risky options.[7] Likewise, $\succ$ denotes strict betterness and $\sim$ equal goodness.

> *Separability for Outcomes*: For any outcomes A, B, and C, $A \succcurlyeq B$ if and only if $A \oplus C \succcurlyeq B \oplus C$.

Here, $\oplus$ is a concatenation operation. $A \oplus C$ denotes an outcome that contains (morally equivalent duplicates of) all of the (morally significant) events within A as well as (such duplicates of) all such

[6]Similar claims hold under versions of egalitarianism as well as under various welfarist theories of betterness on which welfare depends on desire-satisfaction. By the latter, whether it is better to light a candle in the tomb of an ancient Roman general can depend on whether that ancient general preferred for candles to be lit in their tomb long after their death.

[7]Assume that this relation is both reflexive and non-symmetric on both such sets, but not that it is transitive or complete on the set of options.

events within C.[8] Applying this terminology to an idealised present-day decision, A and B might represent the outcomes resulting from the available actions, differing only in the effects one has on the future. C could then represent some relatively impoverished outcome consisting *only* of some events happening in ancient Egypt. Concatenate C with each of A and B—consider modified versions of A and B in which some different morally significant events in ancient Egypt are added—and, intuitively, we would like the comparison of outcomes to remain unchanged. The comparison of $A \oplus C$ and $B \oplus C$ must match that of A and B, as Separability for Outcomes says.

Separability for Outcomes may seem a mild condition to place on our comparisons of outcomes. But it has broad implications. Assume, as I will throughout, that a welfarist view of moral betterness holds—that betterness is determined by the distribution of welfare over individual persons in each outcome. Then, in conjunction with several other basic assumptions,[9] Separability for Outcomes implies an *additive* theory of moral betterness: that the value of any concatenated outcome $A \oplus C$ is equal to the sum of the value of A and the value of C (Thomas 2022, §5; see also Wilkinson n.d.a, §5) .[10] For this reason, I will assume from here on that, if some form of Separability holds (and each of the below forms of Separability do indeed imply Separability for Outcomes) then an additive theory of moral betterness holds.

That tangent aside, how is Separability for Outcomes relevant to the case for Fanaticism? It turns out that an analogous principle, and a case seemingly analogous to that above, allow us to make a very similar argument for fanatical verdicts. That analogous principle is *Separability for Options* (from Russell, 2023, §2).

Separability for Options For any options X, Y, and Z, $X \succcurlyeq Y$ if and only if $X \oplus Z \succcurlyeq Y \oplus Z$.

To interpret this principle, we need to define the $\oplus$ operation over not just outcomes but also over risky options. For options, we can define concatenation as occurring *within* respective states of the world: $X \oplus Y$ is the option that maps each state of the world s to the outcome $X(s) \oplus Y(s)$,

[8]Note that many location-specific events will inevitably be incompatible: e.g., the event (or collection of more fine-grained events that together constitute) "Nelson died at Trafalgar in 1805" and the event (or collection of events) "Nelson survived Trafalgar in 1805". But, in concatenating sets of such events, I assume that we do away with some of their morally insignificant features, such as the exact time and place at which they in fact occurred. For instance, if outcome A contains Nelson dying at Trafalgar in 1805 and another outcome B contains him surviving, then we might construct $A \oplus B$ such that Nelson survives Trafalgar while another person morally equivalent (and perhaps even qualitatively identical) to Nelson perishes.

[9]Those assumptions include *Solipsist's Pareto*, *Anonymity*, and that the set of possible outcomes is sufficiently unconstrained.

> *Solipsist's Pareto*: If two outcomes A and B both contain *only* one person p, $A \succcurlyeq B$ if and only if p's well-being is at least as great in A as in B.
>
> *Anonymity*: If there is a bijection σ from the set of persons who exist in outcome A to the set of those who exist in outcome B such that each person is mapped to someone with precisely the same well-being in their respective outcome, then $A \sim B$.

[10]For similar results, see Blackorby et al. (2005, thm 6.10) and Pivato (2014, thm 1).

where $X(s)$ is the outcome of X in state s and $Y(s)$ the outcome of Y in s.

Here is both an illustration of the $\oplus$ operation over options and an argument for Fanaticism. Take options X and Y, where X has some (perhaps very high) probability p of vast value V and Y has some very small probability ε of only a tiny amount of value v, and both result in value 0 otherwise.

	p	ε	$1-p-\varepsilon$
X	V	0	0
Y	0	v	0

And take a third option B, which gives some small value v in that first state with probability p, and value 0 otherwise. Here is what we get when we concatenate X with B and Y with B.

	p	ε	$1-p-\varepsilon$
B	v	0	0
$X \oplus B$	$V+v$	0	0
$Y \oplus B$	v	v	0

Here, Separability for Options tells us that $X \oplus B$ is at least as good as $Y \oplus B$ *if* X is at least as good as Y. And, indeed, X is better than Y; X gives a high probability of a lot of value, while Y gives a much lower probability of much less value. That the former is better then follows from to a very weak and very plausible principle: *Stochastic Dominance.*

> *(First-Order) Stochastic Dominance*: Let X and Y be any two options. If, for every possible outcome O, $Pr(X(s) \succcurlyeq O) \geq Pr(Y(s) \succcurlyeq v(O))$, then $X \succcurlyeq Y$.
>
> If, as well, there is some possible outcome O such that $Pr(X(s) \succcurlyeq O) > Pr(Y(s) \succcurlyeq O)$, then X is strictly better than Y.

Stated less formally, the principle says that if two options give exactly the same probabilities of the same outcomes (or equally good outcomes) then they are equally good; and if you then swap out any of those outcomes in one option for even better ones, then you make the option strictly better. Looking back at X and Y, we could take Y and swap out the outcome with probability ε for a better one with value V—this would make for a better option, according to Stochastic Dominance—and we could keep swapping out more states, replacing value 0 with value V, until we ended up with an option resembling X (with exactly the same probabilities of the same values). Since this would be equally as good as X, and better than Y, X must be better than Y.

And then, given Separability for Options, we know that $X \oplus B$ must be better than $Y \oplus B$. But here's the catch: if we deny Fanaticism, it can't be, at least not for all such p and ε. The reason it can't comes from Beckstead and Thomas (2023): if it were, we could construct a continuum argument for Fanaticism. We could start with O_{safe} and compare it to another option with a slight probability ε of an outcome with value 0, but the remaining probability $1-\varepsilon$ of an outcome with as much value as you like. And we could then compare that option to another, with a slightly higher probability 2ε of value 0 but vastly more value again otherwise. And we could keep going, comparing each option to yet another, all the way until we are comparing an option to some version of O_{risky}: an option with probability arbitrarily close to 1 of resulting in value 0, and some arbitrarily large amount of value otherwise. But we cannot say that the next option in the sequence is always better—that the option after O_{safe} is better than it, and that the next option is even better, and so on until O_{risky} is better than what came before it. Then, O_{risky} would be better than O_{safe}, giving us Fanaticism. At least, that's the verdict we must reach if the betterness relation $\succcurlyeq$ is transitive (as seems overwhelmingly plausible).[11] So, to deny Fanaticism, while maintaining that $\succcurlyeq$ is transitive, there must be some option in the sequence that is no better than its predecessor, *no matter how much greater* the value the later option could potentially result in. Whatever that later option is, call it $X \oplus B$, and the one before it in the sequence $Y \oplus B$. For some probabilities p and ε and values v and $v+V$, they will match the decision matrix above.

So, if Fanaticism is false, for at least some well-chosen X and Y, we must conclude that $X \oplus B$ is not better than $Y \oplus B$. But X is better than Y. And for consistency—by Separability for Options—the two verdicts must match! So we have a contradiction. If Separability for Options and Stochastic Dominance are true, and if $\succcurlyeq$ is transitive, then Fanaticism must be true.

To make this argument more concrete and more intuitive, we might suppose that the valuable events in B occur in ancient Egypt—with probability p, some event that contributes additional value v occurred on the banks of the Nile several thousand years ago. And suppose that X and Y correspond to options between which you might be choosing today, resulting in overall value V, v, or 0 for the rest of the world, *sans* that one event on the banks of the Nile all those years ago. To make an overall, impartial evaluation of the available options, as various moral theories require, we must consider the prospects of the overall value of the world—we must compare $X \oplus B$ to $Y \oplus B$. But, intuitively, the verdict of this comparison must not differ from that which we reach when comparing X to Y. Although it does *matter* morally whether events in ancient Egypt featured happiness or suffering, our present-day decisions must not be *sensitive* to such unaffected happiness or suffering. Likewise, our decisions must not be sensitive to other events in the world that are unaffected by and unrelated (in any plausible sense) to our actions. And yet, if we deny Fanaticism, our evaluations *are* sometimes sensitive to such events: we must sometimes compare X and Y one way but $X \oplus B$ and $Y \oplus B$ another. Thus, any theory that denies Fanaticism will fall prey to an analogue of the

[11] John Broome (2004, pp. 50-1), at least, considers it a necessary and self-evident truth that comparative relations like $\succcurlyeq$ are transitive. Further arguments in favour of the transitivity of betterness are given by, e.g., Huemer (2008) and Nebel (2018).

Egyptology objection we saw above.[12]

3 The problem with Separability

But there is a serious problem with this version of the Egyptology argument for Fanaticism. As Russell (2023, §2) shows, we cannot accept the argument's conclusion and its premises at the same time: Fanaticism is incompatible with the conjunction of Stochastic Dominance and Separability for Options! Russell's argument goes as follows, applying results from Seidenfeld et al. (2009) and Goodsell (2021).

If Fanaticism holds, then we can construct what can be called a *generalised St Petersburg game*: an option that maps infinitely many different states to different outcomes, in which each outcome has only finite value, but which is better than *any* finite value for sure. (The exact method for specifying those values is laid out in Beckstead and Thomas, 2023: §4 but isn't crucial for my purposes here.) For simplicity, I will just consider the traditional St Petersburg game, as given below. But, in general, different fanatical theories may require much higher values than those listed in this decision matrix.

	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	...
St Petersburg	2	4	8	16	32	...

Consider two modified versions of this option. The first is H: if a flipped coin lands heads then you play the St Petersburg game; if tails, you obtain value 0. The second is T: if that same coin lands tails, you play the St Petersburg game; if heads, you obtain value 0.

	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...
H	2	4	8	...	0	0	0	...
T	0	0	0	...	2	4	8	...

We can easily compare H and T. By even mere Stochastic Dominance, they are equally good.

But then consider the concatenations $H \oplus H$ and $T \oplus H$.[13] And recall that, if Separability for Options holds, they must then be equally good as well.

	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...
$H \oplus H$	4	8	16	...	0	0	0	...
$T \oplus H$	2	4	8	...	2	4	8	...

[12] To my knowledge, a version of this argument for Fanaticism was first given by Beckstead (2013, p. 137), but this early version assumed the axioms of expected utility theory. More general versions of arguments similar to this have since been developed independently by Wilkinson (2022b) and Beckstead and Thomas (2023). The version given above is adapted from the latter (although, in their presentation, they do not appeal to Stochastic Dominance nor Separability directly, but instead to intuitions about particular cases).

[13] This case is adapted from Seidenfeld et al. (2009; cf. Goodsell 2021).

Do these two concatenated options still seem equally good? If the answer is yes, consider what they look like when we rearrange the states in which each outcome occurs, as below. Note that such rearrangements have the exact same probabilities of the same values as the options from which they're rearranged.

	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...
$\sigma(H \oplus H)$	0	4	8	16	...
$\sigma'(T \oplus H)$	2	4	8	16	...

The only difference between the two rearrangements $\sigma(H \oplus H)$ and $\sigma'(T \oplus H)$ is that, in one state with probability $\frac{1}{2}$, one results in value 0 and the other in value 2. In every other state, they result in the same value. Stochastic Dominance applies: $\sigma'(T \oplus H)$ is strictly better than $\sigma(H \oplus H)$. And these are equivalent to $T \oplus H$ and $H \oplus H$, respectively—they have the same probabilities of the same values, so Stochastic Dominance applies in just the same way. $T \oplus H$ must be strictly better than $H \oplus H$.

This is despite H and T being equally good. So, by concatenating each of those options with H, we can change their ranking—we have a violation of Separability for Options. And that is the problem: Fanaticism, Stochastic Dominance, and Separability for Options cannot all hold. Put differently, Stochastic Dominance and Separability for Options together imply that Fanaticism is false. But above, under the assumption that $\succcurlyeq$ is transitive, they implied that Fanaticism is true! So Stochastic Dominance and Separability for Options are incompatible.

So, the Egyptology argument presented above must be unsound. If an argument takes Stochastic Dominance and Separability for Options among its premises, but those same premises cannot both be true (unless $\succcurlyeq$ is not transitive). It seems that the Egyptology argument for Fanaticism—what might seem the most compelling argument for Fanaticism in the literature—fails.

4 Weaker Separabilities

The Egyptology argument given in Section 2 above does indeed fail. Separability for Options and the various other premises lead to Fanaticism no more than they lead to contradiction. But this is not to say that *all* such Egyptology arguments fail—that there is *no* combination of similar premises that imply Fanaticism and are themselves well-supported. In particular, by replacing Separability for Options with a weaker principle, can we salvage the argument?[14] That is what I seek to do below.

[14] Alternatively, we could attempt to replace Stochastic Dominance with a weaker principle and preserve Separability. (Interestingly, (a normative version of) the *regret theory* proposed by Loomes and Sugden (1982) does this, while also doing away with transitivity.) But I take it that Stochastic Dominance in full strength is far more intuitively plausible than Separability for Options.

4.1 Separability only of finite options

One weakening of Separability for Options is considered by Russell (2023, §4): that which only applies to options X and Y that are concatenated with a *finitely-supported* option Z (an option Z with only finitely many different possible outcomes, or perhaps finitely many possible values).

> *Separability for Finitely-Supported Options*: For any options X and Y, and any option Z with finite support, $X \succcurlyeq Y$ if and only if $X \oplus Z \succcurlyeq Y \oplus Z$.

As Russell observes, Separability for Finitely-Supported Options still suffices to give us much the same Egyptology argument as we saw in Section 2: given that X is better than Y, then it implies that $X \oplus B$ is better than $Y \oplus B$, since B has only two possible outcomes. But, unlike the stronger principle, Separability for Finitely-Supported Options avoids the problem we saw in Section 3: even if H and T are equally good, it does not imply that $H \oplus H$ and $T \oplus H$ are equally good; after all, H is an option with infinitely many possible values.

But there are several problems with this version of Separability, at least if we do not also accept a stronger version. The first is that, as Russell notes, it is unclear why this principle would be justified but Separability for Options would not. It seems that any justification we might offer for this principle would also justify the stronger principle and, with it, rule out Fanaticism.

The second, related, problem is that this principle does not capture what we want from a Separability principle: we want to avoid Egyptology problems. We don't want our verdicts in present-day decisions to be sensitive to the details of events causally isolated from and unaffected by those decisions. But this finite version of Separability does not accomplish this. It does not rule out that our present-day verdicts could depend on whether the queen Nefertiti once played the St Petersburg game (and whether we know how that game turned out). Even that implication would be absurd.

The third, closely related, problem is that the principle does not seem to rule out such sensitivity *in practice.* In practice, we will (or at least should) often be uncertain about distant events to such an extent that those events could take on infinitely many different possible values. Most simply, perhaps an epistemically rational agent must assign non-zero probability to Nefertiti once playing the St Petersburg game. Alternatively, even if an agent could pin down the value of all events in ancient Egypt to being between 0 and 1,000 (on whatever cardinal scale they are using) with certainty, there are still infinitely many possible values between those bounds—if every such value is possible, then Separability for Finitely-Supported Options still wouldn't rule out their present-day decision-making being sensitive to their uncertainty about ancient Egypt. And even beyond ancient Egypt, suggestive arguments have been made elsewhere that epistemically rational agents should assign infinitely-supported probability distributions to the moral value of distant regions of the universe (see Tarsney, n.d., §6.1) and to the value of distant civilisations (see Wilkinson, n.d.b,

§2). In all, if Separability for Finitely-Supported Options is true but Separability for Options is not, then in practice our present-day decision-making may *very often* be sensitive to (our uncertainty about) distant, causally isolated events. And, intuitively, this is absurd.

4.2 Separability only of independent options

I propose a different weakening of Separability for Options, one that encounters fewer problems.

> *Separability for Independent Options*: For any options X and Y, and any options Z for which the values of outcomes *are probabilistically independent of both* X and Y, $X \succcurlyeq Y$ if and only if $X \oplus Z \succcurlyeq Y \oplus Z$.

By (the values of the outcomes of) X and Z being *probabilistically independent* I mean that, even if you knew the exact outcome of X, that would never offer you any information about how the events of Z would play out—the probability of Z giving a particular outcome conditional on the outcome of X is precisely the same as Z's *unconditional* probability of giving that outcome. The values resulting from each are not at all correlated (nor anti-correlated), just as the effects of your present-day actions will typically be uncorrelated with what happened in ancient Egypt.

For an example of options that *aren't* probabilistically independent, consider those from the Egyptology argument in Section 2.

	p	ε	$1-p-\varepsilon$
B	v	0	0
X	V	0	0
Y	0	v	0

You may notice that B is not probabilistically independent of either of the two options. (B resulting in value v is perfectly correlated with X resulting in value V, and anti-correlated with Y resulting in value v.) So, Separability for Independent Options falls silent on whether $X \oplus B$ is better than $Y \oplus B$. This blocks the argument we saw earlier—we are no longer led to Fanaticism.

But a similar Egyptology argument can still be made, for which Separability for Independent Options works just fine. That argument, as given below, is adapted from Wilkinson (2022b, §6.a).

Consider O_{safe} and O_{risky}, as defined above (for any ε and any $V > \frac{v}{\varepsilon}$), as *complementary cumulative probability distributions.* For any given value, a cumulative probability is simply a function that outputs that option's probability of resulting in an outcome with at least that much value.

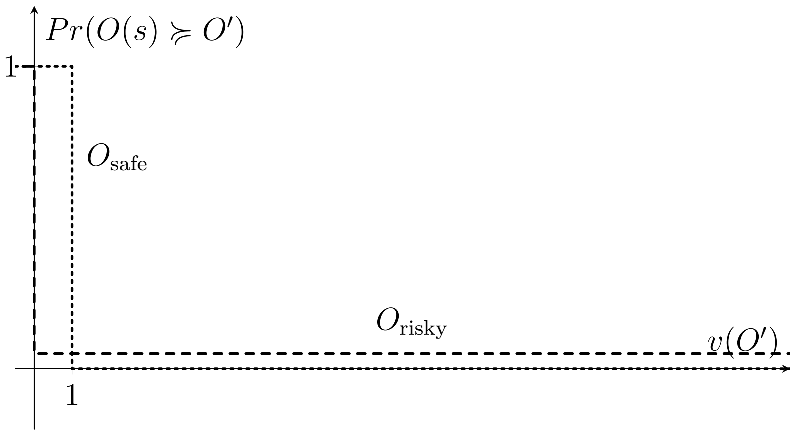

Figure 1: Complementary cumulative probability distributions for O_{risky} and O_{safe}

We can suppose that, here and now, an agent faces a decision between O_{safe} and O_{risky}, and that each represents the value of the entirety of the outcome. If Fanaticism is false, then O_{risky} is not better than O_{safe}. But we might suppose instead that the world were slightly different, in that different events had occurred in ancient Egypt (or at least there were a different probability distribution over those events). Let the hypothetical risky option of how those events would then differ be given by B, as illustrated below. (See note for the precise details of B.)[15]

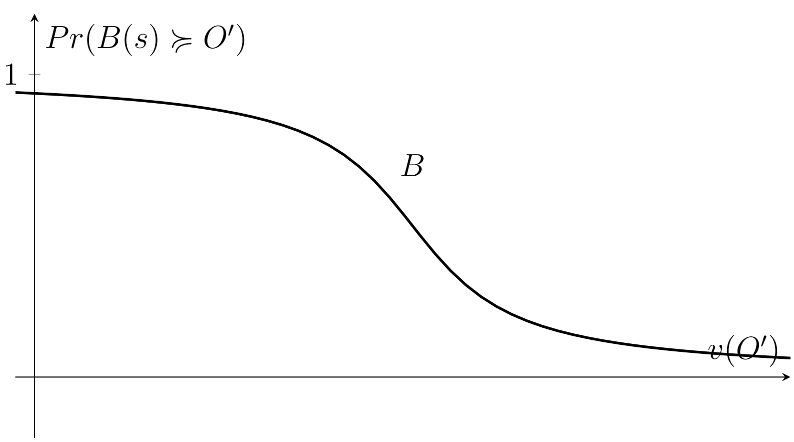

Figure 2: An example of background uncertainty B—a Cauchy distribution

Let the option B also be probabilistically independent of O_{safe} and O_{risky}. Then, if we concatenate each with B, we obtain the options illustrated below.

[15] B can be any Laplace or Cauchy distribution with 'scale factor' at least $\frac{v}{\varepsilon}$. It follows from Theorem 1 of Tarsney (n.d., §5.2) that any such distribution will give us the desired result below: that $O_{\text{risky}} \oplus B$ is strictly better than $O_{\text{safe}} \oplus B$ according to Stochastic Dominance. For more detailed explanations of this mathematical phenomenon, see Tarsney (n.d., §5.1) and Wilkinson (2022b, pp. 470-474).

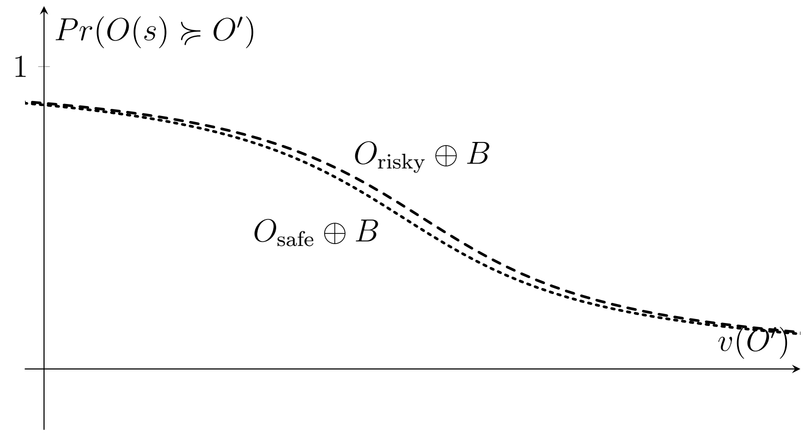

Figure 3: Complementary cumulative probability graphs for $O_{\text{risky}} \oplus B$ and $O_{\text{safe}} \oplus B$

As can be seen here, the cumulative distribution for $O_{\text{risky}} \oplus B$ is always at least as high as, and sometimes higher than, that for $O_{\text{safe}} \oplus B$. In other words, for any possible outcome, the former is at least as likely as the latter to result in an outcome at least that good. And, for some outcomes, $O_{\text{risky}} \oplus B$ is strictly *more* likely to result in an outcome that good. If this relationship sounds familiar, that's because it's the condition for strict betterness that Stochastic Dominance gives us. According to Stochastic Dominance, $O_{\text{risky}} \oplus B$ must be strictly better than $O_{\text{safe}} \oplus B$.

And recall that Separability for Independent Options applies for O_{safe}, O_{risky}, B, and the relevant concatenations. If $O_{\text{risky}} \oplus B$ is better than $O_{\text{safe}} \oplus B$, then so too must O_{risky} be better than O_{safe}. And this is the very definition of Fanaticism. Fanaticism must hold, at least if we accept Stochastic Dominance and Separability for Independent Options. Thus, we have our Egyptology argument.[16]

Separability for Independent Options also allows us to avoid the problem from Section 3. Recall the options involved in the problem case discussed there: H, T, and the concatenations $H \oplus H$ and $T \oplus H$.

	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...
H	2	4	8	...	0	0	0	...
T	0	0	0	...	2	4	8	...

Astute readers may notice that H is not probabilistically independent of H! Of course, the values that result from H are perfectly correlated with the values that result from H. And the values that result from T are anti-correlated. So, Separability for Independent Options does not require our verdict for H versus T to match our verdict for $H \oplus H$ versus $H \oplus T$. The resulting contradiction then

[16]Indeed, an analogous argument can be run if we replace O_{risky} and O_{safe} with any options X and Y, respectively, such that X has greater expected moral value than Y. Given this, Stochastic Dominance and Separability for Independent Options together imply almost all of the verdicts of expected value theory, although that is not my focus here. More on this below.

doesn't arise, nor can we construct an analogous such contradiction. As far as I can tell, Separability for Independent Options is perfectly compatible with Stochastic Dominance and Fanaticism.

5 Justifying Separability for Independent Options but not Separability for Options

But why accept Separability for Independent Options? It seems to face one of the same problems as the other weakening: that it is not obvious why Separability for Independent Options would be justified but the stronger principle of Separability for Options would not. You might think that *any* independent justification we might offer for this principle would also justify the stronger principle and, with it, rule out Fanaticism. But there are some such justifications, which I address below.

5.1 Stochastic Dominance and transitivity

The first such justification comes from the arguments already given above.

From the Egyptology argument of Section 2, we know the following: if $\succcurlyeq$ is transitive, then Separability for Options and Stochastic Dominance jointly imply Fanaticism. And from Russell's (2023) argument, as presented in Section 3 above, we know that: Separability for Options and Stochastic Dominance jointly imply that Fanaticism is false. Given both results, the relevant lesson is not that Fanaticism is not in fact implied by these principles. Rather, the lesson is that, if $\succcurlyeq$ is transitive, then Separability for Options and Stochastic Dominance imply a contradiction—the principles are jointly inconsistent.[17]

Here, then, is one compelling justification for the weaker principle of Separability for Independent Options, but not for the stronger Separability for Options. Considerations of Egyptology are intuitively very compelling: it seems absurd that the correct verdicts in our decisions could rely on distant, unaffected, unrelated events. But even more compelling are Stochastic Dominance and the condition that instrumental moral betterness is a transitive relation. We would like our theory to satisfy our intuitions about Egyptology in as broad a range of cases as possible, but not at the expense of the latter conditions. Separability for Independent Options does so to some extent, while Separability for Options does not. That, by itself, may be sufficient justification for one but not for the other.

[17] See Russell 2023, thm 2.

5.2 Risk sensitivity

A second justification of Separability for Independent Options, which ultimately does not succeed, comes from intuitions about how we should respond to options with different levels of risk.

Consider again the options H and T that caused problems for Separability for Options earlier.

	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...
H	2	4	8	...	0	0	0	...
T	0	0	0	...	2	4	8	...

To concatenate each option with H, and have the resulting options remain equally good, requires that it not matter that $H \oplus H$ is in some sense *riskier* than $T \oplus H$—in $H \oplus H$ we have doubled down and increased our bet on the coin landing heads, while in $T \oplus H$ we have hedged our bet against the coin landing the other way. If we compare these two options to finite analogues H^* and T^*, in which the St Petersburg game is replaced with a finite payoff, then the difference in risk is all the clearer.

	$\frac{1}{2}$	$\frac{1}{2}$
H^*	1	0
T^*	0	1

	$\frac{1}{2}$	$\frac{1}{2}$
$H^* \oplus H^*$	2	0
$T^* \oplus H^*$	1	1

In this case, it may seem intuitive that the cases are evaluatively quite different. In the first, both options are equally risky—both H^* and T^* have equal variance in the value that results. In the second, one option is riskier—although $T^* \oplus H^*$ results in value 1 with certainty, $H^* \oplus H^*$ risks value 0. Note also that, in cases like the second, Stochastic Dominance alone does not say how to compare our options—the case is no longer clear-cut. And, so, it may be less intuitive that a principle of Separability would apply here; certainly less than when the option Z that we concatenate with X and Y is independent of them both, and so does not involve doubling down on a particular bet. In effect, intuition may support Separability for Independent Options more strongly than Separability for Options.

But there is a serious problem with this purported justification. The Egyptology argument given by the weaker Separability for Independent Options principle (in Section 4.2 above) doesn't just give us Fanaticism; it turns out to give us full risk neutrality. In the comparison of $O_{\text{safe}} \oplus B$ and $O_{\text{risky}} \oplus B$ above, we could replace O_{safe} and O_{risky} with *any* pair of options of which one has greater expected value; there will exist *some* option B such that concatenating them with B makes the option with greater expected value stochastically dominate the other (see Tarsney, n.d., §5.2). So, we do not only have the result that Separability for Independent Options and Stochastic Dominance together imply Fanaticism; together, they imply almost all of the verdicts of expected value theory.[18]

[18]It turns out that they do not imply all of expected value theory's verdicts of indifference—if two options X and Y

So, this justification of Separability for Independent Options is an unstable one. It relies on intuitions of risk sensitivity being borne out. And yet, if Separability for Independent Options and Stochastic Dominance hold, then almost all such intuitions must be denied. *Only* in cases like $H \oplus H$ versus $T \oplus H$, where the bet on which we double down is a generalised St Petersburg game, are those intuitions respected. And to deny Separability for Options, but not Separability for Independent Options, on the basis of such intuitions *in only those cases* seems extremely dubious.

5.3 A sure-thing principle

Here is a more promising justification of Separability for Independent Options but not Separability for Options (albeit a justification that still, ultimately, fails). Separability for Independent Options follows, with minimal assumptions, from the Sure Thing Principle, while Separability for Options does not.

Consider the standard Sure Thing Principle (or STP for short) for infinitely-supported options, given below.[19] Note that, here, for any option X and event E, $X_{|E}$ denotes the option obtained from X by conditioning on E. (Equivalently, for any outcome O, $Pr(X_{|E} \succcurlyeq O) = P(X \succcurlyeq O|E)$).

> *Sure Thing Principle*: For any options X and Y and any (jointly exhaustive, pairwise disjoint) set $\mathcal{E}$ of events, each of which has non-zero probability (or, at least, non-zero probability density), if $X_{|E} \succcurlyeq Y_{|E}$ for every $E \in \mathcal{E}$, then $X \succcurlyeq Y$.
>
> If, as well, there is some $E \in \mathcal{E}$ such that $X_{|E} \succ Y_{|E}$, then $X \succ Y$.

Less formally, the Sure Thing Principle tells us the following. Take any options X and Y. And take any question you might have about the world, e.g., "Who will win the next US presidential election?". If, for *every* possible answer to that question—for every possible election victor—the option X would then be at least as good as Y then, overall, X is at least as good as Y. And if there is some possible electoral victor (who has non-zero probability of winning) who, if she were to win, would result in option X being strictly better than Y, then X is strictly better than Y overall. In a sense, with respect to the election result, it would be a 'sure thing' that X will be better than Y. The STP affirms that, if so, then X is indeed better than Y overall.

Notably, the STP implies Separability for Independent Options, at least with an additional assumption. For any options X and Y and any independent option Z, how Z turns out is just the sort of event that the STP refers to. We can let the set $\mathcal{E}$ be the set of all possible outcomes E of Z.

have equal expected value, often there will not be Bs such that both $X \oplus B \succcurlyeq Y \oplus B$ and $X \oplus B \preccurlyeq Y \oplus B$ according to Stochastic Dominance (cf. the exceptions to Theorems 1-4 in Wilkinson, 2022a) .

[19] This version of the principle is taken from Russell and Isaacs (2021, §2). It is stronger than the standard version of the STP in that it allows $\mathcal{E}$ to be infinite.

Then the options $(X \oplus Z)_{|E}$ and $(Y \oplus Z)_{|E}$ (i.e., the options X and Y conditioned on Z resulting in outcome E) are simply $X \oplus E$ and $Y \oplus E$. In terms of value, these are just the same as options X and Y, but with some constant value (the value of E) added onto them. Intuitively, we might think that such background additions to the value of options—in effect, such changes to the zero point from which our options' payoffs deviate—cannot make any difference to our comparison of those options (see Wilkinson, 2022b, p. 467). Assuming this, $X \oplus E \succcurlyeq Y \oplus E$ if and only if $X \succcurlyeq Y$. And, since this holds for all events $E \in \mathcal{E}$, the STP implies that $X \succcurlyeq Y$ if and only if $X \oplus Z \succcurlyeq Y \oplus Z$. Thus, we have Separability for Independent Options.

But, conveniently, the STP *doesn't* imply the stronger principle of Separability for Options. To illustrate, consider again the options from Section 2, in which each option X and Y is concatenated with the option B, which isn't independent of either of them.

	p	q	$1-p-q$
B	v	0	0
$X \oplus B$	$V+v$	0	0
$Y \oplus B$	v	v	0

As above, we might treat the different possible outcomes of B as events and conditionalise on each of them. But, unlike above, those conditional options don't all agree: conditionalise on B having the outcome of value v, and $(X \oplus B)_{|E}$ results in strictly more value than $(Y \oplus B)_{|E}$; but conditionalise on B having the outcome of value 0 and we have the reverse verdict, since $(Y \oplus B)_{|E}$ stochastically dominates $(X \oplus B)_{|E}$. Thus, when B is probabilistically dependent on X and/or Y, the STP need not tell us anything about how the concatenated options compare—we need not accept Separability for Options.

But this justification of Separability for Independent Options has a serious problem, given that Separability for Independent Options leads us to Fanaticism. The STP, in the form given above, is *incompatible* with Fanaticism (in conjunction with Stochastic Dominance)—endorsing the STP doesn't just give us an argument for Fanaticism; it gives us yet another argument for a contradiction. To see why, consider two generalised St Petersburg games, each run independently of one another.[20] They have the same probabilities of the same outcomes; those outcomes just aren't correlated with one another at all.

	$\frac{1}{2} \times \frac{1}{2}$	$\frac{1}{2} \times \frac{1}{4}$	$\frac{1}{2} \times \frac{1}{8}$	...	$\frac{1}{4} \times \frac{1}{2}$	$\frac{1}{4} \times \frac{1}{4}$	$\frac{1}{4} \times \frac{1}{8}$	...	$\frac{1}{8} \times \frac{1}{2}$	$\frac{1}{8} \times \frac{1}{4}$	$\frac{1}{8} \times \frac{1}{8}$	...
St P_1	2	2	2	...	4	4	4	...	8	8	8	...
St P_2	2	4	8	...	2	4	8	...	2	4	8	...

If Fanaticism holds, then we can construct two such generalised St Petersburg games, and they

[20]The result noted here, and the following argument, come from Russell and Isaacs (2021).

will each be better than any sure outcome with finite value (as earlier, from Beckstead and Thomas, 2023, §4). (They may not have exactly the values listed above, but we can just consider the traditional St Petersburg game for simplicity.) If Stochastic Dominance holds, then they are equally good—they are the same options, after all, just run independently of one another. But consider the STP: we can let our set of events simply be the set of outcomes of St P_1; we can conditionalise both options on each such outcome, e.g., St P_1 having the outcome of value 2; and, conditional on that outcome, St P_2 is better, since it is simply a St Petersburg game compared to the value 2; likewise for the other outcomes of St P_1, all of them merely finite and so not as good as a St Petersburg game; conditional on each such event, St P_2 is better; so, the STP then implies that St P_2 is strictly better than St P_1. We have a contradiction!

So, if we accept Stochastic Dominance and we accept Fanaticism (perhaps due to accepting Separability for Independent Options) then we cannot accept the STP. As justifications go, it is an incoherent one.

5.4 An alternative sure-thing principle

The above does not mean that Separability for Independent Options cannot be independently justified by similar means. Note that, much like for Separability for Options, we do not need a principle quite so strong as the STP; we can settle for something weaker (and more plausible).

Recall the way that the STP implied Separability for Independent Options. We can take any options X and Y and any independent option Z, and consider $X \oplus Z$ and $Y \oplus Z$. If we condition those two concatenated options on the event of Z resulting in some event E, we don't just obtain any old pair of options. We obtain the options X and Y with some constant value (the value of E) added to each of their outcomes.

For instance, in place of X, Y, and Z, we can take the options O_{risky}, O_{safe}, and B from earlier. We can carve up $O_{\text{risky}} \oplus B$ and $O_{\text{safe}} \oplus B$ by the events of B resulting in outcomes of value $v_1, v_2, ..., v_i$, and so on. (Call B's probabilities of such outcomes $p_1, p_2, ..., p_i, ...$, respectively.)[21] Carved up in this way, the two concatenated options look like this:

	$\varepsilon \cdot p_1$	$(1-\varepsilon)p_1$	$\varepsilon \cdot p_2$	$(1-\varepsilon)p_2$	...	$\varepsilon \cdot p_i$	$(1-\varepsilon)p_i$	...
$O_{\text{risky}} \oplus B$	$V + v_1$	v_1	$V + v_2$	v_2	...	$V + v_i$	v_i	...
$O_{\text{safe}} \oplus B$	$v + v_1$	$v + v_1$	$v + v_2$	$v + v_2$	...	$v + v_i$	$v + v_i$	...

Conditioned on each of these events—on B resulting in an outcome of each value—we don't just have any old pairs of options $(O_{\text{risky}} \oplus B)_{|E}$ and $(O_{\text{safe}} \oplus B)_{|E}$. Instead, we have the *same* pair options

[21] These p_is may be either probabilities themselves or probability densities, depending on the distribution of the particular option B. The following discussion runs much the same either way.

O_{risky} and O_{safe} over and over again, merely shifted in their outcomes' values by different amounts $(v_1, v_2, ..., v_i, ...)$. It is not just that we might reach agreeing verdicts for each of these pairs (that, in some sense, the two options are *sure* to be ranked a particular way), but that we are effectively comparing the *same* pair over and over again.

For Separability to apply in cases like this, we do not need the full strength of the STP. Instead, we can make do with the weaker (and less problematic) *Sure, Same-But-Shifted Thing Principle* (or SSTP for short).

> *Sure, Same-But-Shifted Thing Principle*: For any options X and Y, if, for every event E of a (jointly exhaustive, pairwise disjoint) set of events $\mathcal{E}$, each of which have non-zero probability (or, at least, non-zero probability density), $X_{|E} \succcurlyeq Y_{|E}$ *and* $X_{|E}$ and $Y_{|E}$ are the *same* pair of options, up to constant shifts in the value of outcomes, then $X \succcurlyeq Y$.
>
> If, as well, $X_{|E} \succ Y_{|E}$ for all such E, then $X \succ Y$.

The SSTP suffices to justify Separability for Independent Options, as illustrated above. And it doesn't justify Separability for Options, by just the same reasoning as for STP. But, crucially, the SSTP *doesn't* conflict with Fanaticism and Stochastic Dominance in the way that the STP did—in the case considered in the previous subsection, of the two generalised St Petersburg games, the SSTP is silent (nor are there analogous cases where SSTP generates such conflict).

But why accept the SSTP but not also the STP? As for Separability itself, is there a justification for the weaker principle that does not serve just as well for the stronger one? Recall that the STP tells us that, if we can condition some options X and Y on any event E in some set $\mathcal{E}$ and always get the same verdict, then we can follow the corresponding verdict for X and Y overall. The SSTP doesn't merely require that conditioning on each E gives us the same *verdicts*; it must give us the *very same options* (in some sense) to compare. The SSTP is thus motivated by an even stronger intuition than that behind the STP: not that we can ignore the differences between events so long as they each lead to the same verdict; but that we can ignore the differences between events when there are not evaluatively relevant differences. And, indeed, there does seem to be no evaluatively relevant difference between two pairs of options where one pair has its outcomes shifted up or down in value by a constant amount—the *differences* in value between their possible outcomes remain the same. And for the purposes of evaluating moral betterness, we are more fundamentally concerned with these differences than with the values themselves—*betterness* is more fundamental than absolute *goodness*, or so you might think. You may disagree with this but, nonetheless, this thought can justify the SSTP but not, by itself, the stronger STP. Thus, we have our justification for one principle that does not also justify the other.

So, we have an intuitive general principle, which is weaker than the Sure Thing Principle but nonetheless has independent justification not shared by it. And this weaker principle itself offers independent justification for Separability for Independent Options, but not the stronger Separability for Options. We have independent reason to accept Separability for Independent Options but not the stronger Separability for Options. And, with it, we then have independent reason to accept the premises of a valid Egyptology argument for Fanaticism.

6 Conclusion

In the existing literature on Fanaticism, perhaps the most *prima facie* compelling arguments for it are Egyptology arguments: that denying Fanaticism leads us to situations where our comparisons of options depend on events in distant parts of the world that are unaffected by our choice between those options (at least if we accept Stochastic Dominance). Intuitively, for our comparisons to be sensitive to such irrelevant events is absurd. What occurred on the banks of the Nile thousands of years ago, for instance, is surely irrelevant to how I should compare the options before me today.

But it has been claimed that those arguments do not succeed—that the principle of Separability on which they rely is itself incompatible with Fanaticism and Stochastic Dominance. Is this true? Do all such arguments indeed fail?

No, it turns out that such an argument can still go through. As shown above, we need only adopt a weaker version of Separability. With it, we face no conflict with Stochastic Dominance or Fanaticism, and we can run much the same Egyptology argument.

But is there any reason to accept that weaker version without also accepting the stronger, problematic version of Separability? Yes, there are several. One is that Stochastic Dominance and transitivity are independently plausible, and they together rule out the stronger version. Another is that the weaker version (and not the stronger version) is implied by a plausible weakening of the Sure Thing Principle, the Same-But-Shifted Thing Principle: the principle that, if two options give the *same* pair of options (up to constant shifts of value) when conditioned on each of a set of jointly exhaustive events, *and* one of those conditional options is always better than the other, then that option is better overall.

Thus, we still have a Separability principle that can serve as the basis for an Egyptology argument in favour of Fanaticism. And we have independent reason to endorse that principle but not a stronger one. In effect, such an Egyptology argument can be salvaged—perhaps the most compelling argument for Fanaticism remains compelling.

References

Beckstead, N., 2013. *On the Overwhelming Importance of Shaping the Far Future*. Ph.D. thesis, Rutgers University Graduate School, New Brunswick. (cited on page 7)

Beckstead, N. and Thomas, T., 2023. A paradox for tiny probabilities and enormous values. *Noûs*, (2023). (cited on pages 6, 7, and 17)

Blackorby, C.; Bossert, W.; and Donaldson, D., 2005. *Population Issues in Social-Choice Theory, Welfare Economics and Ethics*. Cambridge University Press, New York. (cited on page 4)

Broome, J., 2004. *Weighing Lives*. Blackwell. (cited on page 6)

Buchak, L., 2013. *Risk and Rationality*. Oxford University Press, Oxford. (cited on page 1)

Evershed, J., n.d. Super fanaticism. Unpublished manuscript. (cited on page 2)

Goodsell, Z., 2021. A St Petersburg paradox for risky welfare aggregation. *Analysis*, 81, 3 (2021), 420–6. (cited on page 7)

Huemer, M., 2008. In defence of repugnance. *Mind*, 117, 468 (2008), 899–933. (cited on page 6)

Loomes, G. and Sugden, R., 1982. Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92, 368 (1982), 805–824. (cited on page 8)

McMahan, J., 1981. Problems of population theory. *Ethics*, 92, 1 (1981), pp. 96–127. (cited on page 3)

Nebel, J., 2018. The good, the bad, and the transitivity of *better than*. *Noûs*, 52, 4 (2018), pp. 874–99. (cited on page 6)

Parfit, D., 1984. *Reasons and Persons*. Oxford University Press, Oxford. (cited on page 3)

Pivato, M., 2014. Additive representation of separable preferences over infinite products. *Theory and Decision*, 77, 1 (2014), pp. 31–83. (cited on page 4)

Russell, J. S., 2023. On two arguments for fanaticism. *Noûs*, (2023). (cited on pages 2, 4, 7, 9, and 13)

Russell, J. S. and Isaacs, Y., 2021. Infinite prospects. *Philosophy and Phenomenological Research*, 103, 1 (2021), 178–198. (cited on pages 15 and 16)

Seidenfeld, T.; Schervish, M. J.; and Kadane, J. B., 2009. Preference for equivalent random variables: A price for unbounded utilities. *Journal of Mathematical Economics*, 45, 5-6 (2009), 329–340. (cited on page 7)

Tarsney, C., n.d. Exceeding expectations: Stochastic dominance as a general decision theory. Unpublished manuscript. Available at https://globalprioritiesinstitute.org/christian-tarsney-exceeding-expectations-stochastic-dominance-as-a-general-decision-theory/. (cited on pages 9, 11, and 14)

Thomas, T., 2022. Separability and population ethics. *The Oxford Handbook of Population Ethics*, (2022), 271–296. (cited on page 4)

Wilkinson, H., 2022a. Can risk aversion survive the long run? *Philosophical Quarterly*, (2022). (cited on page 15)

Wilkinson, H., 2022b. In defence of fanaticism. *Ethics*, 132, 2 (2022), p. 445–77. (cited on pages 7, 10, 11, and 16)

Wilkinson, H., n.d.a. Can the long term matter if extinction does not? Unpublished manuscript. (cited on page 4)

Wilkinson, H., n.d.b. The unexpected value of the future. Unpublished manuscript. Available at https://globalprioritiesinstitute.org/the-unexpected-value-of-the-future-hayden-wilkinson-global-priorities-institute-university-of-oxford/. (cited on page 9)