Can an evidentialist be risk-averse?[*]

Hayden Wilkinson

Last updated: November, 2022

Comments welcome: hayden.wilkinson@philosophy.ox.ac.uk

Abstract

Two key questions of normative decision theory are: 1) whether the probabilities relevant to decision theory are *evidential* or *causal*; and 2) whether agents should be risk-*neutral*, and so maximise the expected value of the outcome, or instead risk-*averse* (or otherwise sensitive to risk). These questions are typically thought to be independent—that our answer to one bears little on our answer to the other. But there is a surprising argument that they are not. In this paper, I show that evidential decision theory implies risk neutrality, at least in moral decision-making and at least on plausible empirical assumptions. Take any risk-aversion-accommodating decision theory, apply it using the probabilities prescribed by evidential decision theory, and every verdict of moral betterness you reach will match those of expected value theory.

Keywords: *evidential decision theory; risk aversion; risk sensitivity; expected value theory; expected utility theory; risk-weighted expected utility theory.*

[*] I am grateful to Adam Bales, Ina Jäntgen, and Timothy L. Williamson for exhaustive feedback on earlier drafts of this paper.

1 Introduction

When making moral decisions about aiding others, you might think it appropriate to be *risk-averse*. For instance, suppose you face a decision between: rescuing one person from drowning for sure; and spinning a roulette wheel—if the roulette wheel lands on 0 (one of 37 possibilities), you thereby rescue 37 people (with similarly valuable lives) from drowning and, if it lands on any of the other 36 numbers, you rescue no one.[1] On the face of it, it seems plausible that it is better (instrumentally) to rescue the one person for sure, rather than to risk saving no one at all.

In this paper, I will present a novel argument against risk aversion[2] in moral cases, and in favour of *risk neutrality*: that one risky option is instrumentally better than another if and only if it results in a greater expected sum of moral value. The argument starts from a surprising place, usually thought to have no bearing on issues of risk aversion or risk neutrality. It starts from the claim that the probabilities used to compare options are those given by your evidence, *including* the evidence provided by that option being chosen; that we should accept *evidential decision theory* (EDT).

To illustrate what EDT asks of us, consider the much-discussed Newcomb's Problem.

> **Newcomb's Problem**
>
> Before you are two boxes, one opaque and one transparent. You can see that the transparent box contains $1,000. You cannot see into the opaque box, but you know that it contains either $0 or $1,000,000. You can either take the opaque box, or take both boxes. *But* the contents of the opaque box have been decided by a highly reliable predictor (perhaps with a long record of predicting the choices of others who have faced the same problem). If she predicted that you would take both boxes, it contains $0. If she predicted that you would just take just take the opaque box, it contains $1,000,000.

Which is better: to take one or to take both? EDT tells us that taking the one is better. Why? You know that the predictor is highly reliable. So, if you take just the opaque box, you thereby

[1] Assume that, if not rescued, each of those people is guaranteed to drown. So, the possibility of saving no one (in the second option) does not arise because no rescue attempt is necessary; it arises because your rescue attempt would be unsuccessful. Without this assumption, it turns out that what risk aversion recommends is under-determined—see Greaves et al. (n.d.).

[2] More generally, the argument has force against any form of risk *sensitivity* (any deviation from risk neutrality). But, in the moral case, risk aversion seems more plausible than risk seeking (*cf.* Buchak, 2019), so I will focus here on risk aversion.

obtain strong evidence that the $1,000,000 is contained within—we can suppose the probability that it does, conditional on taking just one box, is very close to 1. But, if you take both boxes, you thereby obtain strong evidence that the opaque box is empty—the probability that it contains $0, conditional on taking both boxes, is again close to 1. Using these probabilities, taking both boxes will almost certainly win you only $1,000, while taking just the opaque box will almost certainly win you $1,000,000. The latter then seems far better.

Alternatively, you might endorse *causal decision theory* (CDT): that the probabilities used to compare options are how probable it is that choosing that option will *cause* each outcome; evidence provided by the choice itself is ignored (see Joyce, 1999, p. 4). In Newcomb's Problem, to the causal decision theorist, the probability of the opaque box containing $1,000,000 is the same for both options—making either choice has no *causal* influence on what the predictor puts in the box, so the probability cannot change between options. Using these probabilities, taking both boxes is guaranteed to turn out at least as well as taking just the one. So, the option of taking both must be better than that of taking one.

On the face of it, whether we endorse EDT's or CDT's core claims seems to be independent of whether we should endorse risk aversion or risk neutrality.[3] At its core, the question of EDT or CDT is a question about what notion of probability we take as normatively relevant. And this doesn't seem to bear on how we should respond to said probabilities, and so whether it is appropriate to be risk-averse. Any theory of risk aversion could perhaps be applied to either notion of probability.[4]

But this turns out not to be true. As I will argue, if EDT is true then in practice so too is risk neutrality, at least for moral decision-making. And so we have a novel argument for risk neutrality; that or, if you think risk neutrality deeply implausible, a novel argument against EDT.

[3] One technical, and not very compelling, reason to think otherwise is this: CDT is typically axiomatised in the framework of Savage (1954), while EDT is typically axiomatised in the framework of Jeffrey (1965); and, where risk aversion is accommodated in normative decision theory, it is often done so in the basic framework of Savage (see, e.g., Buchak, 2013, p. 88 & p. 91). But there is no necessary connection between EDT and the Jeffrey framework—EDT can be expressed in Savage's framework (e.g., Spencer and Wells, 2019, pp. 28-9), and CDT can be expressed in Jeffrey's framework (e.g., Edgington, 2011). Nor is there a necessary connection between Jeffrey's framework and risk neutrality—theories accommodating risk aversion can be formulated in that framework too (see Stefánsson and Bradley, 2019).

[4] Risk neutrality is often assumed without argument in existing discussions of EDT and CDT. But I take it that this is typically not for any principled reason, but instead in the interests of brevity (as indicated by, e.g., Williamson, 2021: Footnote 27), or simply due to a lack of imagination.

2 The basic argument

The basic argument resembles so-called 'long-run arguments' for maximising expected value[5], but with a twist. It goes like this.

Consider any (moral) decision problem you like. For any normative decision theory that exhibits or accommodates risk aversion, suppose that that theory says some option O_{safe} is the best of those available, while risk neutrality says that some other option O_{risky} is—that O_{risky} results in the greatest moral value, in expectation.

Consider also a modified version of that decision problem, in which each option is replaced by n repetitions of the same option. O_{safe} is replaced with $O_{\text{safe}_{\times n}}$: the prospect you would get if you ran n independent trials of the prospect associated with O_{safe} and summed the value of every trial. Likewise for O_{risky} (which is replaced with $O_{\text{risky}_{\times n}}$) and each other option. In this modified decision problem, risk *neutrality* will still say that $O_{\text{risky}_{\times n}}$ is better than any other option, including $O_{\text{safe}_{\times n}}$. But risk *aversion*, it turns out, may waver—it may not recommend $O_{\text{safe}_{\times n}}$. Why not? The (Strong) Law of Large Numbers tells us that, as n approaches infinity, the average value of each of the n repetitions of each option is guaranteed to approach the option's expected value.[6] And O_{risky} has greater expected value than O_{safe}. So, as n gets larger, the probability that $O_{\text{risky}_{\times n}}$ turns out better than $O_{\text{safe}_{\times n}}$ gets arbitrarily close to 1. No matter how risk-averse (or risk-seeking) we are (with some provisos, as detailed in the next section), we will reach the verdict that $O_{\text{risky}_{\times n}}$ is better than $O_{\text{safe}_{\times n}}$, so long as n is large enough. Risk neutrality and risk aversion will come to agree.

Suppose an agent only ever faces decisions of this kind—decisions not among standalone, low-stakes options, but among very many n independent repetitions of such options. For any given decision, if n is large enough, the best option available to such an agent will always be the one with the highest expected value (per repetition), no matter whether we are risk-averse or risk-neutral. Risk effectively does not enter into the picture—whichever option has the highest expected value (per repetition) is guaranteed to *actually* result in the highest value. Risk-averse or not, they will prefer that option. So, for such an agent, for practical purposes, risk neutrality would hold.

But by EDT, when making moral decisions, each of us is such an agent! The world contains

[5] Such long-run arguments can be found in Thoma (2019), Stefánsson (2020), Zhao (2021), Hájek (2021, §.2), and Baron (2000, p. 244), among others.

[6] See Feller (1968, p. 258).

vastly many decision-makers similar to us, facing decisions over options with similar moral stakes and similar probabilities. The choices made by those other decision-makers are correlated with our own choices. So, in any given decision, what we choose gives us evidence about what other decision-makers will choose. That evidence may be fairly weak—those decision-makers may be only weakly correlated with us, but there will still be significant correlation. But given how vastly many of them there are, even a small portion of them is still many, many decision-makers. So, when choosing what causal effect to have on the world, we still learn something about how many, many others will affect the world through their own decisions elsewhere. For instance, suppose that you are choosing between options with the value of their causal effects given by prospects O_{safe} and O_{risky}, respectively. (Call these the *causal prospects* of your options.) Choose causal prospect O_{safe} over O_{risky} and you learn that many, many others choose their own causal prospect O_{safe}; choose causal prospect O_{risky} and you learn that those others choose the risky option.

You might wonder if your choice really does provide you with new evidence about those other agents' behaviours. As has been suggested elsewhere in response to so-called 'Medical Newcomb Problems' (e.g., by Ahmed, 2014, §4.3), you might think that your evidence is gained from your *desire* or *impulse* to choose one option or the other, and that your *choice* provides no further evidence. But, whether or not this is true in Medical Newcomb Problems, it is incorrect in the current setting. When considering other agents similar to yourself, they often act similarly to you *not only* because they have similar desires or impulses; they also *respond* to those desires and impulses in a manner similar to you. You do indeed gain evidence about their behaviour from your own desires and impulses prior to making a choice, but you also gain evidence of how they respond to those desires from how you yourself respond—from what you end up choosing. And, so, the evidentialist must take into account that further evidence about how their correlated agents will act.

On top of this, they must also *care* about the effects of those correlated agents' choices, at least when comparing options *morally*. Unlike in prudential decision situations, moral decision-making does not allow us to ignore such distant effects on the world—on any plausible moral theory, good or bad effects occurring far away in space or in time count for just as much as those nearby.[7]Here, I will also assume that such effects count *additively*—that the correct theory of moral betterness is additive (as are totalism, total prioritarianism, and critical-level views), such that any given outcome's value is simply the total additive aggregate of the moral value

[7] For compelling defences of this claim see, for example, Sidgwick (1907) and (Parfit, 1984, §121 & Appendix F).

(considered impartially) of every event within it. [8][9] [10] Given any such additive view, when we consider options with causal prospects O_{safe} and O_{risky}, EDT asks that we compare those options as $O_{\text{safe}_{\times n}}$ $O_{\text{risky}_{\times n}}$.[11] And no matter how risk-averse or risk-seeking we are, as long as n is large enough, the better option will be $O_{\text{risky}_{\times n}}$. Thus, for comparisons of moral betterness, it may seem that EDT implies risk neutrality in practice.

This implication is even more acute if we accept that there are *infinitely* many such other decision-makers spread throughout the universe, facing decisions over options with arbitrarily similar stakes and probabilities. Various leading theories of cosmology predict that our universe will be infinite in its spatiotemporal volume, and contain infinitely many identical instances of every given local physical phenomenon.[12] So, for any particular moral decision we might face, *infinitely many* other decision-makers will face identical decisions with morally equivalent outcomes. Choose one option and we not only learn that *many* more other decision-makers choose the same way; we learn that *infinitely many* more others choose the same way (than would otherwise). By EDT, our overall options are not between, say, O_{safe} repeated n times and O_{risky} repeated n times, for some finite n, but between each option repeated infinitely many times. This means that, plausibly, the average value to result from each option overall is *precisely* the option's expected value. The option with the highest expected value, repeated infinitely many times, is *guaranteed* to actually turn out better than any alternative. (For a more detailed discussion of this, see Section 4 below.) And any form of risk aversion must say that a better outcome with certainty is better than a worse outcome with certainty. Thus, again, in practice EDT implies risk neutrality in moral decision-making.

But does this argument succeed? If we consider specific accounts of risk aversion, do they

[8] This need not be because outcomes themselves have *objective* value (as Ahmed and Spencer, 2020, argue is incompatible with EDT). It may simply be that the prospect in which a given outcome has probability 1 has *subjective* value equal to the total aggregate of value in the outcome.

[9] Similar results would arise for non-additive theories of moral betterness, e.g., averageism and egalitarianism. The main reason is that, in worlds already containing a great deal of moral value, averageism and egalitarianism give verdicts similar to additive theories when evaluating outcomes—see Tarsney and Thomas (n.d.).

[10] Without further assumptions about the correct deontic moral theory, this tells us little of what we *ought* to do. But there are at least some decisions situations in which it is highly plausible that we ought to do whatever is (instrumentally) best—in particular, situations where we aid others, without violating any deontic constraints or neglecting special relationships or demonstrating bad character, and must decide exactly how to do so (see Pummer, 2016).

[11] This is subject to two major complications. The first: there will be other sources of uncertainty in the world, beyond how decisions like this one are made and how they turn out. We can ignore this complication because, if anything, it simply gives us further reason to embrace risk neutrality (see Tarsney, n.d.; Wilkinson, n.d.a, §4.3). The second complication: there may also be agents in the universe whose choices are anti-correlated with your own; I address this in §5.1 below.

[12] This is implied by standard versions of the widely accepted *flat-lambda* model of cosmology (see Wald, 1983; de Simone et al., 2010; Carroll, 2017, for discussion). It is also implied by the *inflationary view* (see Guth, 2007; Garriga and Vilenkin, 2001), and is a likely scenario under Smolin's (1992) model of cosmological natural selection.

really recommend O_{risky} repeated n times over O_{safe} repeated n times, for large enough n? Is the number of correlated decision-makers indeed large enough to give us that large enough n? Does the argument extend neatly to infinite numbers of correlated decision-makers, despite the various complications introduced by infinities? And does the argument break down if we consider the existence of *anti*-correlated decision-makers; or if we consider our uncertainty about the number of other decision-makers; or if we consider the possibility that not just their *choices* depend on ours, but also that how those choices turn out might depend on how ours turn out? In the remainder of this paper, I will consider each of these issues in turn.

3 Long-run arguments

In general, when and how do such long-run arguments in favour of risk neutrality succeed? They succeed when, given a decision problem where each option is simply n repetitions of some simpler option and given a decision theory that exhibits or accommodates risk aversion, that theory implies the same verdicts as risk neutrality. But whether such a theory agrees with risk neutrality will depend on which theory it is, exactly.

3.1 Theories that accommodate risk aversion

One such theory is *expected utility theory*. It says that the best of a set of option(s) is whichever bring(s) about the greatest expectation of *utility*. This may sound like the very definition of risk neutrality. But here I mean *utility* in a decision-theoretic sense, meaning something quite different from (moral) *value*—it also represents one's attitude to risk. Depending on how much better it is, instrumentally, to obtain higher probabilities of good outcomes, the utility of an outcome can be *any* increasing function of its value—one outcome will still have greater value than another if and only if it has greater utility, but *how much greater* its value is can vary.

Expected utility theory allows us to accommodate risk aversion (with respect to moral value) by allowing the agent's utility function to be non-linear. In particular, it allows us to accommodate risk *aversion* by allowing the agent's utility function to be *concave*: the greater the (moral) value of an outcome, the less difference to its utility is made by some additional value. Such a utility function looks like this:

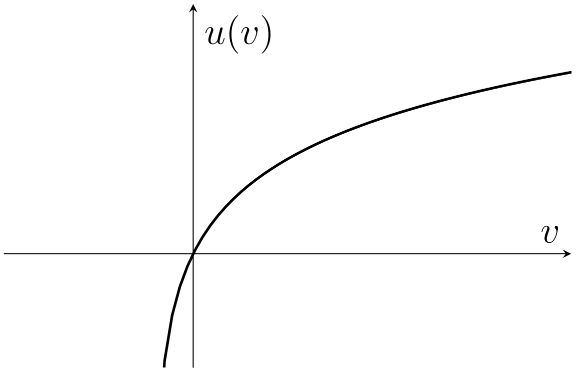

Figure 1: A utility function $u(v)$ that is concave with respect to (cardinal, moral) value v.

If the agent making the decision has a concave utility function like this, expected utility theory will deliver the risk-averse verdict that we expect in cases similar to that from earlier. Suppose you have the option of rescuing one person from drowning for sure (A) and that of spinning a roulette wheel which, if it lands on 0, will result in 38 identical such people being rescued (B). In terms of moral value, we might assume that your options look like this.

A: value 1 with probability 1.

B: value 38 with probability $^1/_{37}$; value 0 with probability $^{36}/_{37}$.

In terms of moral *value*, the difference between zero lives saved and one life saved is much less than the difference between one life saved and 38 lives saved. But the difference in *utility* can be different—with a sufficiently concave utility function, there will be a greater difference between zero lives and one life than between one life and 38. In terms of expected *utility* (or EU for short), saving one life for sure can be better.

Another such theory that accommodates risk aversion is Lara Buchak's *risk-weighted expected utility theory*.[13] According to this theory, the best of a set of option(s) is whichever bring(s) about the greatest *risk-weighted* expected utility (or REU). The REU of a given option O is calculated as follows.

$$REU(O,r) = u_1 + \sum_{j=2}^{n} (u_j - u_{j-1}) r \Big(P(O \geq u_j) \Big)$$

Here, $u_1, u_2, u_3, ..., u_n$ are the utilities (according to some utility function, as above) of the possible outcomes of O, ordered from lowest to highest. The REU formula asks us to start with the utility

[13] The novelty of this theory compared to expected utility theory can be seen in its ability to accommodate the intuitively rational preferences described by Allais, which expected utility theory rules out Buchak (see 2013, pp. 31-4).

of the worst possible outcome, given by u_1, and sum the amounts by which O could exceed u_1 (and u_2, and so on). But we weight those amounts, not by the *probability* that we exceed the previous u_{j-1} by that amount, but by some function *of* that probability.

This risk function, r, represents the agent's risk attitude. By definition, r must be real-valued, non-decreasing, and satisfy $r(0) = 0$ and $r(1) = 1$.[14] And it allows us to accommodate general risk aversion by allowing the agent's r function to be *convex*, and to take values of $r(p)$ less than p for all probabilities between 0 and 1. Such an r function looks like this:

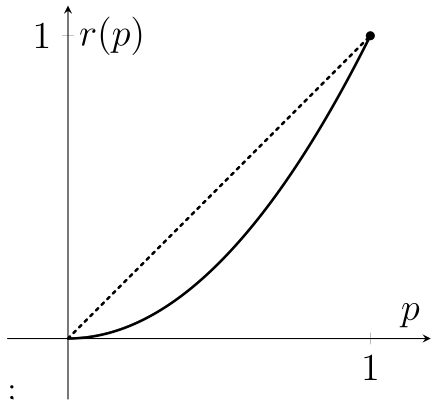

Figure 2: A risk function $r(p)$ that is convex with respect to probability and returns values less than p for all $0 < p < 1$.

If we adopt a convex risk function like this, REU theory can deliver the risk-averse verdict in the case of A versus B.[15] A sufficiently convex r function transforms the probability of saving 38 lives in B to $r(1/37)$ less than $1/37$. The REU of B would be less than 1. So, under REU theory, rescuing one person for sure may be better.

3.2 Long-run arguments under EU and REU theory

Under either of these theories, we can pose long-run arguments. Both EU theory and REU theory often come to agree with expected value theory when comparing repeated prospects such as $O_{a\times n}$ and $O_{b\times n}$, so long as n is large enough (as shown in detail in Wilkinson, n.d.a).

For instance, consider again the options of saving a life for sure (A) and saving 38 with probability $1/37$ (B). Expected value theory says that B is better than A. And, as above, both EU theory and REU theory can say that A is better. For instance, on EU theory, if we let $u(v) = \sqrt{v}$, B has

[14] Buchak, *Risk and Rationality*, 49.

[15] Note that this holds *even if* the utility function $u(v)$ is linear.

an EU of less than 1, and so less than that of A. And on REU theory, if we let $r(p) = p^2$ (and $u(v)$ be linear), B has an REU of less than 1 too. But consider A and B when they are each repeated, say, 1,000 times. On those same utility and risk functions, $B_{\times 10}$ suddenly has (slightly) higher EU and REU than $A_{\times 10}$. So, both theories then agree that the option with the higher expected value is better.

Such agreement emerges more generally. Take *any* causal prospects O_a and O_b such that O_a has strictly higher expected moral value than O_b. EU theory will agree that $O_{a\times n}$ is better than $O_{b\times n}$ for any sufficiently large finite n, *even if* it denies that O_a is better than O_b, so long as the utility function $u(v)$ satisfies certain conditions (more on these below). So too, REU theory will agree that $O_{a\times n}$ is better than $O_{b\times n}$ for any sufficiently large finite n, *even if* it denies that O_a is better than O_b, so long as the utility function and risk function $r(p)$ satisfy some further conditions.[16] So, as long as those conditions are met, for an agent who only faces decisions among repeated prospects like O_a and O_b (and as long as they are repeated sufficiently many times), REU theory and EU theory will never have *anything* to say that isn't already said by expected value theory. Even if we accept such a theory that accommodates risk aversion, there will effectively be no risk aversion at all in the verdicts we reach. In effect, we must be risk-neutral.

But there are limits to this conclusion. We cannot take just *any* pair of prospects each repeated some given number of times, and have these theories (with *any* utility function and *any* risk function) agree in their verdicts. Often, they will still disagree.

Indeed, EU theory will continue to disagree with expected value theory about how to compare some repeated prospects $O_{a\times n}$ and $O_{b\times n}$ whenever they disagree about how to compare O_a and O_b *and* any of:

1. the number n of repetitions simply isn't large enough—in general, the greater the difference in the EU of O_a and O_b, and the more concave the utility function $u(v)$, the more repetitions are needed;

2. O_a and O_b have precisely *equal* expected value—if O_a has lower EU, so too will $O_{a\times n}$ have lower EU than $O_{b\times n}$ for *any finite number* n of repetitions; or

3. the utility function $u(v)$ is non-linear but has what is called constant *absolute risk aversion*[17],

[16] See Theorems 1 and 2 and their proofs in Wilkinson (n.d.a) and the similar theorem sketched by Buchak (2013, p. 238).

[17] Absolute risk aversion is a technical notion from economics, describing the agent's attitude to risk at each level

as do, e.g., *exponential* utility functions (those of the form $u(v) = 1 - a^{-v}$, for some positive a).

And REU theory will continue to disagree with expected value theory about how to compare some repeated prospects $O_{a\times n}$ and $O_{b\times n}$ whenever they disagree about how to compare O_a and O_b *and* any of:

1. the number n of repetitions simply isn't large enough—in general, the greater the difference in the REU of O_a and O_b, and the more convex the risk function $u(v)$, the more repetitions are needed;
2. O_a and O_b have precisely *equal* expected value—if O_a has lower REU, so too will $O_{a\times n}$ have lower REU than $O_{b\times n}$ for *any finite number* n of repetitions; or
3. the risk function $r(p)$ is discontinuous at (or near enough to) $p = 1$.

If the prospects under comparison and the agent's risk attitudes meet any of those six conditions, disagreement will remain. EU theory or REU theory will continue to disagree with expected value theory about the best course of action.

What do these limitations mean for the argument sketched in Section 2, that EDT implies risk neutrality? We will indeed be led to risk-neutral verdicts in many cases (and for many possible utility and risk functions) if our every choice gives us evidence about finitely many $n - 1$ choices by other agents in the world. But not in *all* cases; not even close. There will be vastly many cases where EU theory and REU theory will still say something different to expected value theory. They will continue to offer different verdicts whenever the repeated prospects have precisely equal expected value. And even in other cases, there will continue to be *some* risk attitudes—some utility and risk functions—such that they continue to offer different verdicts, even for some very large numbers n of repetitions. Indeed, for *any* such n—for any number of correlated agents that may inhabit the universe—there are some utility and risk functions such that EU and REU theory will still frequently offer verdicts different to expected value theory.

So, it seems that the argument fails. Even if we accept EDT and consider the evidence given by our choices about how very many other agents will act, it seems that risk aversion can persist in practice. (But, as I show below, appearances may be deceiving.)

of value. It is measured by the expression $\frac{u''(v)}{u'(v)}$, where u' and u'' are, respectively, u's first and second derivatives with respect to v.

4 Long-run arguments and infinities

Despite the above, there is a version that does succeed—the version in which there are *infinitely* many other agents about whose choices we obtain evidence through our own choice.

As noated earlier, our evidence from physics makes it likely that there are infinitely many other decision-makers spread throughout the universe. It also makes it likely that infinitely many of them are near-identical to you or I in their dispositions, and that infinitely many of them face moral decisions among causal prospects that closely resemble our own. If so, we do not face situations in which our options are each repeated some finite n times, but instead situations in which they are repeated *infinitely* many times. (It turns out that the argument does not require that we be *certain* of the existence of infinitely many such agents; even a *non-zero probability* of infinitely many such agents will suffice—see §5.2 below.)

How should we apply long-run results to situations where the number n of repetitions becomes infinite? It is not immediately clear. We know that the average EU and average REU of any repeated prospect $O_{\times n}$ will approach its expected value as n *tends* to infinity.[18] But it isn't obvious that such a limit result tells us anything about what happens when n *reaches* infinity.

One reason it might not tell us anything is that, in general, we cannot take such limit results literally as n reaches infinity. To see why, consider a situation in which we start with a square 1 metre in height and 1 metre in width. And we know that it will be modified: that it will have its width doubled and height halved, both n times. For any finite number n, the square's area will remain $\frac{2^n}{2^n} = 1$. And the limit of its area as n tends to infinity will be 1. But what is its area when n reaches infinity—when it has infinite width and a height of 0? The 'shape' we are left with is simply a straight line. We might claim that its area is 0, or that it is undefined, but either way it would be a mistake to say that the area remains 1 for $n = \infty$. Given cases like this, we should be cautious about over-interpreting limit results when n actually is infinite.

Another reason the limit result may not tell us anything here is that the options under consideration, and at least some of their possible outcomes, will have infinite value. And standard additive theories of moral betterness have difficulty comparing such options and outcomes. For any two options O_a and O_b with positive expected value, there is probability 1 that both $O_{a\times\infty}$ and $O_{b\times\infty}$ will result in a total value that is (positively) infinite. This is guaranteed to occur no matter how vastly different their expected values nor how different the spread of their distribu-

[18] This can be proven along much the same lines as Theorems 1-4 in Wilkinson (n.d.a).

tions. And, by standard cardinal arithmetic, we cannot say that any such infinite total value is greater, so it seems that neither option can be better. But fortunately, as it turns out, there are ways to extend additive theories to say something sensible in the infinite context.

We can think of situations with infinitely many potential differences in value, and/or infinitely many beneficiaries, as follows. Each such beneficiary occupies some physical position in space and time. Let the set of such positions be $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, ...\}$. In any given decision situation, our available options will result in some outcomes, or worlds, say, W_a and W_b. Each outcome is associated with a value function (V_a and V_b) which maps each position in $\mathcal{X}$ to a real number, representing the value at that position.

$$\begin{array}{rccccc} & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \cdots \\ W_a: & V_a(\mathbf{x}_1) & V_a(\mathbf{x}_2) & V_a(\mathbf{x}_3) & V_a(\mathbf{x}_4) & \cdots \\ W_b: & V_b(\mathbf{x}_1) & V_b(\mathbf{x}_2) & V_b(\mathbf{x}_3) & V_b(\mathbf{x}_4) & \cdots \end{array}$$

And we have the machinery to compare such options, even when the set $\mathcal{L}$ contains infinitely many locations. Drawing on recent discussions (e.g., Bostrom, 2011; Wilkinson, 2021a,b; Askell, 2019), we have plausible extensions of additive moral theories to the setting of infinite populations and infinite total value. *Some* such extensions will still fail to compare the outcomes of $O_{a\times\infty}$ and $O_{b\times\infty}$ (and, indeed, fail to compare *any* realistic outcomes—see Wilkinson, n.d.b; Jonsson and Peterson, 2020). But others *can* compare them. And, handily, all such successful proposals satisfy a common principle: *Expansionism*.

> *Expansionism* (from Wilkinson, 2021b, pp. 1935-6): For any outcomes W_a and W_b, W_a is better than W_b if, for every spacetime point $\mathbf{x}_0$ there is some r' such that, for all $r > r'$,
>
> $$\sum_{\mathbf{x} \in E(r, \mathbf{x}_0)} V_a(\mathbf{x}) - V_b(\mathbf{x}) > 0$$
>
> where $E(r, \mathbf{x}_0)$ represents the region of all spacetime points within (spatiotemporal) distance r of $\mathbf{p}$.
>
> And if, for all $\mathbf{x}_0$, there is some r' such that for all $r > r'$ the sum is precisely 0, then W_a and W_b are equally good.

Put simply, to compare two outcomes, Expansionism tells us to: start at any physical position $\mathbf{x}_0$; for each outcome, sum the amount of moral value over all points within distance r of $\mathbf{x}_0$; and let r get larger and larger. If there is some r at which one outcome has more moral value within that region than the other does, and it continues to have more for every larger r as well (and for every other starting point $\mathbf{x}_0$), then the former is the better outcome.

And Expansionism suffices to rank infinitely many repetitions of a prospect with high expected value over infinitely many repetitions of one with low expected value. Suppose you face a decision between causal prospects O_a and O_b, where O_a has strictly greater expected value. And there are infinitely many agents in the world who are correlated with you and are facing similar decisions, in the same positions in each outcome. Let W_a be the outcome that results from them all choosing causal prospect O_a and W_b likewise for O_b. For simplicity (and thanks to the additive nature of Expansionism), we can ignore all extraneous sources of value in the world—we can let the values at each position be only the values arising from the choices of those correlated agents. And we can treat these values at each position, $V_a(\mathbf{x}_i)$ or $V_b(\mathbf{x}_i)$, as independent random variables—as the risky values generated by standalone prospects O_a and O_b respectively.

Applying Expansionism to this pair of outcomes, take any starting point $\mathbf{x}_0$. For any large enough region $E(r, \mathbf{x}_0)$ around it, and any large number n, there will be n such positions within the region hosting one of those correlated decision-makers. The sum $V_a(\mathbf{x}) - V_b(\mathbf{x})$ over all such $\mathbf{x}$ will be given by the random variable $O_{a\times n} - O_{b\times n}$. And, since the causal prospect O_a has strictly greater expected value than O_b, the (Strong) Law of Large Numbers tells us that this difference diverges to positive infinity —with probability 1, there will be some n such that $O_{a\times n} - O_{b\times n}$ is positive and likewise for all greater n. So Expansionism will confirm that W_a will be strictly better than W_b (with probability 1).

This gives us a stronger long-run argument than we had in the finite setting. In that setting, we could only say that the *prospect* $O_{a\times n}$ would eventually be better than the prospect $O_{b\times n}$, on particular views of comparing prospects. But, in the infinite setting, we can say with certainty that the *outcome* of $O_{a\times\infty}$ is better than that of $O_{b\times\infty}$. It does not matter whether we adopt a theory that accommodates risk aversion, nor *how* risk-averse or risk-seeking we might be—one prospect is guaranteed to have a better outcome than the other. By *any* theory that satisfies even an extremely weak principle of dominance[19], including EU theory and REU theory, $O_{a\times\infty}$ must be the better prospect. Thus, any such theory will agree with the risk-neutral verdict.[20]

[19] This may be a principle of *statewise* dominance, *stochastic* dominance, or *eventwise* dominance (also known as the Sure Thing Principle)—see [citations] for details and discussion of such principles.

[20] In the infinite context, it is more difficult to formalise a risk-neutral (or indeed any plausible) decision theory. But it is possible—see Wilkinson (2022b) and Arntzenius (2014) for demonstrations, each of which uphold the risk-

Consider also the situation where the causal prospects O_b and O_c have *equal* expected value. Then, Expansionism says that neither outcome is better than the other. Since the random variable $O_b - O_c$ has expected value 0, the sum $O_{b\times n} - O_{c\times n}$ forms a random walk that is *recurrent*: for any n there is, with probability 1, some greater n' such that the difference $O_{b\times n} - O_{c\times n}$ will return to 0.[21] Likewise, it will continue rising above and below 0, if $O_b - O_c$ is not just a constant 0. So there will be no n (and so no r') to satisfy the conditions of Expansionism. By Expansionism alone, neither is better than the other. Nor by stronger existing proposals (e.g. Wilkinson, 2021b, p. 1946), it turns out, do we obtain the verdict that either is strictly better. But we cannot say that they are equally good either: the two outcomes, and corresponding options, remain incomparable (with some minor exceptions). This is different from what the risk-neutral verdict would be, but it is not *inconsistent* with that verdict; it is simply weaker. It remains true that, whenever Expansionism (or its extensions) gives us a verdict, it will agree with the risk-neutral verdict.

5 Complications

The long-run argument described above faces various further complications and objections that I have so far neglected. It turns out that none of them undermine the core conclusion: that, in practice, EDT implies risk neutrality in moral decision-making. (For simplicity, I deal with each complication separately but the core takeaways carry over to the case where all three complications are present.)

5.1 Partially correlated and anti-correlated decision-makers

The first such complication is that not all of the agents inhabiting the world are either perfectly correlated or entirely uncorrelated with you. Some are partially correlated: how you choose only gives you weak evidence about how they will choose. And others are *anti*-correlated: whenever you choose one option, you gain evidence that they *won't* choose that option.

The first group, partially correlated agents, pose no great problem for the argument. Suppose that you face a decision between any two options: one with causal prospect O_a, which has greater expected moral value; and one with causal prospect O_b, which has greater EU and REU (on any given utility and risk functions). And consider first the situation in which the world

neutral verdicts I point to here.

[21] For a proof of this fact, see Chung and Fuchs (1951).

contains only *finitely* many other agents. Let some fraction $m < 1$ of the n agents in the world be only partially correlated with you[22]. We can assume, for simplicity, that all of these only partially correlated agents are correlated with you to the same degree. Then consider how the overall prospect looks if you choose causal prospect O_a. EDT will have you evaluate that option as the prospect you obtain by summing the payoffs of the following two prospects: the prospect of how the perfectly correlated agents' choices turn out ($O_{a \times n(1-m)}$); and of how the only partially correlated agents' choices turn out (a mixture of O_a with probability $p > 0.5$ and O_b with probability $1 - p$, repeated nm times). Since O_a has greater expected value than O_b, both of these two prospects (and the prospect obtained by summing their payoffs) will be better than the corresponding prospects for O_b, according to either EU theory or REU theory, as long as n is large enough (with the same provisos as given in §3.2 above).[23] So, again, both theories will typically agree with risk neutrality about how to compare the two options available to you.

What if there are infinitely many agents? Again, let some fraction $m < 1$ of the n agents in the world be only partially correlated with you.[24] If we apply Expansionism then the prospect of value in any region with radius r will match the prospect given in the previous paragraph—the prospect obtained by summing the payoffs of how the $n(1 - m)$ perfectly correlated agents' choices turn out and how the nm partially correlated agents' decisions turn out. As r approaches infinity, the number of agents n will approach infinity too, and for the same reasons as above the outcome of choosing O_a will come to be better than that of choosing O_b, with probability 1. And since the outcome of one prospect is guaranteed to be better than that of the other, by any theory that satisfies even a weak dominance principle, the prospect associated with choosing O_a will have a strictly better prospect. Thus, the presence of partially correlated agents does not threaten the result.

The second group, anti-correlated agents, seem to pose a greater problem. It may be that there are vastly many agents out there with capacities similar to yours, but with preferences or perhaps risk attitudes opposed to yours. Every time you choose to save a life rather than let someone die, you may gain evidence that such agents will let others die when given the opportunity. Or, more applicable to the present case, every time you choose an option that maximises the expected moral value of the causal prospect over another that some risk-averse

[22] m must be less than one because there will be at least one agent who is perfectly correlated with you: yourself.

[23] This can easily be proven via the same method as Theorems 3 and 4 in Wilkinson (n.d.a). See also Feller (1968, p. 259).

[24] When I say that some fraction m of an infinite set of agents have some property I mean that, on any natural ordering in which we might enumerate those agents (e.g., orderings given by their positions in spacetime), the fraction of them with that property tends to m as the enumeration approaches the full, infinite set (*cf* the notion of *density* from Wilkinson, 2021b, p. 1930).

agent would favour, you may gain evidence that such agents would choose the other option instead. If there happened to be enough such anti-correlated agents out there, the argument given above would weigh *against* maximising expected value of the causal prospect you bring about, *even if* you accept expected value theory!

Are there more such correlated agents or anti-correlated agents? Cosmology doesn't tell us. But, nonetheless, it seems that you should think that there are many more agents in the world correlated with you than not there are agents anti-correlated with you (to any given degree, including perfect correlation). After all, you yourself are an agent with such and such preferences and reasoning processes, and presumably you know many other agents who have fairly similar preferences and reasoning processes. Those other agents may often choose differently to you, but not in such a way that you choosing an option O gives you positive evidence that they will not choose O. It seems an appropriate use of induction to infer that, for any agent out there in the world whom you haven't yet encountered, they are more likely to be (at least weakly) correlated with you than they are to be anti-correlated with you. If not, you would be in a surprisingly special epistemic position, and it seems rational to assume that you are not so positioned.

And, if we accept that there are fewer agents in the world anti-correlated with you than there are agents correlated with you (to the same degree), the argument still goes through. Suppose again that you face a decision between causal prospects O_a with greater expected moral value and O_b with greater EU and REU (on some given utility and risk functions). And suppose for now that there are only finitely many n other agents in the world (about whom you gain evidence through your choice), some portion $m < 0.5$ of whom are anti-correlated with you. (For simplicity, assume that all such agents are either perfectly correlated or anti-correlated with you; this won't affect the result, for the reasons given above.) Then EDT will have you evaluate the option with causal prospect O_a as the prospect obtained by summing the values resulting from $O_{a\times n(1-m)}$ and $O_{b\times nm}$; and it will have you evaluate the option with causal prospect O_b as the corresponding sum of $O_{b\times n(1-m)}$ and $O_{a\times nm}$. And the (Strong) Law of Large Numbers still grants that the former will have higher EU and REU than the latter, as long as n is large enough (with the same provisos as given in §3.2 above).[25] Thus, anti-correlated agents pose little problem, as long as they are in the minority.

By reasoning analogous to that above, this result can be extended to the setting of *infinitely* many correlated and anti-correlated agents. As long as there is a greater fraction of correlated agents than of anti-correlated ones, Expansionism again says that the outcome of choosing O_a

[25] This follows from Theorems 3 and 4 in Wilkinson (n.d.a).

is guaranteed to be better than that of choosing O_b. Again, anti-correlated agents post little problem, as long as they are in the minority.

5.2 Uncertainty about number of decision-makers

Another complication, specifically about the infinite version of the long-run argument, is that you may be *uncertain* that there are indeed infinitely many agents in the world correlated with you. After all, our empirical evidence about what the distant future (and distant space) hold is weak. We should assign at least some non-zero probability to the world containing only finitely many agents, as well as some non-zero probability to it containing infinitely many agents.

To see how this affects the argument, again suppose that you face a decision between: causal prospect O_a, which has greater expected moral value; and causal prospect O_b, which has greater EU and REU (on some given utility and risk functions). And assume for simplicity that all other agents in the world are perfectly correlated with you (and that there are no sources of value in the world other than the causal consequences of their decisions that mirror your own). Given your uncertainty about whether there are infinitely many such agents, EDT would have you evaluate the option with causal prospect O_a as: a mixture of $O_{a\times n}$ with probability p and $O_{a\times\infty}$ with probability $1-p$, for some $0 < p < 1$ that reflects the probability of there existing only finitely many agents. Likewise, it would have you evaluate the option with causal prospect O_b as: a mixture of $O_{b\times n}$ with probability p and $O_{b\times\infty}$ with probability $1-p$.

As above, we can treat the two overall prospects here as prospects of value at each of a set of individual positions $\mathcal{L}$ in the world. At the first n such positions $(l_1, l_2, ..., l_n)$, the prospect of value is the same as above: an independent repetition of O_a or O_b. But, for all of the infinitely many other other positions $(l_{n+1}, l_{n+2}, ...)$, there is probability p that they do not exist and, in effect, have value 0. For each such location l_i, we can call its prospect $O_{a,i,p}$ or $O_{b,i,p}$: a mixture of value 0 with probability p, and O_a or O_b otherwise. (Note that the values resulting from each of these will not be independent—if there are in fact only finitely many decision-makers, then they will *all* have value 0.)

$$\begin{array}{rccccccc} & l_1 & l_2 & \cdots & l_n & l_{n+1} & l_{n+2} & \cdots \\ \text{If you choose } O_a: & O_{a,1} & O_{a,2} & \cdots & O_{a,n} & O_{a,n+1,p} & O_{a,n+2,p} & \cdots \\ \text{If you choose } O_b: & O_{b,1} & O_{b,2} & \cdots & O_{b,n} & O_{b,n+1,p} & O_{b,n+2,p} & \cdots \end{array}$$

Again, we can also consider the actual *outcomes* of choosing O_a or O_b. We can treat each of the entries in the array above as a random variable rather than a prospect, and consider whether either overall outcome is guaranteed to be better.

Looking at just the first finitely many n locations, either outcome could turn out better. (But, for large n, the outcome of choosing O_a is far more likely to be better.) But, if the outcome of choosing O_b is better over those n locations, then it is at most *finitely* better—the difference in value between it and the outcome of choosing O_a is at most finite. Then, looking at the remaining, infinitely many locations, there are two possibilities. The first: every location has value 0 in both outcomes. If so, the outcomes can be compared by the values at just the first n locations; either could then be better, but by at most a finite amount. The second possibility: those infinitely many other locations have value given by independent random variables corresponding to O_a and O_b (depending on your choice). And we know from above that a prospect O_a with higher expected value than O_b, if repeated independently over *infinitely* many locations, is guaranteed to be better than O_b repeated over the same locations, according to Expansionism. Indeed, by the same reasoning as above, it is guaranteed to be *infinitely* better—the difference between the value in a region in the outcome $O_{a\times\infty}$ and that in $O_{a\times\infty}$ approaches positive infinity as the region grows larger, with probability 1.

So, in the situation where there is probability p of there being only finitely many correlated decision-makers, neither option is guaranteed to have a better outcome. With some probability less than p, choosing O_b will result in a finitely better outcome. And with probability at least $1 - p$, choosing O_a will result in an infinitely better outcome. The question of which option is better comes down to how a particular theory of risk aversion treats probabilities of infinitely or unboundedly valuable outcomes. And both EU theory and REU theory are compatible with denying that probability $1 - p$ of infinite gain is worth probability p of finite loss. Under EU theory, we can adopt a utility function that has some upper bound—that approaches some finite upper limit (as, e.g., Arrow, 1971, proposes). And under REU theory, we can adopt a discontinuous risk function such that $r(p') = 0$ for some non-zero p' (as Buchak, 2013, p. 61 suggests). But such utility and risk functions would need to reflect rather extreme risk attitudes. They would need to imply that the probability $1 - p$ of infinite gain is outweighed by probability p of some finite gain, where $1 - p$ is the probability that the universe contains infinitely many correlated decision-makers. And I would suggest that this probability should not be all that small! As noted earlier, among the most widely accepted theories of cosmology are several that imply the existence of infinitely many such decision-makers. If we are epistemically rational, I suspect that we must assign quite a high probability, perhaps even more than 0.5, to at least one of them being correct in this implication. And for a decision theory to say that even a probability as high as 0.5 or even 0.1 of infinite gain is not as good as a higher probability of finite gain, that decision theory would need to deliver bizarre verdicts elsewhere.[26]

Given this, it is safe to say that even a *moderate* probability of there existing infinitely many correlated decision-makers in the world will still support the argument that EDT implies risk neutrality in practice, at least if we do not adopt a form of risk aversion that implies bizarre verdicts.

5.3 Dependent outcomes

A final complication is that your choices and their outcomes may not only give you evidence about the choices of others; the *outcomes* of those choices may also give you evidence about the outcomes of others' choices. For instance, suppose you are comparing one option A that causes one life to be saved with certainty, and another option B that has probability 0.5 of saving 3 lives (and probability 0.5 of saving no lives). If you choose B, and it does indeed turn out to save 3 lives (or 0 lives), then that may be evidence that B is generally a good strategy—evidence that, if other agents choose B, they too will end up saving 3 lives.

If you do indeed gain evidence of this then, according to EDT, you cannot evaluate the prospect of choosing B as many *independent* trials of B. The respective trials are instead dependent (but not *perfectly* correlated, since you typically won't gain perfect evidence of how those other trials will turn out). And so the results cited above do not apply, as is needed if we are dealing with finitely many correlated decision-makers. And the Law of Large Numbers does not apply, as is needed if we are dealing with infinitely many such agents. So, the argument that, in practice, EDT leads to risk neutrality seems to fail!

Fortunately, there are variants of the Law of Large Numbers that apply for dependent trials. As proven by Bernstein (1927, ch. 3), the average value of n dependent, identically-distributed random variables $O_{a,1}, O_{a,2}, ..., O_{a,n}, ...$ will converge to the expected value of O_a with probability 1, on two conditions. The first condition: O_a has bounded variance (and so is not a pathological prospect like, e.g., the Pasadena or St Petersburg games). The second condition: we can arrange

[26] Any decision theory that denies that some probability of infinite gain outweighs any finite gain will face a variety of further objections that most will find deeply troubling—see Wilkinson (2022a) and Beckstead and Thomas (n.d.).

those variables such that their correlation with $O_{a,1}$ tends towards 0.[27] And, in the practical circumstances with which we are concerned, it seems highly plausible that both conditions hold. At least for decisions with limited stakes, the variance of the available prospects is guaranteed to be bounded. And, as we consider decision-makers further and further away from us in space and time, it seems that the actual outcomes of our own choices should give us less and less evidence (approaching zero evidence) about the actual outcomes of theirs. And, if so, then all of the above will carry over to the context of dependent outcomes.[28]

6 Conclusion

On the face of it, it may seem that the distinction between evidential and causal decision theories is independent of the distinction between risk-neutral and risk-aversion-accommodating theories. But this is not so. If we focus on moral decisions-making, then there are arguments that evidentialists must compare their options as though they were risk-neutral.

As I have argued, whatever moral decision you face, the world contains many other decision-makers who face similar decisions and whose choices are correlated with yours. Whenever you make a choice, you gain evidence about how those agents choose. And EDT dictates that you take this evidence into account when making your own decision. By EDT, the best choice is that which is associated with the best prospect over how *all* of those decision-makers' choose.

For any given risk attitude, on either EU theory or REU theory, taking these other decision-makers into account leads us closer to risk neutrality. If there are finitely many such decision-makers, then any risk attitude, combined with EDT, will imply many more of the verdicts of risk neutrality than it otherwise would. And, if there are *infinitely* many decision-makers, then any risk attitude, combined with EDT, will imply *all* of the verdicts of risk neutrality. And these implications hold even if, as seems realistic: not all of those decision-makers are perfectly correlated with you; some are *anti*-correlated with you (as long as they are in the minority); you are uncertain of whether there are infinitely many or merely finitely many such decision-makers; and if the actual outcomes that result from their choices are at least somewhat dependent on how your choice turns out.

[27] The exact condition is that their covariance satisfies $\lim_{|i-j|\to\infty} Cov(O_{a,i}, O_{a,j}) = 0$.

[28] For the context of only *finitely* many decision-makers, what I have said here is not sufficient to show that the earlier conclusions still hold. But the results cited above—Theorems 1-4 from Wilkinson (n.d.a)—can be reproduced in the context of dependent outcomes with only slight modifications to their proofs. (Simply replace Kolmogorov's Inequality and Chebysev's Inequality with Bernstein's Inequality, from Bernstein, 1927, ch. 3).

Given this, *even if* you assign a rather low probability to the world containing infinitely many such decision-makers—even in the range of 0.01 or 0.001—EDT will still imply perfect risk neutrality in such moral decisions. Likewise if you take your own choices as extremely weak evidence for how those decision-makers act. EDT still implies risk neutrality in practice.

There are some possible conditions under which this conclusion does not hold. One is if we reject my earlier moral assumption: that the correct theory is additive, such that events far away from us in space and time matter morally no less than events here and now. But I strongly suspect that, if we were to rerun the argument with any plausible theory of moral betterness, we would reach a similar conclusion. To avoid it, I suspect, we would need to deny that such faraway events can count morally for anywhere near as much as nearby events. Then, EDT may still be compatible with risk-averse verdicts. But such a moral position is difficult to defend.[29] Another possible escape from the conclusion is to deny that our choices give us evidence about other agents in the cosmos. You might claim that we only gain evidence about those agents from our desires or impulses, not from our choices themselves. But this response, too, seems hard to justify; it would certainly require a defence stronger than that proffered by evidentialists in response to similar cases (see Ahmed, 2014, §4.3).

If you do accept the argument, then what you take away from it is up to you. If you find risk neutrality absurd, you have an argument against EDT. If you find risk neutrality absurd and EDT highly plausible, you have an argument against certain moral theories, and/or against certain claims about when and how we gain evidence. But, for those committed to EDT and to whom the above escapes do not appeal, you can say goodbye to risk aversion; with EDT, in practical moral decisions, you have no choice but to maximise expected value.

References

AHMED, A., 2014. *Evidence, Decision and Causality*. Cambridge University Press, Cambridge. (cited on pages 4 and 21)

AHMED, A. AND SPENCER, J., 2020. Objective value is always newcombizable. *Mind*, 129, 516 (2020), pp. 1157–92. (cited on page 5)

ARNTZENIUS, F., 2014. Utilitarianism, decision theory and eternity. *Philosophical Perspectives*, 28, 1 (2014), p. 31–58. (cited on page 13)

ARROW, K., 1971. *Essays in the Theory of Risk-Bearing*. Markham, Chicago. (cited on page 18)

[29] For discussion see, for example, Parfit (1984, §121 & Appendix F).

Askell, A., 2019. *Pareto Principles in Infinite Ethics*. Ph.D. thesis, New York University. (cited on page 12)

Baron, J., 2000. *Thinking and Deciding*. Cambridge University Press, Cambridge. (cited on page 3)

Beckstead, N. and Thomas, T., n.d. A paradox for tiny probabilities and enormous values. Unpublished manuscript. (cited on page 19)

Bernstein, S. N., 1927. *Teoriia Veroiatnostei [The Theory of Probabilities]*. Gosudarstvennoe Izdatel'stvo, Moscow. (cited on pages 19 and 20)

Bostrom, N., 2011. Infinite ethics. *Analysis and Metaphysics*, 10 (2011), pp. 9–59. (cited on page 12)

Buchak, L., 2013. *Risk and Rationality*. Oxford University Press, Oxford. (cited on pages 2, 7, 9, and 18)

Buchak, L., 2019. Weighing the risks of climate change. *The Monist*, 102, 1 (2019), pp. 66–83. (cited on page 1)

Carroll, S., 2017. Why Boltzmann brains are bad. *arXiv preprint*, (2017). (cited on page 5)

Chung, K. and Fuchs, W., 1951. On the distribution of values of sums of random variables. *Memoirs of the American Mathematical Society*, 6 (1951), p. 12. (cited on page 14)

de Simone, A.; Guth, A.; Linde, A.; Noorbala, M.; Salem, M.; and Vilenkin, A., 2010. Boltzmann brains and the scale-factor cutoff measure of the multiverse. *Physical Review D*, 82, 6 (2010), 063520. (cited on page 5)

Edgington, D., 2011. Conditionals, causation, and decision. *Analytic Philosophy*, 52, 2 (2011), pp. 75–87. (cited on page 2)

Feller, W., 1968. *An Introduction to Probability Theory and Its Applications (3rd ed.)*. Wiley, New York. (cited on pages 3 and 15)

Garriga, J. and Vilenkin, A., 2001. Many worlds in one. *Physical Review D*, 64, 4 (2001), 043511. (cited on page 5)

Greaves, H.; MacAskill, W.; Mogensen, A.; and Thomas, T., n.d. On the desire to make a difference. (cited on page 1)

Guth, A., 2007. Eternal inflation and its implications. *Journal of Physics A: Mathematical and Theoretical*, 40, 25 (2007), pp. 6811. (cited on page 5)

Hájek, A., 2021. Risky business. *Philosophical Perspectives*, 35, 1 (2021), pp. 189–205. (cited on page 3)

Jeffrey, R., 1965. *The Logic of Decision*. McGraw-Hill, New York. (cited on page 2)

Jonsson, A. and Peterson, M., 2020. Consequentialism in infinite worlds. *Analysis*, 80, 2 (2020), pp. 240–8. (cited on page 12)

Joyce, J., 1999. *Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge. (cited on page 2)

Parfit, D., 1984. *Reasons and Persons*. Oxford University Press, Oxford. (cited on pages 4 and 21)

Pummer, T., 2016. Whether and where to give. *Philosophy & Public Affairs*, 44, 1 (2016), pp. 77–95. (cited on page 5)

Savage, L., 1954. *The Foundations of Statistics*. John Wiley & Sons, New York. (cited on page 2)

Sidgwick, H., 1907. *The Methods of Ethics, 7th edn*. Macmillan, London. (cited on page 4)

Smolin, L., 1992. Did the universe evolve? *Classical and Quantum Gravity*, 9, 1 (1992), 173. (cited on page 5)

Spencer, J. and Wells, I., 2019. Why take both boxes? *Philosophy and Phenomenological Research*, 99, 1 (2019), pp. 27–48. (cited on page 2)

Stefánsson, H. O., 2020. The tragedy of the risk averse. *Erkenntnis*, (2020). (cited on page 3)

Stefánsson, H. O. and Bradley, R., 2019. What is risk aversion? *The British Journal of Philosophy of Science*, 70, 1 (2019), pp. 77–102. (cited on page 2)

Tarsney, C., n.d. Exceeding expectations: Stochastic dominance as a general decision theory. Unpublished manuscript. (cited on page 5)

Tarsney, C. and Thomas, T., n.d. Non-additive axiologies in large worlds. Unpublished manuscript. (cited on page 5)

Thoma, J., 2019. Risk aversion and the long run. *Ethics*, 129, 2 (2019), p. 230–53. (cited on page 3)

Wald, R., 1983. Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant. *Physical Review D*, 28, 8 (1983), pp. 2118. (cited on page 5)

Wilkinson, H., 2021a. *Infinite Aggregation*. PhD dissertation, Australian National University. (cited on page 12)

Wilkinson, H., 2021b. Infinite aggregation: Expanded addition. *Philosophical Studies*, 178, 6 (2021), pp. 1917–49. (cited on pages 12, 14, and 15)

Wilkinson, H., 2022a. In defence of fanaticism. *Ethics*, 132, 2 (2022), p. 445–77. (cited on page 19)

Wilkinson, H., 2022b. Infinite aggregation and risk. *Australasian Journal of Philosophy*, (2022). (cited on page 13)

WILKINSON, H., n.d.a. Can risk aversion survive the long run? Unpublished manuscript. (cited on pages 5, 8, 9, 11, 15, 16, and 20)

WILKINSON, H., n.d.b. Chaos, ad infinitum. Unpublished manuscript. (cited on page 12)

WILLIAMSON, T. L., 2021. Causal decision theory is safe from psychopaths. *Erkenntnis*, 86 (2021), pp. 665–85. (cited on page 2)

ZHAO, M., 2021. Ignore risk; maximize expected moral value. *Noûs*, (2021). (cited on page 3)