The unexpected value of the future

Hayden Wilkinson (Global Priorities Institute, University of Oxford)

Global Priorities Institute | September 2022

GPI Working Paper No . 17-2022

GLOBAL PRIORITIES INSTITUTE

UNIVERSITY OF OXFORD

The unexpected value of the future*

Hayden Wilkinson

Abstract

Various philosophers accept moral views that are *impartial*, *additive*, and *risk-neutral* with respect to betterness. But, if that risk neutrality is spelt out according to expected value theory alone, such views face a dire *reductio ad absurdum*. If the expected sum of value in humanity's future is *undefined*—if, e.g., the probability distribution over possible values of the future resembles the Pasadena game, or a Cauchy distribution—then those views say that no real-world option is *ever* better than any other. And, as I argue, our evidence plausibly supports such a probability distribution. Indeed, it supports a probability distribution that cannot be evaluated even if we extend expected value theory according to one of several extensions proposed in the literature. Must we therefore reject all impartial, additive, risk-neutral moral theories? It turns out that we need not. I provide a potential solution: by adopting a strong enough extension of expected value theory, we can evaluate that problematic distribution and potentially salvage those moral views.

Keywords: *Pasadena game; expected value theory; expected utility theory; longtermism; risk aversion; relative expectation theory; principal value theory.*

*I am grateful to Adam Bales, Jacob Barrett, Harvey Lederman, Andreas Mogensen, Toby Newberry, Jeffrey Russell, Christian Tarsney, Johanna Thoma, Teru Thomas, David Thorstad, Timothy L. Williamson, and two anonymous reviewers for their generous comments on various versions of this paper. For helpful discussion, I also thank Tomi Francis and an audience at the 9th Oxford Workshop on Global Priorities Research. And, for his extensive assistance with the mathematical details in §2.2, I thank Alexander Barry.

1 Introduction

Consider three moral claims, each seemingly plausible and, in conjunction, accepted by various philosophers.[1]

The first is *Impartiality*: that the moral value of a life does not depend intrinsically on when or where it occurs; that, for instance, a human life lived millions of years in the future would be no more or less valuable than an otherwise identical life lived today.[2]

The second claim is that what matters for comparisons of moral betterness is the total sum of such value. Call this *Additivity*, the claim that: an outcome is at least as good as another if and only if the former contains at least as great a total sum of (some cardinal measure of) the value within each individual life.[3]

The third claim is that, when comparing *risky* options, *expected value theory* holds. This common view says that the morally best *prospects* over outcomes are those with the highest expected moral value—if we also accept Additivity, those with the highest probability-weighted sum of total moral value.[4]

But, the conjunction of these three claims has troubling implications. If we face options with certain probability distributions over total moral value, those options' expected values will be *undefined*. We will be unable to say that their expected values are greater than, equal to, or less than any other. Nor, by expected value theory (and no stronger principle for comparing risky options) can we say which of our options are best. The probability distributions that generate this result include well-known troublemakers from decision theory, such as: the Pasadena game (originating in Nover and Hájek, 2004), and the Agnesi game (see Poisson, 1824; Alexander, 2012). But the problem is not merely hypothetical, as those trouble-making cases from decision theory are. As I will argue, these three claims lead to a practical ethical problem. In practice, there is reason to think that the total moral value of the future follows a similarly problematic probability distribution. And, if so, *every* option we might ever choose in practice will have undefined expected value.[5]

[1] Note that these three claims do not imply a consequentialist moral theory. They are each claims about moral *betterness*, rather than about what we *ought* to do. They are perfectly compatible with nonconsequentialist views that recognise some notion of betterness *simpliciter*.

[2] Impartiality is defended by Sidgwick (1907, 414), Ramsey (1928, 541), and Parfit (1984, 486), among others.

[3] For strong independent justification of Additivity, see the arguments of Broome (2004, ch. 18) and Thomas (2022b, §5). Note also that Additivity, as defined here, is compatible with critical-level and prioritarian views of how valuable each individual life is.

[4] For compelling arguments in favour of expected (moral) value theory see, for instance, Harsanyi (1955), Tarsney (n.d.), and Zhao (2021).

[5] Crucially, the problem I will discuss is not simply that our actions may have non-zero probability of resulting in *infinite* value. The probability distributions involved assign *no* probability to outcomes of infinite value. And, so, the problem I discuss arises even if we treat infinitely-valued outcomes as a conceptual impossibility. Likewise, an analogous problem arises if we recognise infinitely-valued outcomes but we replace expected value theory with a decision theory that brackets off infinitely-valued outcomes and compares our options only by their finitely-valued outcomes (as is proposed by Bostrom, 2011, 37-8).

If we have one of these problematic probability distributions over the total value of the future, and we accept Impartiality, Additivity, and expected value theory alone, then we face a dire *reductio ad absurdum*. For every option ever available to us in practice, we cannot evaluate it; we cannot compare it to any other such option, not even to options *identical* to itself. We can *never* say how our options compare in terms of moral betterness.

This implication seems absurd. But it is not immediately clear how we might avoid it, at least without abandoning Impartiality or Additivity—without admitting that the time at which a life is lived *can* matter morally, nor admitting that the ranking of outcomes deviates from that of their total values. If we find both claims compelling, can we hold onto them and extend our comparisons to prospects without slipping into absurdity?

One way we might do so—which I will discuss but not ultimately endorse—is by replacing expected value theory with an alternative theory that exhibits sensitivity to risk (e.g., expected *utility* theory with a non-linear utility function, or a version of *risk-weighted* expected utility theory). With the right profile of risk aversion and risk seeking, such theories may effectively replace prospects like the Pasadena game with better-behaved ones. Given this, we may have a novel argument for risk sensitivity in the moral context: it seems we may need to be risk-sensitive to compare our options *at all* in practice.

In this paper, I seek to determine, in effect, whether risk sensitivity is the *only* way out. If you find Impartiality, Additivity, and the risk neutrality of expected value theory convincing, is there some way to salvage them?[6] Can the above *reductio* be avoided without allowing risk sensitivity, or denying Impartiality or Additivity?

To preserve risk neutrality, it is necessary to *extend* expected value theory to compare troublesome options. The literature already features various proposals for how to do so (e.g., Colyvan, 2008; Easwaran, 2008; Easwaran, 2014a; Meacham, 2019). But, as it turns out, many of these existing proposals fail—I argue that they cannot compare the options we face in practice. Despite this, I describe a theory that may succeed in doing so. With such a theory, we can compare the problematic options I describe. And so we may be able to avoid the *reductio* that expected value theory, Impartiality, and Additivity brought upon us, and do so without endorsing risk sensitivity.

[6]This question bears importantly on recent debates concerning our obligations to future generations. It has been argued that Impartiality, Additivity, and expected value theory (or views that imply them) provide a justification for *Axiological Longtermism*: the view that the best options available to us, at least in many important practical decisions, are those that most increase the *ex ante* moral value of the far future (Greaves and MacAskill, 2021, 3). But, if those three claims bring on absurdity, this justification for Axiological Longtermism is undermined. If so, they do not imply the verdicts needed for longtermism; they imply no practical verdicts at all.

2 Why would the expected value of the future be undefined?

Decision theorists have long recognised prospects that lack well-defined, finite expected values. Some prospects lack such expected values because they feature outcomes with *infinite* value, such as in Pascal's Wager. But I will set aside such prospects in this paper, and assume that outcomes must have only finite value.[7]

But even if we exclude infinitely valuable outcomes, some prospects still lack well-defined expected values. One frequently discussed such prospect is that of the *Pasadena game*.[8]

Pasadena game: (An outcome with) value 2 with probability $1/2$;
value -2 with probability $1/4$;
value $8/3$ with probability $1/8$;
...
value $\frac{2^n}{n}(-1)^{n-1}$ with probability $1/2^n$ for each positive integer n.

What is the game's expected value? If we try to calculate it in the order the outcomes are listed, we obtain the series $1 - 1/2 + 1/3 - 1/4 + ... + \frac{(-1)^{n-1}}{n} + ...$ This series, also known as the alternating harmonic series, fails to be absolutely convergent. If we were to naively add it up in one order or another, so long as we picked the right order, we could obtain *any* total we wanted.[9] So, we cannot say that the game has any particular expected value at all (see Nover and Hájek, 2004)—in this sense, the Pasadena game *defies expectations* (or is *expectation-defying*). And so expected value theory cannot tell us how it compares to any outcome, nor to any other option, nor even to itself. If options were to be compared by expected value theory alone, then the Pasadena game would be no better than, no worse than, nor equally good as *any* other option.

A similar prospect is the *Agnesi game*. Unlike the Pasadena game, it gives a continuous (rather than discrete) probability distribution over possible values. It can result in an outcome of *any* real value v; its probability density over value is given by the following function, also known as the *Witch*

[7] My reasons for setting aside such prospects are threefold. The first: it is independently interesting if we can solve the problems raised by prospects over finitely-valued outcomes alone. The second: you might in fact think that outcomes of infinite value are metaphysically or logically impossible, and so assign them probability zero in practice (cf. Al-Kindi, 1974; Craig, 1979). The third: the problems of infinitely-valued outcomes seem solvable, but in a way that leaves intact the problems of the Pasadena game and its kin (see Wilkinson, 2021, 2022b; Tarsney and Wilkinson, n.d.).

[8] This game is typically presented with payoffs in terms of dollars or (decision-theoretic) utility, in amounts matching those below (e.g., Nover and Hájek, 2004; Easwaran, 2014a; Bartha, 2016). Such versions of the game pose problems for expected dollar maximisers and expected utility maximisers. Here, the game is presented in terms of *moral value* and will pose structurally identical problems for expected *value* maximisers.

[9] Since the series is conditionally convergent, this result follows from the Riemann Rearrangement Theorem.

of Agnesi or (an example of) the Cauchy distribution.[10]

$$p(v) = \frac{1}{\pi(1 + v^2)}$$

On a graph, its distribution looks like this, symmetric about 0.

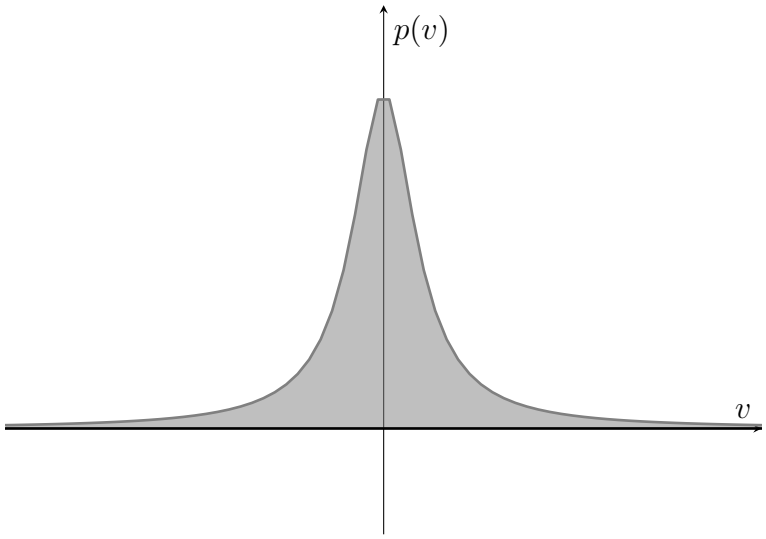

Figure 1: Probability density function $p(v)$ for the Agnesi game

Try to take the expected value of this prospect and you will find that it has none (Poisson, 1824). For continuous distributions like this, the expected value is given by the integral of $v \times p(v)$ from negative infinity to positive infinity (analogous to an expected sum: $v \times P(v)$ over all possible values v). But, for the Agnesi game, that integral between 0 and positive infinity is positively infinite! And, from 0 to negative infinity, it is negatively infinite! Sum these integrals together—equivalently, take the integral over *all* possible values of v—and the total is undefined. Much like the Pasadena game's expected sum, the Agnesi game's expected integral fails to converge absolutely. You might expect its expected value to converge to 0, since the distribution is symmetric about 0. But, no, it too defies expecations. So, expected value theory will fail to compare it to any outcome, to any other option, and even to itself.

You might think that neither of these prospects are realistic—that they are merely contrived, hypothetical options that we are sure never to encounter beyond the pages of a philosophy journal. As Hájek (2014, 565) says of one description of the Pasadena game, you might think that considering such an option is "...a highly idealised thought experiment about a physically impossible game." If so, you might not be troubled that expected value theory cannot compare either of these prospects to any other. You might think that we should simply ignore such prospects, and that expected value

[10]This curve was first described in print by de Fermat (c. 1659) and first analysed as a probability distribution by Poisson (1824). For a discussion of this distribution in the context of decision theory, see Alexander (2012).

theory still suffices for real-world decision-making.

Unfortunately, we cannot—there is reason to think we face such prospects in practice. When evaluating our options morally, if we consider the prospects for the moral value of the distant future and we maintain Impartiality and Additivity, then we have reason to think that every option ever available to us defies expectations. In the remainder of this section, I give two discrete arguments to this effect, the second more compelling than the first.[11]

But, first, a brief note on probabilities. I assume that the notion of an outcome's probability that ultimately determines moral betterness must be one of two notions. The first is its *evidential* probability: how probable that outcome is to result from the given option, on the present evidence of the agent deciding between that option and others (see Williamson, 2000, 209). The second possible notion is the outcome's *subjective* probability: how confident the decision-making agent is that that outcome will result from a given option.[12] If evidential probabilities are the morally relevant ones, and if our evidence prescribes expectation-defying prospects, then we will face difficulties. Or, if subjective probabilities are the relevant ones, agents who form their beliefs rationally given their evidence will face difficulties.

2.1 A possibility of Pasadena

A simple argument that our prospects for the total value of the future defy expectations goes like this.

It seems *possible* that a Pasadena game (or Agnesi game, or similar such game) will be played at some point in the future. It is possible that a stranger may approach you (or some other agent) and offer to toss a coin until it first lands heads, and effect some event with moral value determined by the number of coin flips. Likewise, it is possible that some other mechanism will produce moral value according to the same distribution of objective chances. Although perhaps physically unrealistic, we can at least *conceive* of this happening. It would be no (logical or metaphysical) impossibility for this to occur. And, given how little we know about the far future, you might think it overconfident to assign probability zero to any agent ever being subjected to such a game. So, the evidential probability of a Pasadena game someday being played, it seems, must be greater than zero.[13]

[11] I focus on the prospects of our causal future rather than of the world as a whole, for three reasons. The first is simplicity. The second is that there are moral views on which the proper objects of comparison are not worlds as a whole but instead consequences—the portion of the world that it is (nomologically) possible to influence in a given decision (see, e.g., Bostrom, 2011, §3.2). And the third is that, if our future prospects have undefined expected value, then so too will the prospects of the world as a whole (unless the value of events inside and outside our causal future are very strongly anti-correlated, and we have no reason to think that they are). So, it suffices to focus on the value of our causal future.

[12] Other notions, of *objective chances*, may also sometimes be morally relevant, but only insofar as they constrain the evidential and subjective probabilities. They do not *ultimately* determine moral betterness, I assume.

[13] This line of thinking might be captured in the much-discussed principle of *Regularity*: that only logically (or perhaps metaphysically, or doxastically) impossible propositions have evidential probability zero (see Edwards et al.,

And, as has often been discussed before, *any* prospect with real, non-zero probability p of the Pasadena game, no matter what other prospects it is mixed with, inherits the problems of the game itself—like the game itself, having any such probability p of the Pasadena game brings undefined expected value (Hájek and Smithson, 2012: pp. 39-42; Bartha, 2016: pp. 802-3). So, as long as we have some probability p of such a Pasadena game over moral value being run somewhere in the future, the overall prospect for the total value of the future will be undefined.

But is there such a probability of the Pasadena game someday being played? I do not think it clear that the answer must be yes. One reason for doubt is that the correct theory of epistemic rationality may be *knowledge-based*: it may include as evidence everything the agent *knows*, and so require that evidential probabilities be assigned only after conditionalising on the agent's knowledge (see Williamson, 2000, §10.3).[14] And you might think that we *know* that no one will ever be subjected to the Pasadena game. Why? Perhaps you know that some particular physical law holds, and any version of the Pasadena game that you can imagine would violate it.. Or perhaps you note that there are infinitely many different *possible* games that future people might face in their lives, but at most finitely many that anyone actually faces—from this, perhaps you can know that the Pasadena game, specifically, won't be among them. Or perhaps you simply think it so implausible or subjectively improbable that the Pasadena game is ever played that you conclude that you know it will not be. Whatever the reason, you might then conditionalise on this knowledge and assign the game evidential probability zero.

Another reason to doubt that the evidential probability of the Pasadena game is non-zero is this. It's one thing to think that any possible *outcome* should be assigned non-zero probability. But it's quite another to think that any possible probability distribution *over* outcomes should be assigned non-zero probability. It may be too overconfident to assign probability zero to the future having value v or greater, for any v.[15] But it would be a strictly stronger, and so less plausible, claim to say the same of assigning probability zero to the future having any possible *probability distribution* over values v and above. Perhaps doing the latter would not be too overconfident. Or at least, given the dire implications if you do so, perhaps epistemic rationality should not require that you entertain every such possible probability distribution (even if it *does* require you to entertain every possible *outcome*).

For either of these reasons, or perhaps others, you might be unconvinced of this argument for us facing expectation-defying moral prospects in practice. To be truly worried that expected value theory is not up to the task of comparing our moral prospects, we may need a more compelling

1963; Easwaran, 2014b). But this principle is controversial (see, for instance, Pruss, 2013).

[14]To similar effect, you might instead think that the correct *decision* theory is knowledge-based: that, when comparing prospects, we can evaluate each prospect once we conditionalise on our knowledge (see Liu, n.d.) .

[15]This claim could be treated as a weakened form of Regularity (see Footnote 13), such as: that, only for a logically (or perhaps metaphysically, or doxastically) impossible outcome O can the proposition "Outcome O occurs." have an evidential probability of zero.

motivation—more compelling than the observation that facing the Pasadena game is merely *possible*.

2.2 A model of the distant future

Here is a more compelling argument that we face expectation-defying prospects in practice.

Consider some point in the distant future after which our empirical evidence tells us almost nothing about what will occur when. Specifically, let **t** be some future time (or more accurately, for reasons to do with general relativity, a point in spacetime)[16] such that all of our specific predictions of events *after* **t** are merely the uniform continuation of continuous physical trends from *before* it. In effect, **t** is a point after which all of our particular predictions of valuable future events are exhausted. Perhaps **t** is a billion years in the future; perhaps just 1,000 years in the future.[17]

However late **t** is, it is possible that humanity survives until then (or at least that *some* form of morally valuable life in our causal future survives until then). Regardless of how pessimistic you are about humanity's prospects, it seems wildly overconfident to assign probability zero to us not making it until after **t**, or to say that we *know* that we will not survive until then. (Indeed, it seems *far more* overconfident than assigning probability zero to the Pasadena game someday being played, or claiming knowledge that it won't be.) Then, conditional on us surviving until **t**, what of the prospects for life *beyond* that, as time stretches out indefinitely? What is the conditional probability of a further value v arising after **t**? Since we have no empirical evidence about events beyond **t**, by definition, the answer is not so clear.

Consider one way we might model value after **t**, albeit a very speculative one. In broad strokes, it will be a reasonably plausible one, but certainly not the only plausible model we might adopt. (For reasons explained below, the existence of other plausible models won't detract too much from the lessons we can draw from this one.)

We might model the moral value ocurring after **t** as the sum of value at discrete, isolated, and reproducing *clusters* of life. Focusing on humanity and other Earth-bound life, at present, we are

[16]The general theory of relativity tells us that there is no absolute notion of a *time* t, nor of the period before time t, nor the period after it—the set of events that we carve out as occurring at the same time t (or, equivalently, as being simultaneous with one another) is sensitive to the velocity at which we do the carving. But, when talking of a point in spacetime such as **t**, there is a set of events that occur *after* it when observed at any velocity. This set corresponds to those events within **t**'s *future lightcone*: the region to which, if you started from **t**, you could hypothetically reach while travelling at the speed of light or slower. Note that, in what follows, "after **t**" is meant as intuitive shorthand for "in the future lightcone of **t**" and "before **t**" as shorthand for "outside the future lightcone of **t**".

Note also that there are many possible points that we could define as **t** here; indeed, infinitely many! And with different choices of **t** may come different prospects over the value arising after **t**. Fortunately, what I say below will hold on *any* choice of **t**.

[17]Perhaps **t** lies after the so-called heat death of the universe. But note that even that predicted heat death is a continuation of a long-running trend of cosmological expansion—of the universe increasing in entropy which, beyond some point, it qualifies as having undergone heat death. Still, the universe will never quite reach a state of perfect entropy, so there is no genuine categorical difference between the time before heat death and the time after it (Dyson et al., 2002).

clustered together at one location, on a single planet. If we were to stay in this situation, it would be appropriate to assign a constant probability to all such life going extinct each year (or, since the risk of extinction may vary over time, at least a minimum, non-zero probability). But, more realistically, humanity might *not* remain so clustered; perhaps we will spread through space into many such clusters. As we spread further and further, some such clusters will be more and more isolated from others. For instance, if we imagine life spreading to different planet-like bodies throughout space (perhaps in different galaxies, or as far from each other as we like), the maximum spatial distance between one planet and its most distant counterparts will become greater and greater. Each such planet thereby becomes more and more isolated from its most distant counterparts—its inhabitants become better and better protected from calamities that arise on the most distant planets.

Indeed, given enough time, it most likely becomes *physically impossible* for events within one such cluster of life to affect other discrete clusters. This is implied by the most widely-accepted cosmological model (the 'flat-lambda' model), which predicts that, as our universe evolves in the distant future, it will continue to expand at an ever-accelerating rate—many star systems, galaxies, groups of galaxies, and other bodies about which civilisation might cluster will be pulled apart. Eventually, such clusters will be moving away from another so quickly (and continuing to accelerate) that events in one cluster will never be able to affect any other cluster, even if their effects travel at the speed of light (Nagamine and Loeb 2003; Busha et al. 2003; see Ord 2021 for an accessible survey).[18] And, *if* our descendants successfully isolate themselves from one another in this way, their extinction then seems far less likely. The extinction of humanity as a whole (and indeed all morally valuable life) would then require great calamities to happen *independently* in each of many isolated clusters of civilisation. This is far less likely than any individual calamity.[19] And, the more clusters, the lower the probability of overall extinction in a given time period.[20]

In this model of the future, absent such calamities, the number of clusters increases over time, at least in expectation. We can assume that each existing isolated cluster has the same (independent) probability of 'reproducing' by settling a new location that will eventually be isolated from it, and thereby creating a new cluster. I will also assume, as seems at least possible, that the probability of

[18]This is not the only way that clusters of life may become completely physically isolated—for instance, such isolated clusters would be generated if humanity created and populated new 'baby universes'. The possibility of doing this is somewhat supported by the prominent *inflationary view* of cosmology, under which our own universe was created by a quantum tunnelling event (see Vilenkin, 1983). It is far from settled whether inflationary cosmology would indeed allow this but, on our current understanding, it is certainly a live possibility (Farhi et al., 1990). And, independently, there is theoretical support for it being possible to create new universes via the formation of black holes, and that universes created in this way may be temporarily accessible to their creators (Brandenberger et al., 2021; Frolov et al., 1990). The science is far from settled but, based on our current evidence, it is a live possibility. (For an accessible survey of this topic, see Merali, 2017, .)

[19]Cf. Sandberg and Armstrong (2012).

[20]As above, relativity makes things more complicated here (see also footnote 16). Our carving up of events in spacetime into time periods, and our measurement of the duration of such time periods, is sensitive to the velocity at which we do the carving up and the measuring. But still, at any such velocity, it will hold that more clusters means a lower probability of overall extinction. Here and in what follows, by "at a time" or "in a given time period" or "the number of years", I mean this as determined by some observer travelling at *any* given velocity.

a cluster reproducing in a given time period is at least as great as its probability of dying off.

And, the more clusters, the more moral value there plausibly is. We can assume—conservatively, as it ignores growth within each cluster—that the total moral value arising in the world in a given year is proportional to the number of such clusters that then exist. The total (absolute) value after **t** then, again assuming Impartiality and Additivity, will be roughly proportional to the sum of the lifetimes of every such cluster to ever exist. But that total value may be positive or negative—there is some risk that the future of life in our universe may be one of immense misery. Or, at least, we should be uncertain about the relation between total number of cluster-years and total value—uncertain of the average value of a year of such a cluster existing. For simplicity, I will assume that there is a simple distribution over what this average value will be: probability 0.5 that it is some value v and probability 0.5 that it is $-v$; and this is (roughly) independent of our uncertainty of how *many* clusters there are. (This distribution is unrealistic but, with some further tweaks below, will be realistic enough to draw some useful lessons.)

If we combine these assumptions, the arrangement of clusters forms a stochastic process known as a *birth-and-death* process (or, more specifically, a *Kendall process*—see Kendall, 1948). Individual clusters reproduce and die off independently, much like members of a population. And what we care about is the total number of cluster-years that are ever lived, weighted by the average moral value of each cluster-year. (By assumption, it is equiprobable that the average cluster-year is positive or negative in value.) This gives us a rather complicated probability distribution over value.[21] But, fortunately, there is a prospect with a simpler distribution that shares its key properties: the Aquila game.[22] For simplicity, I will focus on the Aquila game, as given by the equation and plot below.

$$p(v) = \frac{a}{b + |v|\sqrt{|v|}} \qquad \text{for some constant } a, b > 0$$

[21] The above model, as a Kendall process with death and birth rates of μ and λ respectively, gives the following probability distribution over total cluster-years, multiplied by the average moral value of each (from McNeil, 1970, §5.b).

$$p(v) = \sqrt{\frac{\mu}{\lambda}}\, \frac{I_1(2|v|\sqrt{\mu\lambda})}{2|v|e^{|v|(\mu+\lambda)}}$$

Here, $I_1(x)$ is the first-order modified Bessel function of the first kind, which is equivalent to $\frac{1}{\pi}\int_0^\pi e^{x\cos\theta}\cos\theta d\theta$.

Crucially, when $\mu \leq \lambda$, that distribution lacks a defined expectation. It also matches the equation for the Aquila game above in that the variants of it raised in the following section behave the same as the corresponding variants of the Aquila game, both under expected value theory and under the various theories I introduce in Section 5. For my purposes, then, it suffices to focus on the (much simpler) Aquila game.

[22] Given its connection to the St Petersburg game and its cosmic motivation, the game takes its name from the location of the Petra system in our night sky: the Aquila constellation.

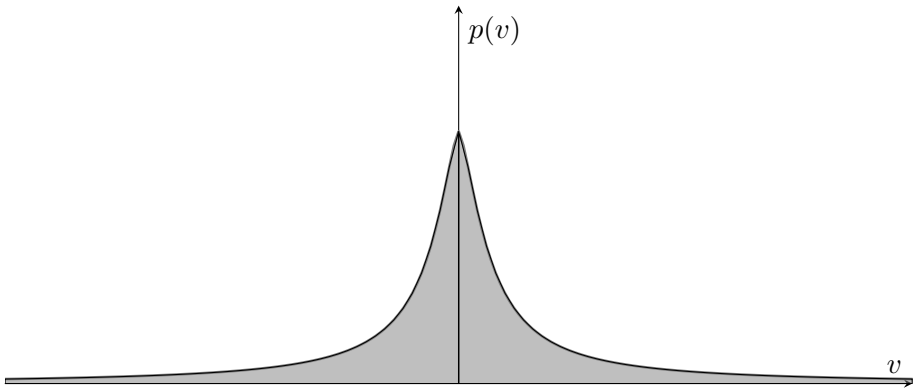

Figure 2: The probability density function over value for the Aquila game

Just like Pasadena and Agnesi, attempt to take the Aquila game's expected value and you will find that it has none. Its distribution is symmetric about 0, so you might expect it to have expectation 0. But, like the Agnesi game above, the probability density in its tails—as v approaches $\pm\infty$—approaches 0 sufficiently slowly that the expected value integral is undefined.

The same goes for the prospect for the total value of the world *overall*, even if we take into account other very different models of what occurs after $\mathbf{t}$, and even if we include value occuring both before and after $\mathbf{t}$. Why? Like Pasadena, we can mix the Aquila game with any other prospect and the overall prospect will defy expectations too. Similarly, we can add the payoff of Aquila to any other prospect (such as the prospect over the value of events before $\mathbf{t}$) and the prospect over the overall payoff will defy expectations.[23] So, if (a prospect that behaves like) the Aquila game is at least one minimally probable prospect for what happens after $\mathbf{t}$, then expected value theory will fail to compare *every* pair of options we might ever come across in practice.

But is the model described above even a *plausible* model for the value of the distant future? Must we really assign *any* probability to a prospect like the Aquila game being generated, such that the overall prospect inherits its expectation-defying property? You might be sceptical. Here are three reasons why, and why I do not think they undermine the claim that prospects for the total value of the world will behave like the Aquila game.

The first reason for scepticism: perhaps the number of clusters of, and value of, civilisation simply couldn't continue growing forever. Perhaps eternal exponential growth, whether it is achieved by spreading outwards in an ever-expanding cosmos or by creating baby universes, is physically (or metaphysically) impossible. This may well be true! But, most plausibly, we do not *know* that it is. And that could be enough to make it rational to assign at least *some* non-zero probability to an average such growth rate (at least in the absence of catastrophes). But even if we did know that eternal exponential growth were impossible, the above model does not require it. Indeed, the Aquila

[23] As above, I assume that events before and after $\mathbf{t}$ are not strongly anti-correlated (see Footnote 13).

game assigns probability 0 to the total survival time, or the total number of cluster-years, being infinite. We cannot rule out the model on these grounds.

The second reason why the above model may seem unrealistic: you might think that some possible extinction scenarios would strike every cluster of civilisation at once; perhaps some exotic physical phenomenon could simultaneously remove the conditions necessary for morally valuable life everywhere. If so, the annual probability of extinction of each cluster would not be entirely independent of others'. And, given this, the annual probability of overall extinction would not be brought arbitrarily close to 0 by simply adding more and more clusters. But still this does not prevent the prospect of overall future value from resembling the Aquila game. Even if there is some annual probability of civilisation-wide extinction, whether we avoid extinction in one year (conditional on having survived until the previous year) is not independent of whether we avoid it in every other year (conditional on having survived until the year before). In some states of the world, phenomena that extinguish all life at once are physically possible; in some states they are not. In states of the latter kind, having arbitrarily many isolated clusters of life does provide arbitrarily much protection from extinction. So, we should assign at least *some* non-zero probability to such extinction-causing phenomena being physically impossible. And so we can treat the overall prospect as a mixture of the prospect in which such phenomena are impossible and the prospect in which they are possible—in effect, a gamble between the Aquila game and something else. And so the overall prospect we obtain will still have tails resembling the Aquila game, since it offers some non-zero probability of playing such a game. And, since the Aquila game defies expectations, then the overall prospect will too. So it suffices to analyse the Aquila game in place of the more complicated overall prospect.

The third reason: it seems implausible that the average life is just as likely to be negative in value as it is to be positive, and of equal absolute value (on whichever interval scale we use to represent value). It seems to me at least that any future civilisation will more likely aim to make its descendants happy than aim to make them miserable (or, more generally, to have valuable experiences rather than disvaluable ones), and that its probability of success in this goal is better than chance. This probability of success seems *far* better than chance once we recognise that humanity in the far future, if it's still around, will likely have access to far more advanced technologies and greater resources than we do. Or perhaps you are pessimistic about humanity's future technological level, its available resources, or its inclination to benefit posterity. Perhaps our descendants are particularly likely to succumb to scenarios of widespread misery (for discussion of such possibilities, see Baumann, 2017). If you think so, you might think the prospect for the average life skews towards misery rather than happiness. Either way, my earlier assumption that the average life has probability 0.5 of having value some $v > 0$ and probability 0.5 of $-v$ would be false. Rather, one of these possibilities will have higher probability than the other, and so the distribution will skew one way or the other.[24]

[24]The distribution will likely also be far more spread out than this, but I will put that complication aside, as it will simply result in an overall distribution with tails that approach 0 even more slowly than the Aquila game. The same

Given this skew, the true distribution over future moral value will not be symmetric like the Aquila game. It will be skewed in either the positive or negative direction, as illustrated below. This more general *Skewed Aquila Game* has a probability distribution given by the following equation (for some positive a_1 and a_2, representing the relative probabilities of total value being positive or negative).

$$p(v) = \begin{cases} \frac{a_1}{b+|v|(\sqrt{|v|}} & \text{for } v > 0 \\ \frac{a_2}{b+|v|(\sqrt{|v|}} & \text{for } v < 0 \end{cases}$$

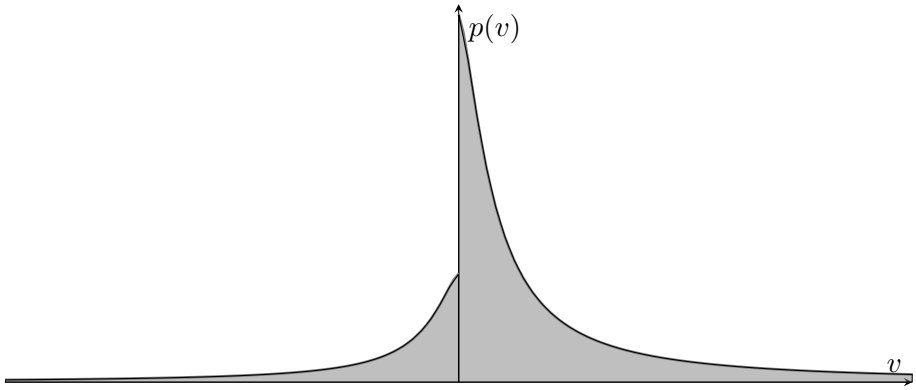

Figure 3: A probability density function over value for (a version of) the Skewed Aquila game

For simplicity, in much of what follows, I will focus on the more basic Aquila game. The problems I will describe that arise for comparing the Aquila game to alternatives will arise with equal force if we substitute in the Skewed Aquila Game, a mixture of the Aquila game with something else, or other more complicated variations.

3 Challenges for decision-making

If you model the value of the distant future involves *any* non-zero probability of the Aquila game, the Skewed Aquila game, or any similar such game, then you face a serious challenge. In practice, you cannot assign an expected moral value to any of the options ever available to you. So, if expected value theory (and nothing stronger) were the correct theory of moral betterness under risk, then no option ever available to you would be morally better or worse than any other. But to accept this implication would be absurd.

To plausibly compare any of our available future prospects, we must replace expected value theory with some stronger alternative. In later sections, I will discuss such alternative theories. But, first, what do we want them to achieve?

problems as below will arise and the same solutions will hold.

I propose five problem cases that those theories must be able to deal with—and deal with in the intuitively *correct* way—to be extensionally adequate.[25] These cases will be crude simplifications of the options we face in practice: they exclude all sources of value in the world other than the possibility of an Aquila game generated after **t**. In practice, we face options in which many valuable events will occur before **t**, in which there is perhaps only a small probability of life surviving until **t** to generate something resembling an Aquila game, and in which the prospect for events after **t** is far more complicated than the Aquila game. Nonetheless, it will largely suffice to consider simplified cases like these—as noted above, our available options will inherit the problems of the Aquila game. If expected value theory fails in the cases below, it will fail in practice. And it turns out that it does fail, as do many stronger theories designed to deal with the original Pasadena and Agnesi games.

The first problem case, *No Change*, is (a simplification of) the decision scenario an agent faces when their available actions all produce exactly the same future prospect. For instance, an agent may choose between eating Sugar Puffs for breakfast and eating Frosties, but have no evidence for either option being more or less likely to influence the future in any particular way. (Agents with great foresight may have access to evidence supporting some story of why one cereal is more likely to produce better long-run outcomes, but suppose that the agent here lacks any such evidence.)

Scenario 1: No Change

Sugar Puffs: The Aquila game with particular values of a, b.

Frosties: The Aquila game with the same a, b.

Note that both options have identical probability distributions over value. But, still, bare expected value theory cannot say how they compare—neither option has well-defined expected value, so that value cannot be equal to itself. (The same goes if we swap the Aquila game for the Skewed Aquila game.) And this is all the more troubling when, intuitively, the correct ranking of options seems clear: Sugar Puffs and Frosties are equally good. It would be desirable for our theory to say this, that any instance of the Aquila game with such and such parameters is equally as good as any other with the same parameters.[26]

The second problem case, *Improving the Present*, is that which an agent faces when they can improve some aspect of the world with certainty,[27] without otherwise changing the prospect. For

[25] Note that dealing with these five problem cases is *necessary* but perhaps not *sufficient* for extensional adequacy. Extensional adequacy may require dealing with cases even more complicated than those I consider here.

[26] Failure to rank these options as equally good can also be characterised as a violation of *Stochastic Equivalence*.

> *Stochastic Equivalence*: For any options O_a and O_b, if for every possible outcome O both O_a and O_b have the same probability of resulting in an outcome equally as good as O, then O_a and O_b are equally good.

This principle is both intuitively very plausible and one that expected value theory easily satisfies for finitely-supported prospects.

[27] If that improvement is less-than-certain, we have a slightly different scenario. Fortunately, each of the proposed theories below that give the correct verdict in Improving the Present happen to give the same verdict in this different scenario, so I will not dwell on that scenario here.

instance, an agent may choose whether to save the life of a child in the present day. And, regardless of whether they do so or do not, their evidence may entail an identical probability distribution over what happens in the very distant future. If so then, for my purposes, their options are equivalent to the following.

Scenario 2: Improving the Present

Do Nothing: The Aquila game (with particular a, b).

Save a Life: The Aquila game (with the same a, b) with value $s > 0$ added to every outcome.

Here, both options are identical *except* that the latter has its probability distribution shifted by some bonus value s.[28] But, again, expected value theory cannot compare any options of this sort. (And again, the same goes if we swap the Aquila game for the Skewed Aquila game here.) And, again, this is all the more troubling given that the intuitively correct ranking is clear: that, as long as s is positive, Save a Life should be better than Do Nothing. Improving every outcome should improve the option overall (so long as the outcomes' probabilities are held fixed).

The third problem case, *Improving the Future*, is one that an agent may face if they attempt to improve events occurring after $\mathbf{t}$, conditional on us surviving until then. In this case, the agent's choices don't make it more or less likely that we survive but, if we do survive until then, those choices make it more or less likely that the average life afterwards will have positive value (on whichever interval scale we represent value). It is a decision of whether to alter the skew of the Skewed Aquila game one way or the other. And, to an approximation, this is the sort of case an agent might face when they can affect humanity's long-run prospects in some manner that is extremely persistent. Perhaps it applies when a political activist decides whether to campaign for a change to political institutions that would foreseeably improve decision-making. Doing so may make it ever so slightly more likely that humanity at large has better political institutions indefinitely far into the future, perhaps increasing the probability that future lives (and clusters of civilisation) have positive value on average.

Scenario 3: Improving the Future

Campaign: The Skewed Aquila game with some $\frac{a_1}{a_2}$ and b.

[28] This case is an analogue of the widely-discussed comparison of the Pasadena game to the *Altadena* game (introduced by Nover and Hájek, 2004, 241). In both cases, a failure to rank the latter option as better is a violation of *Weak Stochastic Dominance*.

> *Weak Stochastic Dominance*: If, for every possible outcome O, one option O_a has a strictly higher probability than another option O_b of an outcome at least as good as O, then O_a is strictly better than O_b.

Like Stochastic Equivalence, this principle is both intuitively very plausible and one that expected value theory easily satisfies for finitely-supported prospects.

Don't Campaign: The Skewed Aquila game with a *lower* $\frac{a_1}{a_2}$ (and the same b).

Again, expected value theory alone cannot compare the two. Nor can it say that Campaign is better than Don't Campaign—it cannot say that it is better to make it more likely that future lives are very good and less likely that they are very bad. Intuition demands that Campaign be ranked as better than Don't Campaign.[29]

The fourth problem case, *Reducing Extinction Risk*, is that which an agent faces when they can affect humanity's probability of long-term survival (and, *a fortiori*, the probability of morally valuable life surviving). If the agent does nothing, humanity will have some probability of surviving to **t** and beyond. If they intervene, humanity will have a *greater* probability of doing so. For my purposes, both options can be represented by some mixture of a low-value outcome (which, for simplicity, we can set to value 0) and the prospect obtained conditional on surviving the near term. For our purposes, those options are equivalent to the following.

Scenario 4: Reducing Extinction Risk

Intervene: A mixture of the (Skewed) Aquila game (with some a or a_1 and b) with probability $p > 0$ and an outcome of value 0 with probability $1 - p$.

Do Nothing: A mixture of the (Skewed) Aquila game (with the same a or a_1 and b) with probability $q < p$ and an outcome of value 0 with probability $1 - q$.

Here, both options are equivalent to having some probability of playing the Aquila game or Skewed Aquila game (with such and such parameters), with Intervene giving the higher probability. But, again, expected value theory cannot compare any two options fitting these descriptions. Expected value theory cannot say that Intervene is better; it cannot say reducing the risk of extinction is an improvement. Again, this is troubling.

Take the Aquila-game versions of Intervene and Do Nothing—each gives a probability of Aquila and a probability of value 0. It seems all the more troubling for expected value theory to say nothing in this case, given that the correct verdict may seem obvious. Both options seem equally good, in so far as expected value theory is plausible in the first place. After all, the Aquila game's distribution is *symmetric* about 0. For any value v, it has the same probability (density) of v and $-v$. These should cancel out. According to expected value theory, they would do if we were dealing with a prospect whose expectation were defined. To uphold the spirit of expected value theory, they must cancel out for the Aquila game too, such that the game is valued at 0. What, then, of Intervene and Do Nothing? Each is a mixture of that option valued at 0 and a further outcome with the same value. They should each be valued at 0 overall, and so be equally good.

[29]Similar to the failure in Improving the Present, a failure to rank Campaign as better than Don't Campaign is a violation of Weak Stochastic Dominance (see Footnote 32).

Or consider the *Skewed*-Aquila-game versions of the two options. If the Skewed Aquila game in question is skewed in the *positive* direction, then it can be obtained from the Aquila game by shifting probabilities such that it is more likely that future lives are very good and less likely that they are very bad. This is clearly an improvement, and so must be better than obtaining value 0. Then, Intervene would surely be better than Do Nothing—between the Skewed Aquila game and the outcome of value 0, it provides the higher probability of the better one. Or, if the Skewed Aquila game in question is skewed in the *negative* direction, it can be obtained from the Aquila game by making it more likely that future lives are very *bad* and less likely that they are very good. This is clearly worse than the original Aquila game, and so must be worse than an outcome of value 0. Then, Do Nothing must be better than Intervene—between the Skewed Aquila game and the outcome of value 0, it provides the higher probability of the better one.[30]

The fifth and most challenging problem case, *Multifarious Changes*, is a combination of the previous three. The agent does not merely improving/worsening the present with certainty, or changing the probability of human extinction before $\mathbf{t}$, or changing the probability of a good future conditional on survival. They have all three effects, or any subset of them, at once. This, I think, is a more realistic representation of many of our options. For instance, attempts to improve the long-term future often have some moral cost in the present—e.g., the opportunity cost of spending one's resources on lobbying for institutional change is that the same resources aren't directly used to help the poor. Or, when attempting to reduce the risk of extinction, there is often a further effect on the well-being of future people in the event of survival—e.g., implementing some measure to reduce the incidence of deadly pandemics not only reduces the risk of extinction, but likely also causes future people to experience fewer pandemics in general, whether or not they rise to the level of threatening extinction. Likewise, attempting to make future lives better conditional on survival will often affect the probability of extinction—e.g., if one succeeds in changing political institutions to better respond to the public's interests, those institutions would then likely also be better at responding to threats of extinction.

If an agent has any options have at least two of those effects in one then, for my purposes, we can model their decision as follows. (Recall that $\frac{a_1}{a_2}$ represents the skew of the Skewed Aquila game—the greater the fraction, the greater the game's skew towards positive value.

Scenario 5: Multifarious Changes

Intervene: A mixture of 1) the Skewed Aquila game with some $\frac{a_1}{a_2}$ and b, with value s added to every outcome, with probability p, and 2) an outcome of value s with probability $1-p$.

Do Nothing: A mixture of 1) the Skewed Aquila game with some (perhaps different) $\frac{a_1}{a_2}$

[30] A failure to rank these options in the ways described would also violate Weak Stochastic Dominance (see Footnote 32).

and b with probability $q \neq p$, and 2) an outcome of value 0 with probability $1 - q$.

From above, *a fortiori*, we know that expected value cannot compare (at least some) options fitting these descriptions. But nor, it turns out, can it compare *any* such options—any two such options will defy expectations. And again, this silence is troubling. It is not merely troubling because the correct ranking of the options is intuitively obvious; the correct ranking often won't be. But it is troubling that our normative may fall silent in a decision that we plausibly face in practice. If an agent ever has the opportunity to influence humanity's long-term future, it is plausible that they face this scenario, and they need guidance. For a decision theory to be plausible, it must offer such guidance in at least the cases we actually face in practice. But expected value theory cannot.

4 One escape: risk sensitivity

Given its failure in all five problem cases above, expected value theory alone cannot be the correct theory of instrumental moral betterness. If it were, no option ever available to us would be better than (or even comparable to) any other. And that would be absurd.

In later sections of this paper, I will argue that this absurdity can be avoided without rejecting the verdicts of expected value theory altogether—that the theory can be *extended* to deal with the problem cases raised above. But, before that, it's worth briefly considering an alternative solution. That solution is to reject expected value theory altogether, not in favour of some extension of it, but rejecting even the verdicts it makes in less troublesome cases. In its place, we could adopt a *risk-sensitive* decision theory. This would allow us to avoid absurdity in the above problem cases, as I will show in this section. In effect, the discussion up to this point might be seen as forming a surprising argument *in favour* of risk sensitivity.

To illustrate how risk sensitivity avoids absurdity in the above cases, consider one theory that exhibits it: *expected utility theory* (specifically, a risk-sensitive version of it). This theory works much like expected value theory does. Where expected value theory says that the best options are those with the highest expected moral *value*, expected utility theory says that the best options are those with the highest expected *utility*.

What is utility? For my purposes, it is some representation of the betterness ranking over outcomes. But it need not be the *same* representation as the moral value function. Utility here is not the same thing as what moral theorists sometimes call utility—a cardinal measure of total welfare—but instead a purely decision-theoretic construct.[31]

[31] As von Neumann and Morgenstern (1953, 28) put it, utility is simply "...that thing for which the calculus of mathematical expectations is legitimate."

In general, the utility of an outcome may be *any* increasing real-valued function of its moral value (at least when determining instrumental moral betterness), risk-sensitive or not, so long as that function is strictly increasing. In particular, the correct utility function for use in comparing options morally might sometimes be *concave*: the higher the *value* of outcomes, the less their *utility* increases for each additional unit of value that is added to them. This tends to lead to risk-averse preferences. And/or the utility function may sometimes be *convex*: the higher the value of the outcomes, the *more* their utility increases for each additional unit of value. This tends to lead to risk-*inclined* preferences. One possible function, $u(v)$, that is sometimes concave and sometimes convex is plotted below.

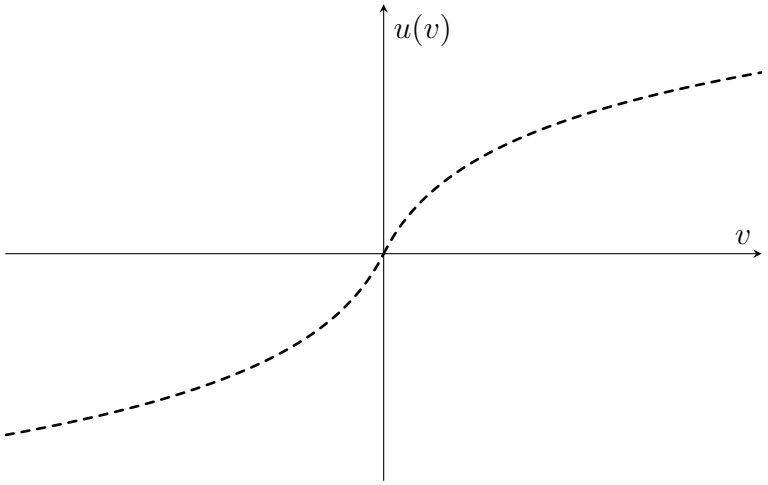

Figure 4: A utility function that is concave for $v > 0$ and convex for $v < 0$

But how does switching from expected *value* theory to expected *utility* theory, with a non-linear utility function, affect our comparisons of expectation-defying options? To see how, note that such options posed a problem for expected value theory only because the probability densities of outcomes didn't approach zero quickly enough as value approaches positive and negative infinity. If those extreme outcomes just had lower (perhaps *much* lower) absolute values, the options would no longer defy expectations, and expected value theory could evaluate them. But, in effect, they *do* have lower absolute 'values' if we switch to expected utility theory with a utility function like that plotted above—we lower the contribution that those extreme outcomes make to the expected utility calculation. Then, for the purpose of calculating expected utility, an expectation-defying option no longer defies expectations!

Take, for example, the Aquila game. Its troublesome distribution was given by $p(v) = \frac{a}{b+|v|\sqrt{v}}$ (for some $a, b > 0$). With a utility function that is concave enough for large positive values and convex enough for large negative values, we can turn that expectation-defying distribution *over value* into a much tamer distribution *over utility*. For instance, set $u(v) = \sqrt{|v|}$ such as that given by $p(u) = \frac{4a|u|^2}{b+|u|^3}$ (which gives an expected utility of 0).

This works in all five of the problem cases described above. In the first (No Change), where we must compare the Aquila game to an identical prospect, a utility function as above lets us say that the two options are equally good—each gets an expected utility and, since the prospects are identical, those expected utilities are equal. So, not only can expected utility theory say something here, but it says the intuitively correct thing. In the second case (Improving the Present), expected utility theory with a utility function as above lets us say that the Aquila game sweetened by value s is indeed better than the same Aquila game without that sweetener s. In the third case (Improving the Future), it says that increasing the Skewed Aquila game's skew in the positive direction is indeed an improvement. In the fourth case (Reducing Extinction Risk), where we compare two mixtures of the (Skewed) Aquila game, expected utility theory can again provide a comparison (although what it says will depend on the exact utility function). And, in the fifth (Multifarious Changes), again, it can compare any (perhaps sweetened mixture of) one Skewed Aquila game to another. In all five cases, it satisfies the desiderata I gave above.

And we can do the same with *any* pair of expectation-defying options; we need only adopt a utility function that is concave (convex) *enough* for large positive (negative) values. We need only accept a certain sensitivity to risk and the problem is solved. Thus, expected utility theory can deliver verdicts in those scenarios where expected value theory did not.[32]

And, thus, we seem to have a new and surprising argument in favour of risk sensitivity. Risk-sensitive decision theories can be compatible with Impartiality, Additivity, and the various empirical claims above, while expected value theory alone is not. Following this argument, perhaps we must reject expected value theory in favour of risk sensitivity.

But whether this argument succeeds depends on whether this is the *only* way to deal with those problem cases. Ultimately, it turns out that risk sensitivity is unnecessary to solve the problem. As I show below, we can extend expected value theory to deal with those problem cases. But it is worth keeping in mind what is at stake here: the result not only exonerates expected value theory from being incompatible with Impartiality and Additivity; it also undermines an otherwise compelling argument *in favour* of alternative theories that accommodate risk sensitivity.[33]

[32] A similar result could be achieved with a modified version of Buchak's (2013) *risk-weighted expected utility* (REU) theory (even with utility linear with respect to moral value). That theory says that an option O_a should be evaluated by $\text{REU}(L) = u_0 + \sum_{j=1}^{n}(u_j - u_{j-1}) \cdot r\big(P(L \geq u_j)\big)$, where the utilities of possible outcomes are given in ascending order by $\{u_0, u_1, ..., u_n\}$ and $r : [0,1] \to [0,1]$ is some non-decreasing function describing a particular risk attitude. When applied to prospects with continuous distributions and over outcomes with unbounded values, we might adjust the theory in two ways: 1) replace with the discrete sum with an integral; and 2) take separately the REU of the conditional prospects i) O_a, conditional on $u \geq 0$, and ii) O_a, conditional on $u < 0$, with the latter calculated 'in reverse', using the equation $\text{REU}(L|u < 0) = u_n - \sum_{j=1}^{n}(u_j - u_{j-1}) \cdot r^*\big(P(L < u_{j-1})\big)$ (and a suitable r^* function). Doing so has an effect similar to that under expected utility of adopting the utility function illustrated above. But proponents of REU theory may baulk at this modification of their theory—particularly (2)—which may seem ad hoc, arbitrary, and poorly motivated.

[33] There are other possible objections to this argument for risk sensitivity. The first is that the risk sensitivity needed to solve the problem will lack independent justification. Unlike the view advocated by Buchak (2013), the risk sensitivity necessary here will not arise from nor match the agent's own preferences over different means to their

5 Preserving risk neutrality

Can we deal with prospects like the Aquila game without embracing risk sensitivity? Rather than rejecting expected value theory altogether, can we extend the theory to deal with such problem cases? And, if so, how?

In this section, I consider several possible extensions. As we will see, many of the extensions so far proposed in the literature cannot deal with a prospect as troublesome as the Aquila game. But some can.[34]

5.1 Relative Expectation Theory

The first such extension is *Relative Expectation Theory*, first proposed by Colyvan (2008). Here, I will focus on the strengthened version suggested by both Colyvan and Hájek (2016, 837-8) and Meacham (2019, 13-7).

According to Relative Expectation Theory, we no longer attempt to assign some value to each option separately and compare those values. Instead, for each *pair* of options, we evaluate a *relative expectation* (RE): the expected *difference* in value between the two options; but, in calculating this difference, we match up the outcomes of each option by how far along the option's probability distribution they are. For any options O_a and O_b, we match up the lowest value of the possible outcomes of O_a to the lowest possible value for O_b; we match up the median values of each; we match up the best possible values of each; and likewise for every other possible value, matching each value from O_a with the value in O_b that is equally far along O_b's distribution. Put differently, we match each possible value in O_a to the value lying at the same *quantile* in O_b.

Formally, we identify the value that is fraction P of the way along the probability distribution of O with the quantile function $v_O(P)$—the function that, for each probability P, gives you the largest value v such that O has probability P (or less) of resulting in value v or less. For instance, $v_O(0.5)$

desired ends. Here, it must be imposed externally, and will often diverge from the agent's own attitudes. If anything, the standard motivation for risk sensitivity may tell *against* this argument.

A second, related objection is methodological. Even if we do accept that there is a correct universal, agent-neutral attitude to risk, we might think that the correct method to set this is by considering simple, idealised cases in which our normative intuitions are especially clear (cf. Buchak 2013, §2.3 and to reason from those to more complicated practical cases. If we instead determine the correct risk attitude based on the presence of options we *in fact* face, as we may need to to solve the problem described here, it may seem that we make a methodological mistake. (I am grateful to Johanna Thoma for suggesting this objection.)

A third objection is that, particularly in the moral setting, risk *neutrality* has some powerful arguments in its favour. These include Harsanyi's (1955) classic social aggregation theorem and various others (e.g., Thoma, 2019; Tarsney, n.d.; Zhao, 2021; Thomas, 2022a; Wilkinson, 2022a, n.d.a). By such arguments, if we adopt an aggregative theory of moral betterness but admit sensitivity to risk, we must violate one or another highly plausible principles.

[34]Note that I am interested here only in whether we *can* accommodate plausible verdicts without giving up risk neutrality—whether we can find an extensionally adequate theory—not the further question of whether we can independently motivate such a theory, which is beyond the scope of this paper. For independent motivation for the theory I endorse below, see Wilkinson (n.d.b).

would be the median, and $v_O(0.9)$ would be the value that O has only a probability 0.1 of exceeding. (Equivalently, $v_O(P)$ is the inverse of O's cumulative probability distribution; for an illustration, see below.) With this function, Relative Expectation Theory can be stated as follows.

Relative Expectation Theory: A option O_a is at least as good as another option O_b if

$$\mathrm{RE}(O_a, O_b) = \int_0^1 \big(v_{O_a}(P) - v_{O_b}(P)\big)dP \geqslant 0$$

Relative Expectation Theory agrees with all of the verdicts given by expected value theory. (To see this, note that, an option O_a's expected value can always be expressed with the integral $\int_0^1 v_{O_a}(P)dP$. So, when expectations exist, $\mathrm{RE}(O_a, O_b)$ simply becomes their difference.) But how does the theory fare in cases where expected value theory says nothing, such as those from earlier? Recall, for instance, the case of No Change. Relative Expectation Theory says that both options are equally good. Where O_a and O_b are both the Aquila game—precisely the same distribution—both will have the same quantile function v_O (matching the function labelled "Aquila game" in the figure below). So $v_{O_a}(P) - v_{O_b}(P)$ will always be 0, the integral from 0 to 1 will be 0, and they will be equally good.

Or consider Improving the Present. Relative Expectation Theory says that Save a Life is better than Do Nothing. Recall that Do Nothing was simply the Aquila game, while Save a Life was the same Aquila game but with every outcome sweetened by value $s > 0$. These options will have quantile functions as plotted below—functions that are identical, except that Save a Life's function is shifted up by value $s > 0$ for all P. The difference between the functions for Save a Life and Do Nothing is always positive, so the integral of $v_{O_a}(P) - v_{O_b}(P)$ from 0 to 1 (matching the area between the two graphs below) will be positive too, and Save a Life will be better. So, not only can Relative Expectation Theory compare the two, but it gives the intuitively correct verdict. For similar reasons, it also gives the intuitively correct verdict in the third case, Improving the Future.

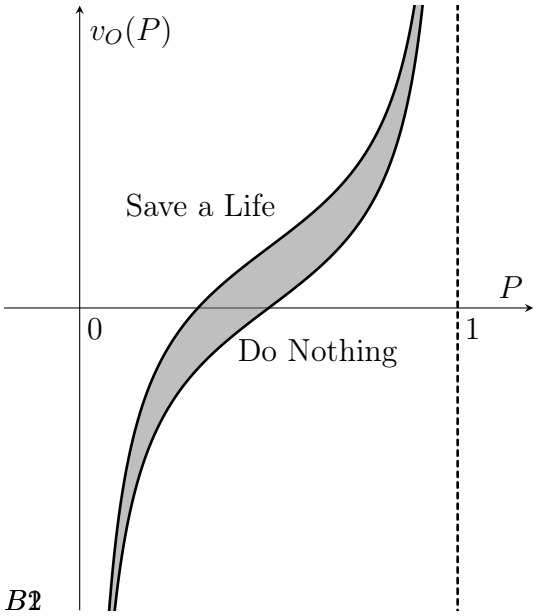

Figure 5: The quantile functions v_O of the options in Improving the Present: Do Nothing (the Aquila game); and Save a Life (the same Aquila game with each outcome sweetened by $s > 0$)

But Relative Expectation Theory cannot say anything in the fourth and fifth scenarios (Reducing Extinction Risk and Multifarious Changes). As has been noted elsewhere, the theory cannot compare an expectation-defying option to a sure outcome of value 0 (Colyvan and Hájek, 2016; Meacham, 2019)—$\mathrm{RE}(O_a, O_b)$ becomes the expected value of the expectation-defying option which, by definition, is undefined. And, so, options that differ by increases or decreases in the probability of the Aquila game (as do the options in both Reducing Extinction Risk and Multifarious Changes) will have undefined $\mathrm{RE}(O_a, O_b)$ too. Where other authors have observed this implication, they have accepted it—Pasadena and its kin are peculiar prospects, so it is not clear how we should compare them to the status quo, nor how good they are. But it cannot be a proper implication of decision theory that it falls silent in a wide range of decisions we actually face in practice, particularly moral decisions. And yet, with that verdict, Relative Expectation Theory does—it must fall silent in Reducing Extinction Risk and, *a fortiori*, in Multifarious Changes. Whenever an agent faces a decision that affects the probability that humanity survives rather than perishes, this theory fails us. So, I suggest, it proves inadequate.

5.2 Weak Expectation Theory

Another such extension comes from Easwaran (2008). By this proposal, an option is assigned a value called a *weak expectation*, where it exists, and options are ranked according to these values. Specifically, an option's weak expectation $\mathrm{WE}(O_a)$ is the value that, if the option of O_a were rerun in arbitrarily many independent trials, its *average* payoff (given by $\frac{O_a^{\times n}}{n}$) would be arbitrarily likely

to be close to. Put formally:

> *Weak Expectation Theory*: An option O_a is at least as good as an option O_b if $\mathrm{WE}(O_a) \geq \mathrm{WE}(O_b)$, for some $\mathrm{WE}(O_a)$ and $\mathrm{WE}(O_b)$ such that, for any small $\varepsilon > 0$:
>
> $$\lim_{n\to\infty} Pr\Big(\Big|\frac{O_a^{\times n}}{n} - \mathrm{WE}(O_a)\Big| < \varepsilon\Big) = 1 \qquad \text{(and similarly for } O_b\text{)}$$

But many options that defy expectations also defy weak expectations. For instance, both the Aquila and Skewed Aquila game do. (As does any mixture, or independent sum, of them with other prospects.) You could run arbitrarily many independent trials of one of these games, sum together the values of their outcomes and average them out, and the probability distribution you end up with for the average value will be just as spread out as the Aquila game or Skewed Aquila game you started with. There is no weak expectation to which the average is guaranteed to converge.

As a result, Weak Expectation Theory cannot evaluate *any* of the options featured in the five cases above. Nor can it compare any of those options to any other. Whenever we face any moral decision whatsoever, the theory fails us. Like Relative Expectation Theory, it is inadequate.

5.3 Invariant Value Theory

But the silence of the the previous two proposals does not mean that *no* extension of expected value theory that can sensibly compare the Aquila game to alternatives. Several proposals do so but, for ease of exposition, I will focus on just one here: *Invariant Value Theory*.[35]

Under this theory, like Relative Expectation Theory, we focus on the quantile function of each option. (That is, for each probability $0 \leq P \leq 1$, the largest value v such that the option has probability P of outcomes with greater value.) The quantile function for the Aquila game, for instance, is plotted below. We also take advantage of the fact that an option's expected value can

[35]Other proposals include Easwaran's (2014a) *Principal Value Theory*, which can deal with the first three cases, and Meacham's (2019) further extension of *Difference Minimising Theory*, which can deal with the latter two cases. Principal Value Theory operates much like Invariant Value Theory, except it truncates the option's probability distribution rather than its quantile function, and takes the limit of the option's expectation as the truncated distribution approaches the true distribution. If we define $O_a^{|v|\leqslant n}$ as the prospect that assigns the same probability as O_a to every possible value with absolute value up to n, and redistributes the remaining probability mass (taken from values below $-n$ and above $+n$) to value 0, then the theory can be characterised as follows.

> *Principal Value Theory*: A prospect O_a is at least as good as another prospect O_b if $\mathrm{PV}(O_a) \geq \mathrm{PV}(O_b)$, where
>
> $$\mathrm{PV}(O) = \lim_{n\to\infty} \mathbb{E}(O^{|v|\leqslant n})$$
>
> and O_a and O_b each satisfy a technical condition called *stability* (see Easwaran, 2014a, 524-5).

It turns out that, in the five problem cases from earlier, each option has a defined principal value PV (and satisfies Easwaran's stability condition). As a result, the theory can successfully compare all five pairs of options. (Meacham's proposal can too, as it is a further extension of Principal Value Theory, effectively combining it with Relative Expectation Theory.)

be given as the integral of its quantile function (and so the area between the quantile function and the horizontal axis) from 0 to 1.

For options like the Aquila game, of course, the expected value is undefined. But we can consider *truncated* versions of options: we can ignore the portion of the quantile function close to 0 (to the left of some small ε) and the portion close to 1 (to the right of $1 - \varepsilon$), as illustrated below, and take the expected value as the area under the quantile function between those endpoints.[36] We can also consider how that expected value changes as we truncate closer and closer to 0 and 1 (as ε approaches 0). If, as we get closer to the full quantile function, the expected value approaches some finite limit, then that limit seems an appropriate value to assign to the option. We can call that limit the *invariant value* of the option, and it is by this value that Invariant Value Theory has us evaluate options.

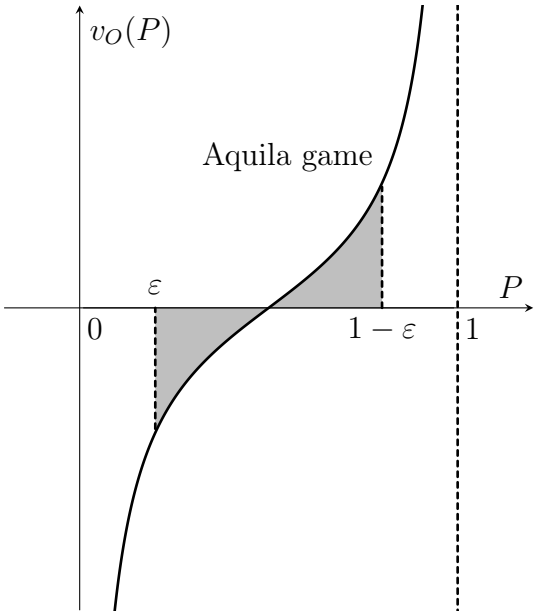

Figure 6: The quantile function v_O of the Aquila game

Put precisely, the theory says the following.[37]

> *Invariant Value Theory*: An option O_a is at least as good as another option O_b if $\mathrm{IV}(O_a) \geq \mathrm{IV}(O_b)$, where:
>
> $$\mathrm{IV}(O) = \lim_{\varepsilon \to 0^+} \int_{\varepsilon}^{1-\varepsilon} v_O(P) dP$$

This proposal maintains all of the verdicts of expected value theory. (Since the expected value of an option is the integral of its quantile function from 0 to 1, where that integral is defined, it will

[36] Smith (2014) proposes a similar method of truncation, although his final proposal is very different.

[37] For a fuller discussion of the advantages and disadvantages of this theory relative to its rivals, as well as an axiomatic treatment of it, see Wilkinson (n.d.b).

match the limit given above.) It also happens to maintain all of the verdicts of Weak Expectation Theory (see Wilkinson, n.d.b).

But it goes much further than either of those theories. For instance, it assigns a value to the Aquila game—it assigns the game an invariant value of 0. To see this, note that the Aquila game's probability distribution is symmetric about value 0. So too, its quantile function is rotationally symmetric about probability 0.5 (where its value is 0). We can take any probability $\varepsilon < 0.5$ and $v_O(\varepsilon)$ will be the same as $v_O(1-\varepsilon)$, but negative. So too, for any such ε, the integral of $v_O(P)$ from ε to 0.5 will be the same as that from 0.5 to $1-\varepsilon$, but negative; they will perfectly cancel out. No matter how small ε gets, even approaching 0, the integral will be 0. So the invariant value of the Aquila game will be 0—as it seems it should, since the game's probability distribution is symmetric about 0. Likewise, in general, any option with probability distribution symmetric about some value v will be given invariant value v.

This tells us all we need to know to deal with the first and second problem cases from earlier. In No Change, Invariant Value Theory assigns invariant value 0 to both options. Both options are evaluable, and both are equally good. And, in Improving the Present, it assigns value 0 to Do Nothing and value s to Save a Life (i.e., the Aquila game with value s added to every outcome). If $s > 0$ then Save a Life is better than Do Nothing, as intuition suggests it must be.

To deal with the latter three cases, we can extend the theory slightly. In effect, we can combine it with Relative Expectation Theory.[38] Instead of taking the invariant value of each option, we can take a *relative invariant value* between any two options—the relative expectation of the two (as described above), but truncated at ε and $1-\varepsilon$, and taking the limit as ε approaches 0.

*Invariant Value Theory**: An option O_a is at least as good as another option O_b if

$$\mathrm{IV}^*\big(O_a, O_b\big) = \lim_{\varepsilon \to 0^+} \int_{\varepsilon}^{1-\varepsilon} \Big(v_{O_a}(P) - v_{O_b}(P)\Big) dP \geq 0$$

Take the third problem case, Improving the Future. In it, we must compare two Skewed Aquila games: one corresponding to Campaign; and another, corresponding to Don't Campaign, with a lower probability of the average future life having positive value and a higher probability of negative value (i.e., a lower ratio $\frac{a_1}{a_2}$). As illustrated below, the quantile function of Campaign is *always* higher than that of Don't Campaign. That is, the difference $v_{O_a}(P) - v_{O_b}(P)$ will always be positive, and so too must be the integral of that difference. Hence, so will the IV* of Campaign relative to Don't Campaign—the theory will judge Campaign as better, as intuition demands.

[38] This is analogous to Meacham's (2019, 1021) method of extending Easwaran's (2014a) Principal Value Theory.

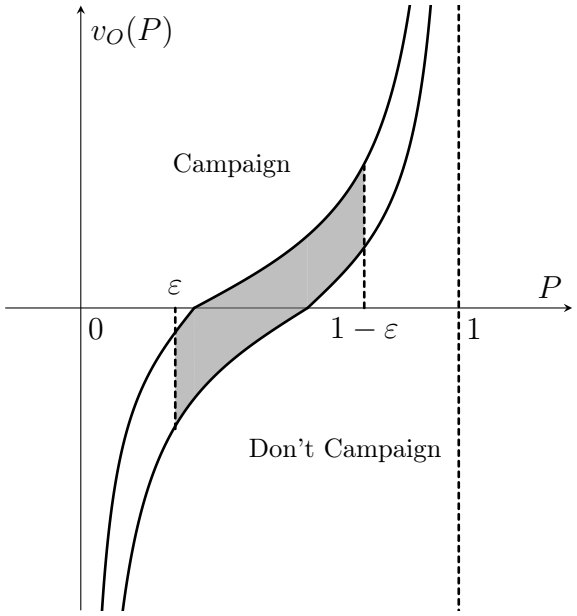

Figure 7: The quantile functions of the options in Improving the Future: Campaign (the Skewed Aquila game) and Don't Campaign (the Skewed Aquila game with a lower $\frac{a_1}{a_2}$)

Or consider the fourth problem case, Reducing Extinction Risk. In it, we must compare an option called Do Nothing, with some probability of the (perhaps Skewed) Aquila game, to another option called Intervene, with a *lower* probability of the same (perhaps Skewed) Aquila game, each of which otherwise result in an outcome of value 0. If we are dealing with the standard, unskewed Aquila game then, for the same reasons as above, both options have invariant value 0—the theory will say that they are equally good. If we are dealing with a Skewed Aquila game, the situation is more complicated, but can still be dealt with. If the Skewed Aquila game in question is skewed towards positive values, as illustrated below, IV*(Intervene, Do Nothing) will be positive—the area between the curves where Intervene has higher quantile function will be counted more quickly than the area where Do Nothing has the higher quantile function. So, the theory will say that Intervene is better, as seems intuitively correct. Similarly, where the Skewed Aquila game is skewed towards negative values, the theory will say that Do Nothing is better.

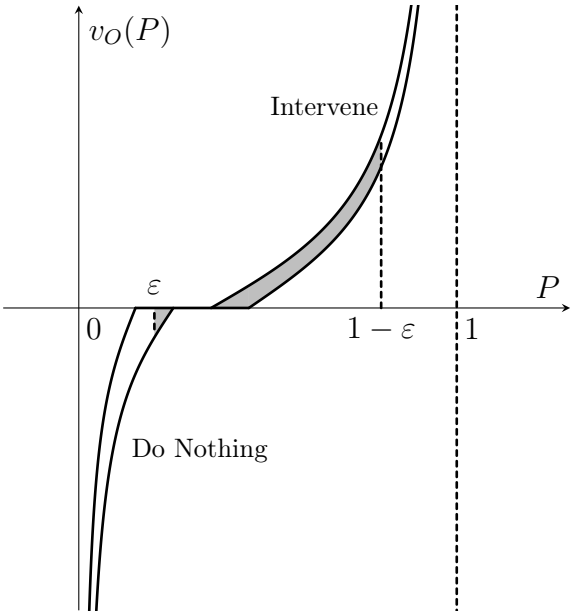

Figure 8: The quantile functions of the options in a version of Reducing Extinction Risk. Both options, Do Nothing and Intervene, are mixtures of a Skewed Aquila game with positive skew and an outcome with value 0.

What of the fifth case, Multifarious Changes? Again, we must compare different mixtures O_p and O_q of (perhaps Skewed) Aquila games, but those mixtures may be sweetend by different constant amounts; they may also involve *different* Skewed Aquila games (with different values of $\frac{a_1}{a_2}$). We are no longer comparing transformations of the same underlying option; we must now compare transformations of entirely *different* Aquila games. But, again, it turns out that Invariant Value Theory* can do so.[39] Even in this most challenging of the problem cases, the theory succeeds in providing guidance.

Thus, expected value theory can be extended to provide verdicts in all of the problem cases raised above; and not just any verdicts, but those aligning with intuition. As Invariant Value Theory(*) demonstrates, we need not abandon expected value theory and resort to risk sensitivity to evaluate our future prospects, at least not due to the models of the future described above.

Before moving on, it is worth noting that Invariant Value Theory(*) (as well as other extensions of expected value theory that can deal with these cases) faces certain objections. If fatal, these objections may mean that we *cannot* deal with these cases while preserving the verdicts of expected value theory, at least not in any plausible way. But I think that, fortunately, they are not fatal.

The first such objection is that Invariant Value Theory* violates a prima facie very plausible principle, and common axiom of expected value theory: *Independence*.

[39]This very general claim follows from Theorems 2 and 3 in Wilkinson (n.d.b). Any pair of perhaps sweetened, perhaps skewed, and perhaps mixed versions of the Aquila game in this case will satisfy the conditions given therein, and so be comparable by Invariant Value Theory*.

> *Independence*: For any options O_a, O_b, and O_c and any probability p, O_a is at least as good as O_b if and only if a mixture of O_a with probability p and O_c with probability $1-p$ is at least as good as a mixture of O_b and O_c with the same probabilities.

By Independence, it does not matter what we mix O_a and O_b with; if one is better, it remains better even when we mix each of them with some further option. In so far as we find the verdicts of expected value theory plausible, this principle seems highly plausible too. But it turns out to be violated by Invariant Value Theory*. (For a demonstration of this, see Wilkinson n.d.b, §6.2.)[40] This may be reason to reject the theory. Or it may instead be seen as a feature, rather than a bug—impossibility results given elsewhere (see Wilkinson, n.d.b, §6.2) tell us that a decision theory *cannot* satisfy Independence without violating one of several other very plausible conditions. Nonetheless, this is one reason we might doubt that this extension of expected value theory is satisfactory.

A second objection is that the definition of invariant value may be unavoidably arbitrary.[41] Like expected values, invariant values are a form of probability-weighted sum of the value an option might result in. But, unlike expected values, they are obtained only by summing in a particular order (or, equivalently, by taking a particular limit)—we start from the median of the option's distribution, and take the probability-weighted sum of value according to its quantile, summing outwards towards quantiles ever closer to 0 and 1. Why sum in this order rather than any other?[42] Why not start from the quantile of 0.33 and sum twice as quickly towards 1 as towards 0? The choice may seem arbitrary. Perhaps it is. But there is one reason to think that it is not. There is *only one* order in which we can sum that will consistently evaluate symmetric prospects in the intuitively correct way—that assigns value v to any option with probability distribution symmetric about v. The only order in which we can sum to get this result is that used by invariant value theory: starting at the median and approaching the quantiles 0 and 1 at equal speed. So, if we accept that we ought to be indifferent between any value v and a symmetric spread about v—as I think we should, in so far as we find expected value theory plausible in the first place—this order is not arbitrary after all.

6 A notable implication

As demonstrated above, expected value theory can be extended to deal with the problem cases from earlier—even if our prospects for the total value of the future are described by versions of the (Skewed) Aquila game, we can still compare the options available to us. But the comparisons we reach are

[40] Independence is also violated by other extensions of expected value theory that can deal with the above cases, such as Meacham's (2019) Difference Minimising Theory—see Wilkinson (n.d.b).

[41] I am grateful to two anonymous reviewers for pressing me to address this objection.

[42] A different order is used in Easwaran's (2014a) proposal of Principal Value Theory: start at value 0, take the probability-weighted sum of value between values $-n$ and n, and let n tend to infinity. Much the same objection is raised against that theory by Alexander (2012, 720-1) and Easwaran (2014a, 528).

perhaps surprising.

Recall Multifarious Changes. In that case, our options may be *any* two versions of the Skewed Aquila game, sweetened by some value s and/or mixed with an outcome of value 0. Specifically, consider a version of the case where we must choose between a) sweetening the game by some large amount s, and b) slightly increasing the positive skew of the Skewed Aquila game and/or slightly increasing its probability. According to Invariant Value Theory*, the latter will *always* be better (provided that it is skewed in the positive direction).[43] And this holds *no matter* how large the sweetener in (a) is, and *no matter* how slight the changes in probability in (b). In effect, changes of the sort made in (b) are *infinitely* more valuable than any finite sweetening as in (a). This is perhaps surprising and counterintuitive. If the Skewed Aquila game does describe our prospects over the total moral value of the future, then some interventions focussed on the long-term future will be, in effect, infinitely valuable. Increasing the probability that future lives are positive and/or get lived at all will *always* be more valuable than improving (finite numbers of) present lives. And this will still hold no matter how small those changes in probability, and no matter how many present lives you might otherwise improve.

Notably, this implication is not peculiar to Invariant Value Theory*. It is also implied by the other extant extensions of expected value theory that can deal with each of these cases (i.e., those of Easwaran, 2014a; Meacham, 2019). And this is to be expected—the differences in the probability distributions of (a) and (b) are roughly analogous to the St Petersburg game, which any genuinely risk-neutral theory must say is better than any finite value (see Hájek and Nover, 2006, 706). I suspect that *no* faithful extension of expected value theory will be able to avoid saying that it is always more valuable to increase the skew or probability of a (positively skewed) Skewed Aquila game than to gain any finite value for sure.

Even if the Skewed Aquila game *doesn't* accurately describe our options in practice, we may still encounter a similar implication. Invariant Value Theory* says much the same for analogous changes to *any* probability distribution with undefined expected value, so long as it can evaluate those changes at all. (So, I suspect, will other faithful extensions of expected value theory.) So we might accept a model of the future very different to the one from earlier that gave us the Skewed Aquila game, or we might assign only a small probability to the earlier model alongside many others. And still, if the resulting probability distributions have undefined expected values (and if Invariant Value Theory* can compare them at all), then much the same will hold—it will be more valuable to increase the probability of future lives overall being positive or, if they are positive, the probability of future lives being lived at all than to improve (finite numbers of) present lives. Admittedly, this result is only suggestive—there might be a single correct model of the future which gives a defined expected value or, more plausibly, the collection of plausible models might result in probability distributions so

[43] Similarly, if the latter option is equally skewed and in the negative direction, then it will always be worse.

troublesome that even Invariant Value Theory* cannot compare them. In either such case, the lessons drawn here from the Skewed Aquila game would have no bearing on practical decision-making. But, if neither is the case, we have quite radical implications for how to evaluate actions that have a small probability of greatly altering the future.

7 Conclusion

This discussion started with three seemingly plausible normative claims: Impartiality, Additivity, and expected value theory. In practice, are these claims compatible? Or, in practice, do they lead to absurdity?

As I have argued, there is some reason to think that the total expected moral value of the future is undefined. There is at least one plausible model of morally valuable events in the distant future that, if we accept Impartiality and Additivity, gives a probability distribution (the Aquila game or Skewed Aquila game) over moral value that has undefined expectation. Assign *any* non-zero probability to this model and, no matter what other models we might consider nor what we might expect to happen in the near-term future, our overall prospects for total moral value will inherit that undefined expected value. And this is bad news for expected value theory. If it alone were the correct decision theory, then no option ever available to us in practice would ever be morally better than any other. And this would be absurd.

One possible response to this absurdity is to abandon the verdicts of expected value theory altogether, in favour of some alternative theory that exhibits risk *sensitivity*. As demonstrated above, by doing so, we can effectively turn any expectation-defying option into a better-behaved one. If this is the only way to avoid absurdity, while holding onto Impartiality and Additivity, we may have a surprising argument in favour of risk sensitivity. But is this the only possible solution? Or can we preserve the risk-neutral verdicts of expected value theory somehow?

It turns out that we can—that risk sensitivity may not be necessary. We can *extend* expected value theory to deal with the expectation-defying options described here. Admittedly, not just any old extension of expected value theory will do—some proposals are insufficient (e..g., Relative Expectation Theory and Weak Expectation Theory). But other proposals do better, including Invariant Value Theory(*). As demonstrated above, with such an extension, we can deliver comparisons even in those various problem cases involving the (Skewed) Aquila game. (Doing so also brings on some surprising implications, as described in the previous section.)

Does this mean that Impartiality, Additivity, and (some extension of) expected value theory are perfectly compatible in practice; that we need not accept risk sensitivity? Maybe; maybe not. If the five problem cases given above accurately describe the decisions we face in practice, then yes. Even if

they loosely describe our real-world decisions—if our real-world options have probability distributions that behave sufficiently like the (Skewed) Aquila game—then the answer is yes. But our real-world options may also be far more complicated. For instance, the model of the distant future described above is just one possible model. There may be far more complicated models to which we should assign some probability. And, if those models give us even more challenging probability distributions, perhaps neither Invariant Value Theory(*) nor even further extensions of expected value theory will be able to compare the options we then face. If so, we may have a compelling argument for risk sensitivity once more.

This somewhat limits the conclusions that can be drawn here. We cannot conclude that Impartiality, Additivity, and expected value theory are *guaranteed* to be compatible in practice. But at least one seemingly troublesome argument against their compatibility has been undermined—on one fairly plausible model of our future, on which they seemed to conflict, they have been shown to cohere perfectly well. Perhaps there are other plausible models of the future on which they do still conflict, but this remains to be seen. For now, absent such models being proposed, it seems that we can safely endorse all three principles.

References

Al-Kindī, 1974. *Al-Kindī's Metaphysics: A Translation of Ya'qūb ibn Ishāq al-Kindī's Treatise 'On First Philosophy'*. State University of New York Press, Albany. (cited on page 3)

Alexander, J. M., 2012. Decision theory meets the Witch of Agnesi. *Journal of Philosophy*, 109, 12 (2012), pp. 712–27. (cited on pages 1, 4, and 28)

Bartha, P. F. A., 2016. Making do without expectations. *Mind*, 125, 499 (2016), pp. 799–827. (cited on pages 3 and 6)

Baumann, T., 2017. S-risks: An introduction. *Center for Reducing Suffering*, available at: https://centerforreducingsuffering.org/research/intro/ (accessed March 2022). (cited on page 11)

Bostrom, N., 2011. Infinite ethics. *Analysis and Metaphysics*, 10 (2011), pp. 9–59. (cited on pages 1 and 5)

Brandenberger, R.; Heisenberg, L.; and Robnik, J., 2021. Through a black hole into a new universe. *International Journal of Modern Physics D*, 30, 14 (2021), 2142001. (cited on page 8)

Broome, J., 2004. *Weighing Lives*. Blackwell. (cited on page 1)

Buchak, L., 2013. *Risk and Rationality*. Oxford University Press, Oxford. (cited on pages 19 and 20)

Busha, M. T.; Adams, F. C.; Wechsler, R. H.; and Evrard, A. E., 2003. Future evolution of cosmic structure in an accelerating universe. *The Astrophysical Journal*, 596, 2 (2003), 713. (cited on page 8)

Colyvan, M., 2008. Relative expectation theory. *Journal of Philosophy*, 105, 1 (2008), pp. 37–44. (cited on pages 2 and 20)

Colyvan, M. and Hájek, A., 2016. Making ado without expectations. *Mind*, 125, 499 (2016), pp. 829–857. (cited on pages 20 and 22)

Craig, W. L., 1979. Whitrow and Popper on the impossibility of an infinite past. *British Journal for the Philosophy of Science*, 30, 2 (1979), pp. 165–70. (cited on page 3)

de Fermat, P., c. 1659. De aequationum localium transmutatione et emendatione ad multimodaum curvilineorum inter se vel cum rectilineis comparationem, cui annectitur proportionis geometricae in quadrandis infinitis parabolis et hyperbolis usus. In *Œuvres de Pierre Fermat* (Eds. P. Tannery and C. Henry), p. 216–37. Gauthier-Villars. (cited on page 4)

Dyson, L.; Kleban, M.; and Susskind, L., 2002. Disturbing implications of a cosmological constant. *Journal of High Energy Physics*, 2002, 10 (2002), 011. (cited on page 7)

Easwaran, K., 2008. Strong and weak expectations. *Mind*, 117, 467 (2008), pp. 633–41. (cited on pages 2 and 22)

Easwaran, K., 2014a. Principal values and weak expectations. *Mind*, 123, 490 (2014), pp. 517–31. (cited on pages 2, 3, 23, 25, 28, and 29)

Easwaran, K., 2014b. Regularity and hyperreal credences. *Philosophical Review*, 123, 1 (2014), pp. 1–41. (cited on page 6)

Edwards, W.; Lindman, H.; and Savage, L. J., 1963. Bayesian statistical inference for psychological research. *Psychological Review*, 70, 3 (1963), pp. 193–242. (cited on page 5)

Farhi, E.; Guth, A. H.; and Guven, J., 1990. Is it possible to create a universe in the laboratory by quantum tunneling? *Nuclear Physics B*, 339, 2 (1990), 417–90. (cited on page 8)

Frolov, V. P.; Markov, M. A.; and Mukhanov, V. F., 1990. Black holes as possible sources of closed and semiclosed worlds. *Physical Review D*, 41, 2 (1990), 383. (cited on page 8)

Greaves, H. and MacAskill, W., 2021. The case for strong longtermism. Global Priorities Institute Working Paper Series. (cited on page 2)

Hájek, A., 2014. Unexpected expectations. *Mind*, 123, 490 (2014), pp. 533–67. (cited on page 4)

Hájek, A. and Nover, H., 2006. Perplexing expectations. *Mind*, 115, 459 (2006), pp. 703–20. (cited on page 29)

Hájek, A. and Smithson, M., 2012. Rationality and indeterminate probabilities. *Synthese*, 187 (2012), pp. 33–48. (cited on page 6)

Harsanyi, J. C., 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63, 4 (1955), pp. 309–21. (cited on pages 1 and 20)

Kendall, D. G., 1948. On the generalized "birth-and-death" process. *The Annals of Mathematical Statistics*, 19, 1 (1948), 1–15. (cited on page 9)

Liu, S., n.d. Don't bet the farm: Decision theory, inductive knowledge, and the St. Petersburg paradox. Unpublished manuscript. (cited on page 6)

McNeil, D. R., 1970. Integral functionals of birth and death processes and related limiting distributions. *The Annals of Mathematical Statistics*, 41, 2 (1970), p. 480–5. (cited on page 9)

Meacham, C., 2019. Difference minimizing theory. *Ergo*, 6, 35 (2019). (cited on pages 2, 20, 22, 23, 25, 28, and 29)

Merali, Z., 2017. *A Big Bang in a Little Room: The Quest to Create New Universes*. Hachette UK. (cited on page 8)

Nagamine, K. and Loeb, A., 2003. Future evolution of nearby large-scale structures in a universe dominated by a cosmological constant. *New Astronomy*, 8, 5 (2003), 439–448. (cited on page 8)

Nover, H. and Hájek, A., 2004. Vexing expectations. *Mind*, 113, 450 (2004), pp. 237–49. (cited on pages 1, 3, and 14)

Ord, T., 2021. The edges of our universe. Unpublished manuscript. Available at https://arxiv.org/abs/2104.01191. (cited on page 8)

Parfit, D., 1984. *Reasons and Persons*. Oxford University Press, Oxford. (cited on page 1)

Poisson, S. D., 1824. Sur la probabilité des résultats moyens des observations. In *Connaissance des Temps pour l'an 1824*, p. 273–302. (cited on pages 1 and 4)

Pruss, A. R., 2013. Probability, regularity, and cardinality. *Philosophy of Science*, 80, 2 (2013), 231–40. (cited on page 6)

Ramsey, F. P., 1928. A mathematical theory of saving. *The Economic Journal*, 38, 152 (1928), pp. 543–59. (cited on page 1)

Sandberg, A. and Armstrong, S., 2012. Indefinite survival through backup copies. Future of Humanity Institute Technical Report 2012-1. Available at https://www.fhi.ox.ac.uk/reports/2012-1.pdf,. (cited on page 8)

Sidgwick, H., 1907. *The Methods of Ethics, 7th edn.* Macmillan, London. (cited on page 1)

Smith, N., 2014. Is evaluative compositionality a requirement of rationality? *Mind*, 123, 490 (2014), pp. 457–502. (cited on page 24)

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 1 and 20)

Tarsney, C. J. and Wilkinson, H., n.d. Longtermism in an infinite world. In *Essays on Longtermism*. Available at https://globalprioritiesinstitute.org/longtermism-in-an-infinite-world-christian-j-tarsney-and-hayden-wilkinson/. (cited on page 3)

Thoma, J., 2019. Risk aversion and the long run. *Ethics*, 129, 2 (2019), 230–253. (cited on page 20)

Thomas, T., 2022a. The asymmetry, uncertainty, and the long term. *Philosophy and Phenomenological Research*, (2022). (cited on page 20)

Thomas, T., 2022b. Separability and population ethics. *The Oxford Handbook of Population Ethics*, (2022), 271–296. (cited on page 1)

Vilenkin, A., 1983. Birth of inflationary universes. *Physical Review D*, 27, 12 (1983), 2848. (cited on page 8)

von Neumann, J. and Morgenstern, O., 1953. *Theory of Games and Economic Behavior, 2nd edn.* Princeton University Press, Princeton. (cited on page 17)

Wilkinson, H., 2021. *Infinite Aggregation*. Ph.D. thesis, Australian National University. (cited on page 3)

Wilkinson, H., 2022a. In defence of fanaticism. *Ethics*, 132, 2 (2022), p. 445–77. (cited on page 20)

Wilkinson, H., 2022b. Infinite aggregation and risk. *Australasian Journal of Philosophy*, (2022). (cited on page 3)

Wilkinson, H., n.d.a. Can an evidentialist be risk averse? Unpublished manuscript. (cited on page 20)

Wilkinson, H., n.d.b. Flummoxing expectations. Unpublished manuscript. (cited on pages 20, 24, 25, 27, and 28)

Williamson, T., 2000. *Knowledge and Its Limits*. Oxford University Press, Oxford. (cited on pages 5 and 6)

Zhao, M., 2021. Ignore risk; Maximize expected moral value. *Noûs*, (2021). (cited on pages 1 and 20)