Unbounded Utility

Zachary William Lee Goodsell

June 2023

In memory of Chester Goodsell.

Acknowledgements

I would like to thank my committee: Jeff Russell, Andrew Bacon, John Hawthorne, Jacob Nebel, and Isabelle Brocas.

Office hours with Jeff in my first year at USC were a revelation to my and my fellow regular attendees Kin and Siddharth. His example showed me that I would not have to forego my passion for mathematics in pursuing philosophy, and that it could become my greatest strength. Jeff is responsible for sparking and stimulating my interest in the topic of this dissertation. He more than anyone else has helped my ideas on the topic develop.

Andrew has along with Jeff had the greatest impact on my education at USC, having patiently mentored me since my second year on a variety of topics, the most important of which was higher-order logic, which is now the central tool of all my philosophical theorising. Like Jeff, Andrew exhibits a rare combination of technical wherewithal and philosophical sophistication that I can only hope to someday emulate.

John, by contrast, professes to be innocent of any logical or mathematical knowledge. He makes up for this with a preternatural ability to make just the right suggestions to improve any philosophical views he comes across, and to ask just the right questions to reveal any weaknesses (suspiciously, this ability works even on obscure technical philosophy). A discussion where John works his magic is some of the most fun one can have as a philosopher. I would also like to thank John for being an especially steadfast ally in the difficult search for a job.

Despite his youth, Jake has been a font of wisdom, having very deep knowledge and ability in ethics, as well as any other subject we discuss. Feedback from Jake in these last few months has had an especially big impact on the dissertation. I would also like to thank Jake for offering me the opportunity to be his research assistant in this final semester, which has provided great intellectual stimulation and has freed up additional time for getting the dissertation done.

Isabelle has not been so involved in my education, but she has been very diligent in her duties as external committee member, and her perspective as an economist is greatly valued. Thank you!

Juhani Yli-Vakkuri is not on my committee, but I would be remiss not to mention his great philosophical influence on me, brought about through years of collaboration. Juhani: I hope this dissertation makes a contribution to "science". Thanks also for all of those meals and drinks on Smith St and in town.

I would also like to thank my colleagues in the graduate program at USC, who have made my time here so much more enjoyable and intellectually stimulating. I will miss ambushing you or being ambushed with a philosophical question in the courtyard (Frank Hong was particularly common ambushee; he contributed greatly to my interest in the topic of this dissertation), playing Coup at house parties or video games late into the night, getting a drink at The Lab, walking to a local concert, the lavish home-cooked dinners, the racquetball, the fancy tea and coffee, and the good company in general. I am especially grateful for the material, intellectual, and/or moral support provided by my friends Jaime Castillo-Gamboa, David Clark, Brian Haas, Ayala Haddad, Frank Hong, Mahmoud Jalloh, Philip Li, Siddharth Muthukrishnan, Weng Kin San, Levy Wang, Nadja Winning, Matthew Wiseman, and Jin Zeng. I have pestered Kin with sufficiently many bad arguments and fallacious proofs that he deserves particularly special thanks. All the good ideas in this dissertation have been run by him at some point. He has had an enormous influence on my development as a philosopher.

Thanks must also go to the professors at USC not on my committee, for many stimulating discussions and enlightening courses, and also for being so friendly and approachable. Jeremy Goodman in particular has contributed with insightful discussion on Chapter 3. The administrative staff at USC also made a big difference to my time there. I would especially like to thank Natalie Schaad and Nikolai. Prior to USC, I studied at the University of Sydney under Mark Colyvan. He was an excellent supervisor, and to my surprise, the theories proposed in this dissertation turn out to vindicate much of Mark's writing on the same topic. At Sydney I also had the luck to work with Kristie Miller. Collaborating with her gave me the confidence to publish my work.

This dissertation has greatly benefited from discussion with many other people over the years. I am especially grateful for in-depth feedback from the aforementioned Mark Colyvan, as well as Kenny Easwaran, Hayden Wilkinson, and Will Combs. Parts of the dissertation have been presented at a meeting of the American Philosophical Association (Pacific division), as well as at the Global Priorities Institute, the University of Texas at Austin, and New York University. Thank you to all the audience members, and especially to the organisers: respectively Christian Tarsney, Johan Gustafsson, and Snow Zhang. I have also received comments on some chapters from anonymous referees. The comments I have received from *Analysis*, *The Philosophical Review*, *Mind*, and *Noûs* have been very helpful.

My family deserves the most thanks for supporting me throughout my degree and my life. My cousin Samuel Lee is to blame for my obsession with philosophy, when he introduced me to Newcomb's problem around a decade ago. Our discussions, usually interspersed by some sort of racquet sport or lifting (unfortunately only Samuel could claim to be weightlifting), made me realise that philosophy was my true passion. Samuel was the only one to recognise that switching from mathematics to philosophy would be a good idea. My grandparents, Paul and Loani "Wani" Lee, have been a great support, especially during the year I was at Sydney, where Wani's delicious and long-lasting fruitcakes constituted a worryingly large proportion of my diet. My uncle Bob Goodsell has had a very practical impact throughout my life and especially in these last few years. I have had great fun learning how to help Bob install insulation, cut boulders, lay concrete, and other things of that nature. This taste of hard work has made me appreciate philosophy much more.

My mother, Erica, and my brothers, Joseph and Declan, have been the most important people in my life since the death of our father. Joe and Deccy: I have been kept sane by the fun we've had over the years—getting stuck on Fraser Island, cruising in the trimaran or sailboarding with Ben, cooking toasties on the Trangia or something fancier at home, teaming up online or becoming mortal enemies. Thanks also for driving me around (Deccy can I get a lift from the airport please?). Mum: I wouldn't have been able to cope without your unwavering support. Thanks for always welcoming me home, and for all the cakes and cups of tea.

This dissertation is dedicated to my father, Chester, who died in 2020. His curiosity and analytical mindset is the root cause of all my intellectual activity, and his drive to contribute positively to the world has been my greatest inspiration. Dad, I wish you were still here.

Abstract

Some risks are worth taking, and some are not, depending on the possible outcomes and the probabilities of their occurrence. *Decision theory* is the branch of ethics which concerns how probability and value interact to determine whether a risk is worth taking. The hypothesis of *unbounded utility* says that for every probability, no matter how minuscule, there is some outcome so fantastically important that it is worth sacrificing one's life for that minuscule probability of that outcome. Unbounded utility is a consequence of many plausible ethical theories, but causes significant structural difficulties in decision theory. Most notably, the standard view in decision theory, which evaluates risks on the basis of the mathematical expectation of the utility of their outcome, breaks down.

This dissertation develops decision theory in the presence of unbounded utility. The structural problems arising from unbounded utility are used as tools to investigate the fundamental principles of the discipline. I argue in favour of an orthodox approach to decision theory, but show how orthodox principles lead to many surprising consequences in the presence of unbounded utility. Nevertheless, care is taken to establish that the resulting theories are consistent, ensuring that no formal paradox arises.

Contents

List of Theorems

List of Figures

List of Tables

Chapter 1

Unbounded Utility in Decision Theory

1.1 Introduction

Some gambles are good to take, and some are unacceptable, depending on the value of the possible outcomes and the likelihood of them coming about if the gamble is taken. *Decision theory* concerns how value and probability interact to determine the acceptability of a gamble. Since the inception of the discipline (e.g., Bernoulli 1954 [1738]), it has been recognized that gambles with no known finite upper or lower bound on how good their outcome might be pose significant problems in decision theory. This dissertation develops decision theory in the presence of arbitrarily good or bad outcomes. Using the technical notion of utility to measure the goodness or badness of outcomes, this is decision theory where the range of utilities had by outcomes is unbounded.

The difficulty which this dissertation confronts manifests itself in two ways. First, standard approaches to decision theory leave uncomfortable gaps where unbounded utility is concerned. Most saliently, the standard *Expected Utility Theory* (Section 1.2) says disappointingly little when it comes to gambles whose payoffs have unbounded utility. Second, very natural extensions and justifications of Expected Utility Theory—including the classical theories of John von Neumann and Oskar Morgenstern (1944), Leonard Savage (1954), and Richard Jeffrey (1965)—rule out the existence of gambles with payoffs of unbounded utility altogether. Unbounded utility brings in many of the highly counterintuitive properties of the infinite, and it is correspondingly difficult to countenance unbounded utility without contradicting oneself.

These difficulties make it tempting to deny that a gamble's payoffs could be unbounded in utility. There are two natural ways of cashing out this denial. First, one might follow Daniel Bernoulli in positing a finite upper limit on the utility of any possible outcome. This strategy is endorsed by Savage (1954: 76-82), as well as more recently by Barry Nalebuff (1989) and Jeff Russell and Yoaav Isaacs (2021). The second option is to admit that there is no limit on the utility of outcomes, but that no one gamble could have payoffs of unboundedly large utility. This approach is taken by von Neumann and Morgenstern, Jeffrey, and more recently by Frank Hong (forthcoming). Both options allow pleasingly neat pictures of decision theory, but they have their own philosophical costs, which are discussed in Section 1.3. These costs are serious enough that the complications of unbounded utility start to look more palatable.

Many previous authors have agreed with this assessment, and have developed decision theory with unbounded utility in a variety of directions. Some highlights of recent literature on the topic are: Arntzenius, Elga, and Hawthorne 2004, Nover and Hájek 2004, Dietrich and List 2005, Fine 2008, Colyvan 2008, Easwaran 2008, Seidenfeld, Schervish, and Kadane 2009, Hájek and Smithson 2012, Buchak 2013, Gwiazda 2014, Hájek 2014, N. J. J. Smith 2014, Easwaran 2014a, Easwaran 2014b, Lauwers 2016, Lauwers and Vallentyne 2016, Lauwers and Vallentyne 2017, Meacham 2019, and Kosonen 2022. What sets the present work apart from previous literature is that the theories developed are at once *orthodox*, *fundamental*, and *metatheoretically developed*:

Orthodox The theories generally agree with and extend the classical theories of von Neumann and Morgenstern, Savage, and others, as well as with Expected Utility Theory; with small modifications made to accomodate unbounded utility. The attitude motivating this orthodoxy is a recognition that many principles of classical decision theory enjoy powerful motivation that is not clearly inconsistent with unbounded utility, so should be respected as much as possible. By contrast, Fine, Smith, Kosonen, Seidenfeld, Schervish and Kadane, Lauwers and Vallentyne, and Buchak endorse or explore radical failures of orthodox principles as a way to avoid the problems of unbounded utility. Developing an orthodox way of accommodating unbounded utility suggests that these radical theories are unnecessary.

Fundamental The approach taken here is axiomatic; theories are presented by axioms that are as far as possible obvious, or at least very natural and plausible. This is in reaction to much of the literature arising after Nover and Hájek 2004, including Colyvan 2008, Gwiazda 2014, Easwaran 2014b, Meacham 2019. These papers propose extensions of Expected Utility Theory that handle certain prospects of unbounded utility, and which have certain pleasing properties. However, although the proposed extensions are plausibly true, they are not plausibly fundamental: it would be unreasonable to take them as primitive assumptions of decision theory, even if one might reasonably expect some more fundamental axioms to ultimately vindicate them. This literature has generally failed to give produce any decisive arguments for the proposed extensions as a result.

Metatheoretically developed Any sufficiently developed theory accommodating unbounded utility will have some counterintuitive consequences; an inevitable feature of reasoning about the infinite. For this reason, it is important to thoroughly understand the consequences of a proposed theory before it is accepted. Here, care is taken to establish the consistency of any proposed theory, as well as intertheoretic connections such as independence or interderiveability between theories. Less radical contributors to the literature, such as Arntzenius, Elga, and Hawthorne, Dietrich and List, and Easwaran (2014a) will find many of their recommendations echoed in this dissertation. The intention is to improve our understanding of what these theories entail.

1.2 Utility and Expected Utility

Utility is a central concept of decision theory, and one that will be appealed to throughout this dissertation. It is a measure of how much a chance of an outcome affects the value of a prospect that gives that outcome with that chance. The utility of an outcome is a real number, chosen so that utility is proportional to the contribution a chance of that outcome instead of an arbitrarily chosen "zero point" contributes to overall value. For example, an outcome of utility 6 is twice as important as an outcome of utility 3, in the sense that a $1/4$ chance of the former makes an equally great contribution as a $1/2$ chance of the latter (making the arbitrary zero point explicit: $1/4$ chance of utility 6 and $3/4$ chance of utility 0 should be equally good as $1/2$ chance of utility 3 and $1/2$ chance of utility 0).

A central assumption of Expected Utility Theory, as well as the classical theories of Savage and von Neumann and Morgenstern, is that different possible outcomes make *independent* contributions to the overall evaluation of a prospect. Suppose a prospect has a $1/4$ chance of utility 4, a $1/2$ chance of utility 2, and a $1/4$ chance of utility -4. The $1/4$ chance of utility 4 and the $1/2$ chance of utility 2 would each contribute as much as a 100% chance of utility 1 on their own, and the $1/4$ chance of utility -4 would contribute -1 on its own; so by the classical hypothesis of independence the prospect is overall equally good as a 100% chance of utility 0.

Simple Expected Utility Theory says that when a prospect can yield outcomes o_1 through o_n with respective probabilities x_1 through x_n, then the value of that

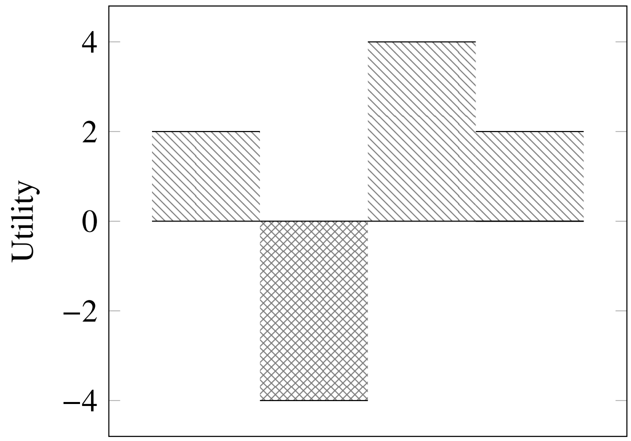


Figure 1.1: Geometric representation of a prospect which has $1/2$ chance of utility 2, and $1/4$ chance each of utility 4 and -4. The expected utility is proportional to the area of the northwest-hatched part minus the area of the crosshatched part

prospect is computed by adding together the probability-weighted utilities of the outcomes. Where U, the *utility function*, maps an outcome to its utility, this sum is expressed as

$$x_1 U(o_1) + \cdots + x_n U(o_n).$$

This sum is the *expected utility* of the prospect; and Simple Expected Utility Theory says that it is equally good as the outcome whose utility is equal to the prospect's expected utility. Simple Expected Utility Theory is the combination of the two aforementioned ideas: the definition of utility as a measure of the contribution a chance of an outcome makes to overall value, and the independence of the different contributions made by different possible outcomes. For this reason, Simple Expected Utility Theory will be included in almost all theories considered in this dissertation.

Suppose that a certain line is to have one of its points selected by a fair random process, and that all prospects have their outcome determined by the result of this process. A prospect may be specified by specifying for each point on the line what utility would result if that prospect where taken and that point were randomly selected. The prospect from earlier can now be graphically represented as in Figure 1.1, where its overall value is proportional to the northwest-hatched area minus the crosshatched area. In general, when a prospect represented this way is the union of finitely many rectangles, Simple Expected Utility Theory says that its overall value is proportional to the northwest-hatched area minus the crosshatched area.

Expected Utility Theory proper says that the overall value of a prospect is always proportional to the northwest-hatched area minus the crosshatched area, even when

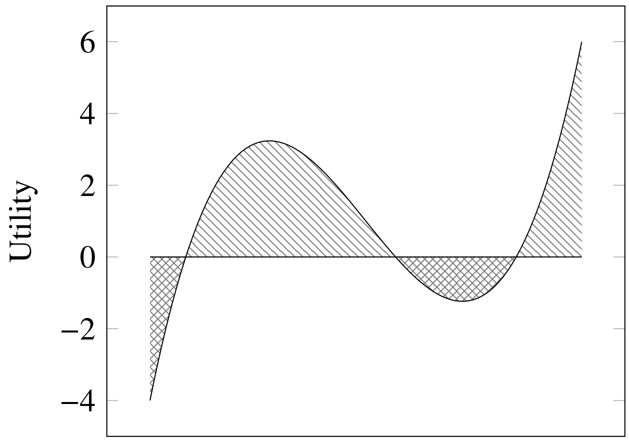


Figure 1.2: Geometric representation of a prospect which has probability zero of any particular outcome. The expected utility is proportional to the north-west hatched area minus the crosshatched area.

the prospect is not the union of finitely many rectangles. For example, the prospect represented in Figure 1.2 has probability zero of any particular utility, because the utility of the outcome varies continuously with the point selected from the line. According to Expected Utility Theory its value is to be computed by subtracting the crosshatched area from the northwest-hatched area, so it is better than the prospect in Figure 1.1 if and only if that subtraction of areas is greater.

Expected Utility Theory is a natural strengthening of Simple Expected Utility Theory, but it importantly goes beyond the two central ideas concerning the definition of utility and the independence of the contribution of different possible outcomes. For this reason, the dissertation will remain neutral on Expected Utility Theory (some considerations against are raised in Chapter 2).

1.3 Unbounded Utility

1.3.1 What it is

For utility to be unbounded is for there to be no finite positive or negative bound on the utilities had by outcomes. This is to say: for any probability ε, no matter how small, there is some outcome with a finite utility for which an ε chance of that outcome outweighs a 100% chance of utility 1. Geometrically, this means that it is possible to for the graph of a prospect as in Figure 1.1 to be infinitely tall.

The unboundedness of utility should be carefully distinguished from outcomes of infinite utility. An outcome might be said to have positive infinite utility if, for any positive probability ε, an ε chance of that outcome outweighs a 100% chance of utility 1. The difference is in the order of the quantifiers: utility can be unbounded when every outcome has a finite utility, because for successively smaller choices of ε we choose successively more important outcomes, whereas for an outcome of infinite utility one outcome is chosen for which any probability of it outweighs the 100% chance of utility 1. Throughout the dissertation it will typically be assumed that every outcome has a finite utility, which is to say that for every outcome o there is a small enough probability ε that an ε chance of o makes a smaller contribution than a 100% chance of utility 1. This should be thought of as a simplifying assumption; by taking it we are ignoring any outcomes of infinite utility there might be, rather than committing ourselves to their nonexistence.

Even if utility is unbounded in this sense, there might be gaps in utility. For example, there might be no outcome of utility 3, which would mean that there is no outcome for which a $1/3$ chance of that outcome is equally important as a 100% chance of utility 1. This possibility will generally be ignored in this dissertation: the unboundedness of utility will be identified with the claim that *every* real number, positive, negative, or zero, is such that some outcome has that number as its utility.

1.3.2 Arguments for Unbounded Utility

Whether utility is unbounded is an important ethical question. To understand its consequences, suppose that utility is unbounded and consider any two outcomes. For concreteness, let the first outcome be one where humanity flourishes, and the second be one where we are all wiped out by an asteroid. We may suppose the first outcome to have utility 1 and the second to have utility 0 (because the units and zero point of the utility scale are arbitrary). For utility to be unbounded, then, is for there to be outcomes such that for any probability ε, a 100% chance of humanity flourishing is worth trading for an ε chance of that outcome and a $1 - \varepsilon$ chance of an asteroid wiping us all out. These are outcomes so important that the survival of humanity is almost completely negligible in importance comparatively. When ε is very small, it is not at all obvious that there could be such outcomes, and it may even seem implausible that there would be any. Nevertheless, this section argues that there are outcomes of unbounded utility. The technical problems arising from unbounded utility become especially pressing in light of these arguments; these are the problems that this dissertation undertakes to resolve.

A typical way of arguing for unbounded utility correlates utility with a quantity for which there is no physically necessary upper or lower bound on the amount of that quantity. For early decision theorists working on the theory of gambling, the size of one's bank account often played this role, with it being assumed, in effect, that an outcome where you have twice as much money has twice the utility. As D.

Bernoulli pointed out, this assumption is implausible:

> There is no doubt that a gain of one thousand ducats is more significant to a pauper than to a rich man though both gain the same amount. (Bernoulli 1954: 24)

It should also be remembered that money is not an end in itself. The "outcome" of receiving a thousand ducats is itself risky, depending on whether those ducats are stolen, squandered, spent wisely, etc.

Nevertheless, considerations in ethics frequently converge upon two plausibly unbounded quantities—population size and lifespan—being linearly correlated with utility.

According to *Expected Total Utilitarianism*, the utility of an outcome is given by the sum total of welfare in that situation, where welfare is measured by a real number. Supposing that a happy life has welfare 1, the Expected Total Utilitarian will say that two happy lives have total welfare 2, three have total welfare 3, and so on. Supposing that there could be any finite number of people, while everyone has a happy life, utility is unbounded given Expected Total Utilitarianism.

Expected Total Utilitarians generally also think that the welfare of a life that remains equally happy throughout depends linearly on its duration. 100 happy years results in twice as much welfare as 50 happy years. Supposing that there could be just one person with a happy life that could last any finite amount of time, utility would again be unbounded.

Expected Total Utilitarianism is a controversial view in ethics. However, the full strength of the the theory is not needed to argue for unbounded utility. For example, if utility increases only with the square root of the number of happy lives, so that 100 times as many happy lives is worth only 10 times the utility, utility is still unbounded if the possible number of happy lives is. So long as the function converting number of happy lives into utility is unbounded in the real numbers, and there could be any number of lives while everybody is happy, utility is unbounded. Similarly for the function that converts the duration of a single happy life into the utility contributed to the outcome.

There are a range of arguments that make it plausible, independently of Expected Total Utilitarianism, that the number or duration of happy life increases the utility of the outcome without bounds. Here we will repeat an argument due to Vann McGee (1999), which is also used by Beckstead and Thomas 2021, and was presented to me by Caspar Hare.

The argument is as follows. Suppose that the utility contribution of the number of happy lives is bounded (McGee goes via the duration of happy life instead; which can be subsituted for number of happy lives in the following with no significant difference to the effectiveness of the argument). Then there is some n such that n

happy lives is near the upper bound of utility that can be contributed by happy lives. Consider four outcomes:

w_0 A world with a solitary life, which is happy.

w_0^+ A world with a solitary life like w_0, but where the person is little bit happier because they receive one extra chocolate bar.

w_n A world where $100 \times n$ people are born, but all but n of them die painlessly as babies, with the survivors all having happy lives.

w_{100n} A world like w_n but where all $100 \times n$ people are saved, and all have happy lives.

Let us set the zero point of the utility scale at w_0, and the unit at w_0^+. And let us suppose that n is large enough that the utility of w_n is within some very small number $\varepsilon \ll 1$ of the limit of utility that can be contributed by happy lives. Now consider two prospects:

Chocolate Fifty-fifty between w_0^+ and w_n.

No Chocolate Fifty-fifty between w_0 and w_{100n}.

According to Simple Expected Utility Theory, Chocolate is the better prospect. This is because, in the case where there is one person, Chocolate gains a net 1/2 a point of utility, and in the case where there are lots of people, No Chocolate only gains at most $\varepsilon/2$ points of utility, which is much less.

McGee, Hare, Beckstead, Thomas, and I agree that this judgement is ridiculous: No Chocolate is obviously better. It is obviously better because it trades a trivial benefit—a chance of a chocolate bar—for an equal probability of an incredibly weighty one—$99n$ lives saved. To say that utility is bounded requires us to say that $99n$ lives saved is not very weighty, and is indeed as trivial as a chocolate bar, when sufficiently many other people have already been saved. This is what we find difficult to believe.

A similar point can be made in a range of ways. Suppose that $100n$ people are to ever live, and they will either painlessly die young or go on to live happy lives depending on your decision. Surely, for some number k much less than 100 but at least 2, saving one person for sure is worse than a $1/2$ chance of saving k people. Now suppose that we divide kn of the people into n groups of size k, and we consider two options: saving one person from each group for sure, or a $1/2$ chance of saving everybody. Since the utility of saving n people is assumed to be near the limit of utility of number of lives saved, the first option must be better. But looking at the people group-by-group, we would prefer to sacrifice the one for the chance of

Table 1.1: The St Petersburg prospect.

$1/2$	$1/4$	$\dots$	$1/2^n$	$\dots$
2	4	$\dots$	2^n	$\dots$

Table 1.2: The St Petersburg prospect sweetened.

$1/2$	$1/4$	$\dots$	$1/2^n$	$\dots$
3	5	$\dots$	$2^n + 1$	$\dots$

Table 1.3: The alternating St Petersburg prospect.

$1/2$	$1/4$	$\dots$	$1/2^n$	$\dots$
-2	4	$\dots$	$(-2)^n$	$\dots$

Table 1.4: The Pasadena prospect.

$1/2$	$1/4$	$\dots$	$1/2^n$	$\dots$
2	$-4/2$	$\dots$	$-(-2)^n/n$	$\dots$

Table 1.5: The symmetrical St Petersburg prospect.

$1/4$	$1/4$	$\dots$	$1/2^{n+1}$	$1/2^{n+1}$	$\dots$
2	-2	$\dots$	2^n	$-(2^n)$	$\dots$

saving the k. So having decided that it is right to sacrifice the one for a chance of saving the k, we cannot infer that the policy of sacrificing one for a half chance of saving k is right in general. But again it seems bizarre to think that once n people are saved, saving k times as many people is all of a sudden not very important.

1.3.3 Two Puzzles

Arguments for unbounded utility can easily be multiplied. This section presents two of many structural problems for the hypothesis of unbounded utility. These problems threaten to show that unbounded utility is not compatible with a plausible decision theory. Given the previous arguments for unbounded utility, this would be an uncomfortable discovery. The point of this dissertation is to establish that this is not so, by developing decision theory in the presence of unbounded utility.

Divergent and Undefined Expectations

If the possible outcomes of a prospect can be unbounded in utility, then that prospect need not have a well-defined expected utility. Consider the *St Petersburg prospect*, attributed to Nicolaus Bernoulli in Bernoulli 1954 [1738]. For this prospect, a coin is fairly and independendently tossed until it lands tails for the first time, and the prize is determined by how many tosses it takes for this to happen. If it lands tails right away, you get utility 2, if it lands heads then tails, you get utility 4, if it lands heads heads tails you get utility 8, and in general if it takes n tosses to get tails for the first time you get utility 2^n. Since the coin is fair and independent, the probability of n tosses is $1/2^n$, so the expected utility would be given by the sum

$$\begin{aligned}&\frac{1}{2} \times 2 + \frac{1}{4} \times 4 + \cdots + \frac{1}{2^n} \times 2^n + \dots \\ =&1 + 1 + \cdots + 1 + \dots\end{aligned}$$

which diverges to infinity.

Consider also the variants in Tables 1.2 through 1.5. Trying to compute the expected utility of these prospects variously yields divergence to infinity (for the sweetened St Petersburg) or divergence due to the partial sums oscillating, or mere conditional convergence (in the case of Pasadena—see Nover and Hájek 2004). In no case is there a well-defined real-valued expected utility, so Expected Utility Theory does not on its own say how to compare these prospects.

In the case of the St Petersburg and sweetened St Petersburg prospects, we may assign a non-real expected "utility", ∞, to both. Such a value keeps track of the fact that they are both better than any finite utility, but is of limited use, because it does not discriminate between the two prospects when the sweetened St Petersburg is clearly better, and there is in any case no similar value to assign to the alternating and symmetrical St Petersburg prospects. Expected Utility Theory, then, is of little help in evaluating these prospects.

Some comparisons between the prospects are obvious: sweetened St Petersburg is better than plain St Petersburg, which is better than the rest. But it is unclear whether alternating St Petersburg would be better or worse than the Pasadena prospect or the symmetrical St Petersburg prospect, and it is unclear also how to compare these latter three prospects to finite levels of utility. The first puzzle, then, is that acknowledging unbounded utility seems to render standard tools of decision useless in many cases. The puzzle is to be solved by establishing which well-motivated decision-theoretic assumptions would yield verdicts on these cases and others, and by figuring out when those assumptions leave a comparison undecided.

Inconsistencies

The second puzzle is that many very natural decision-theoretic assumptions turn out to be inconsistent with unbounded utility. For example, it is natural to suppose that the value of a prospect must be in between the values of the possible outcomes it might yield. This is Savage's axiom P7 (1954: 77), and as Savage recognizes, it conflicts with unbounded utility. The St Petersburg prospect is a witness to this, because it is better than every one of its possible outcomes. This is a strange situation: taking the St Petersburg prospect leads to certain disappointment.

Suppose there are two St Petersburg-like prospects available, whose results are determined by independent sequences of coin tosses, and where one is sweetened by an additional point of utility in every outcome. The sweetened prospect is clearly better, even though the unsweetened prospect is better than any outcome the sweetened one could yield. As a result, if you take the sweetened prospect and then learn its outcome, you would prefer to switch to the unsweetened prospect. We end up in the shocking situation where one prospect is better than another even though the worse prospect is better conditional on any outcome of the better prospect.

This bizarre situation has been discussed at length in the literature following Barry Nalebuff's *two-envelope paradox*. Arntzenius, Elga, and Hawthorne (2004) ameliorate this strange consequence of unbounded utility somewhat, but the main problem in the background is that inconsistency and paradox lurk very close when theorising about unbounded utility. Supposing that we can get used to the pathologies just mentioned, it remains to be seen whether there is a satisfactory decision theory to be had that accomodates unbounded utility, or whether any reasonably strong theory collapses back into inconsistency. Some further inconsistencies will be explored throughout the dissertation, and care will be taken to ensure that the theories proposed are consistent.

1.3.4 Infinite Possibilities

If every possible prospect has only finitely many possible outcomes, then every prospect will have a well-defined expected utility if there are outcomes of arbitrarily great but finite utility. More generally, no very significant problems can arise if every prospect has a well-defined expected utility, since Expected Utility Theory is very well-behaved in the cases where it applies.

Jeffrey (1965) is an early proponent of this strategy for avoiding the problems just mentioned:

> Put briefly and crudely, our rebuttal of the St. Petersburg paradox consists in the remark that anyone who offers to let the agent play the St. Petersburg game is a liar, for he is pretending to have an indefinitely large bank. (Jeffrey 1965: 154)

Of course it is not really money that is at issue but utility. Jeffrey acknowledges that happy life-years might vary linearly with utility (p. 155, Example 7), but the response is the same: nobody has the ability to guarantee happy life years as the certain result of coin tosses, so the prospect cannot arise. Michael Huemer (2016: 213-18) takes a very similar line.

Jeffrey's argument misses the mark. To face a St Petersburg prospect, or other prospect without a well-defined expectation, there need be no guarantee that the bookmaker honors the bet. Rather, the probability that they honor the bet can tend towards zero, conditional on the prize owed, as the owed prize gets larger. For the expected utility to be undefined only requires the probabilities to decay sufficiently slowly. Jeffrey does not provide any reason to think that they do.

Suppose, as was suggested earlier, that utility grows unboundedly with the number of happy lives in the world. The number of happy lives is, from a physical perspective, not dissimilar to other quantities, like the total mass of gold, or the number of stars. For Jeffrey's argument to work requires us to have a well-defined expectation for the number of happy lives there will be independently of our intervention, but without an argument that is specific to happy lives (which Jeffrey does not supply), it seems Jeffrey would be equally well-advised to hope for a well-defined expectation for the mass of gold ever in the universe. It seems rather optimistic to reach such physical conclusions on the basis of decision theory.

Even though Jeffrey's argument is inadequate, the question of whether our actions could have consequences without well-defined expectation for the amount of utility or gold they produce is a difficult and interesting one. Thought experiments where coin tosses determine the fate of unboundedly many lives are only very plausible given outlandish theological hypotheses. If the theological hypotheses have nonzero probability independently of the outcomes of the coin tosses, then we are off to the races. Otherwise, one would expect a sophisticated application of physics would yield some insight. It seems plausible to a layperson such as myself that our actions could have ill-defined expected utility from a physical perspective, but the important thing for a decision theorist is not hitching the hopes of decision theory on the difficult physics and philosophy that would be needed to prove that our actions always have well-defined expected utility.

Another strategy that philosophers have tried in order to avoid ill-defined expected utility is to make surprising claims about probability that are almost independent of what physics or theology might suggest. This is the strategy of Martin Smith (forthcoming) and Frank Hong (forthcoming), who make the following two claims:

(a) Everything some given agent knows has probability 1.

(b) That agent knows an upper and lower limit on the utility that might result from any available prospect.

As a result: every prospect gives probability 1 to the utility of the outcome lying in some bounded range, so unbounded prospects like the St Petersburg prospect are ruled out.

To apply these principle to the St Petersburg prospect more directly: these authors would plausibly claim that before the coin is tossed, you know it won't land heads 100 times in a row; which is to say you know the outcome of the prospect won't have utility more than 2^{100}. By principle (a), the probability of utility 2^{100} or more is 0 exactly, and not 2^{-100} as it would be if the coins were probabilistically fair and independent.

I take issue with both principles. Once we are in the frame of mind where we ascribe advance knowledge about coin tosses to ordinary people, it becomes considerably less plausible that ethically relevant betting probabilities correspond to knowledge as principle (a) says. Surely even if you know in advance that the coin won't land heads 100 times in a row, you still get to know it better, with greater certainty, after you see tails on the third toss. And if there is such a thing as "certain" knowledge over and above ordinary knowledge, it seems implausible that ordinary knowledge would imply probability 1.

Furthermore, imagine betting on the landing place of an apparently fairly thrown dart on a dartboard. Either there is a set of finitely many points that it will land on with probability 1, or there are infinitely many disjoint subregions of the dartboard that each have nonzero probability. Even once we are used to the idea that we know a bit about coin tosses before they happen, the former possibility seems outlandish, and decision theory would implausibly recommend risking everything for a chocolate bar if the dart lands on one of those finitely many points. But if there are infinitely many disjoint subregions of nonzero probability, then we may enumerate them as E_1, E_2, etc., and imagine a gamble which gives utility $1/\mathbf{P}(E_n)$ if the dart lands on E_n, whose expectation will be undefined (McGee 1999 shows that this is sufficient to get a two-envelope paradox going). Hong and Smith would have to show that it is impossible for this prospect to be available when $E_1, E_2 \ldots$ get those probabilities. But I see no reason to suppose that your advance knowledge of where a dart is going to land would correlate so nicely with the facts about how what happens next depends on where the dart lands.

1.4 Dissertation Summary

1.4.1 Chapter 2: Decision Theory Unbound

A minimal set of decision-theoretic axioms are presented from which the two-envelope paradox may be derived. The two most plausible culprits are found to be the assumption of unbounded utility and the *Countable Sure-Thing* principle, labelled thus by Russell and Isaacs (2021), and called *Eventwise Dominance* by Dietrich and List (2005). Countable Sure-Thing corresponds to a plausible diachronic coherence principle, which as a result of the two-envelope paradox must surprisingly be rejected if unbounded utility exists.

The chapter explores the question of what happens to decision theory in the absence of Countable Sure-Thing. A theory without Countable Sure-Thing, called DTU, is proposed in broad agreement with Dietrich and List. The notable principles of DTU include the standard Sure-Thing principle, as well as a stochastic dominance axiom as well as a principle saying that there is no incomparability among prospects (that is: one is always better than the other, if they are not equally good). The main technical contribution of the chapter is to show that DTU is consistent, establishing that no further paradox can arise without going beyond these strong and standard principles.

The chapter also includes commentary on potential extensions of DTU. DTU includes Simple Expected Utility Theory (see Section 1.2), but it is shown that full Expected Utility Theory is independent of DTU. It is argued that Expected Utility Theory is not clearly in good standing, since the most natural argument for the principle would appeal to Countable Sure-Thing. The other potential extension mentioned in this chapter consists of the symmetry principles addressed at length in Chapter 4; it is shown in Chapter 2 that these principles do *not* follow from DTU, so are genuine extensions of it.

1.4.2 Chapter 3: Adding Lotteries

A major competitor to the approach taken with DTU in Chapter 2 is due to Seidenfeld, Schervish, and Kadane (2009), who build on an idea in Colyvan 2008. Seidenfeld, Schervish, and Kadane endorse a principle, which they call *Coherent Indifference*, which roughly speaking says that if prospects X and Y are equally good, then it is indifferent to gain X and then lose Y. They show that Coherent Indifference is incompatible with one of the core assumptions of DTU, called Stochastic Equivalence, which says that prospects that give the same probability distribution over outcomes are equally good.

This chapter compares the relative merits of the two approaches to decision theory. It is shown that a very natural strengthening of Coherent Indifference, here called *Addition Invariance*, has very desirable unifying consequences in decision theory, allowing various principles that are primitive in DTU or independent of DTU to be derived in a very minimal theory containing Addition Invariance. Moreover, some potential objections to Addition Invariance concerning the good-standing of the notion of gaining a prospect and then losing another prospect are defused.

Seidenfeld, Schervish, and Kadane's proof of the incompatibility with Stochastic

Equivalence is then rehearsed, and the chapter turns to emphasising the costs of giving up Stochastic Equivalence. It is found that despite the pleasing features of Addition Invariance, this cost is likely too great to bear.

1.4.3 Chapter 4: Symmetries of Value

When a prospect has a well-defined expected utility, doubling the utility of each outcome doubles its expected utility, meaning that if you take any two such prospects X and Y with X better than Y, then double X is better than double Y. This suggests that this principle should hold in general, regardless of whether X and Y have expected utilities; principles of this sort are the topic of Chapter 4. The chapter begins by precisifying exactly how and why the principle ought to generalize, building on a very brief argument of Hájek (2014). This investigation suggests a pair of principles together called the *affine symmetry principles*

Attention is then turned to the consequences of the affine symmetry principles in decision theory. In addition to some desirable structural consequences, they give rise to some fairly shocking results concerning the exact value of some particular prospects.

These surprising results make it especially important to establish the consistency of the principles, which is the next task of the chapter. A class of models for the theory is developed to establish consistency, as well as some salient independence results. The chapter concludes by discussing how the affine symmetry principles must be modified once we relax the assumption that every outcome has a real-valued utility.

1.4.4 Chapter 5: A St Petersburg Paradox for Risky Welfare Aggregation

This chapter concerns the principle of *Anteriority*, which says that if two prospects are stochastically identical from the point of view of each possible person's welfare, then they are equally good. Anteriority is a weakening of the *Ex Ante Pareto* principle, which gives rise to an influential argument for Expected Total Utilitarianism originally due to John Harsanyi (1955). More modestly, principles like Ex Ante Pareto have also been appealed to in arguing for unbounded utility in ethics.

The main result of the paper is that Anteriority is incompatible with very minimal decision-theoretic principles if there are infinitely many possible people. Despite its plausibility, Anteriority plays a very similar role to the Countable Sure-Thing principle, and should similarly be rejected.

Two lessons are drawn from this paradoxical result. The first is that we must be very careful in deploying principles like Ex Ante Pareto in favour of unbounded utility or Expected Total Utilitarianism, since these principles are false in full generality. As Jeff Russell has put the point (using 'Fanaticism' for the claim that the ethical utility scale is unbounded):

> Arguments for Fanaticism based on these modified and truncated principles don't have nearly the same immediate grip as the ones involving the sweeping clean versions. So even though the fanatical position is still there, it no longer stands out as the bold, austere, and systematic ethical framework that it once seemed. (Russell 2021: 29)

The second lesson, is the observation that certain arguments deployed in population ethics are in bad company, most famously Parfit's (1984) 'Egyptology' argument: it is shown that if Parfit's argument, if any good, would justify Anteriority, meaning that the argument cannot be any good because Anteriority is false.

These results put some pressure on the hypothesis of unbounded utility, since some of the most natural arguments for it fall through. The arguments in Section 1.3 have been carefully constructed to avoid relying on Anteriority.

Chapter 5 has been published previously in *Analysis*, with the exception of the short technical postscript.

Bibliography

Arntzenius, Frank, Adam Elga, and John Hawthorne (2004). "Bayesianism, Infinite Decisions, and Binding". In: *Mind* 113.450, pp. 251–283.

Arntzenius, Frank and David McCarthy (1997). "The Two Envelope Paradox and Infinite Expectations". In: *Analysis* 57.1, pp. 42–50.

Askell, Amanda (2018). "Pareto Principles in Infinite Ethics". PhD thesis. New York University.

Bales, Adam, Daniel Cohen, and Toby Handfield (2014). "Decision Theory for Agents with Incomplete Preferences". In: *Australasian Journal of Philosophy* 92.3, pp. 453–70.

Bartha, Paul (2016). "Making Do Without Expectations". In: *Mind* 125.499, pp. 799–827.

Beckstead, Nick and Teruji Thomas (2021). "A Paradox for Tiny Probabilities and Enormous Values". In: *GPI Working Papers* 7.

Bernoulli, Daniel (1954). "Exposition of a New Theory on the Measurement of Risk". In: *Econometrica* 22.1, pp. 23–36.

Blackwell, David A and Meyer A Girshick (1954). *Theory of games and statistical decisions*. Wiley.

— (1979). *Theory of games and statistical decisions*. Courier Corporation. *Theory of games and statistical decisions*. Wiley, 1954.

Bolker, Ethan D. (1967). "A Simultaneous Axiomatization of Utility and Subjective Probability". In: *Philosophy of Science* 34.4, pp. 333–340.

Bostrom, Nick (2011). "Infinite Ethics". In: *Analysis and Metaphysics* 10, pp. 9–59.

Broome, John (1995). "The Two-Envelope Paradox". In: *Analysis* 55.1, pp. 6–11.

Buchak, Lara (2013). *Risk and Rationality*. Oxford University Press.

Chalmers, David (2002). "The St. Petersburg Two-Envelope Paradox". In: *Analysis* 62.2, pp. 155–157.

Colyvan, Mark (2008). "Relative Expectation Theory". In: *Journal of Philosophy* 105.1, pp. 37–44.

Colyvan, Mark and Alan Hájek (2016). "Making Ado Without Expectations". In: *Mind* 125.499, pp. 829–857.

de Finetti, Bruno (1972). *Probability, Induction, and Statistics*. New York: John Wiley.

Dietrich, Franz and Christian List (2005). "The Two-Envelope Paradox: An Axiomatic Approach". In: *Mind* 114.454, pp. 239–248.

Easwaran, Kenny (2008). "Strong and Weak Expectations". In: *Mind* 117.467, pp. 633–641.

Easwaran, Kenny (2014a). “Decision Theory Without Representation Theorems”. In: *Philosophers’ Imprint* 14.

— (2014b). “Principal Values and Weak Expectations”. In: *Mind* 123.490, pp. 517–531.

Fine, Terrence (2008). “Evaluating the Pasadena, Altadena, and St Petersburg Gambles”. In: *Mind* 117.467, pp. 613–632.

Fleurbaey, Marc and Alex Voorhoeve (2013). “Decide As You Would With Full Information! An Argument Against Ex Ante Pareto”. In: *Inequalities in Health: Concepts, Measures, and Ethics*. Ed. by Ole Norheim et al. Oxford University Press.

Gibbard, Allan and William L. Harper (1978). “Counterfactuals and Two Kinds of Expected Utility”. In: *Foundations and Applications of Decision Theory*. Ed. by A. Hooker, J. J. Leach, and E. F. McClennen. D. Reidel, pp. 125–162.

Goodsell, Zachary (2022). “Tossing Morgenbesser’s Coin”. In: *Analysis* 82.2, pp. 214–221.

Gwiazda, Jeremy (2014). “Orderly Expectations”. In: *Mind* 123.490, pp. 503–516.

Hájek, Alan (2014). “Unexpected Expectations”. In: *Mind* 123.490, pp. 533–567.

Hájek, Alan and Michael Smithson (2012). “Rationality and Indeterminate Probabilities”. In: *Synthese* 187.1, pp. 33–48.

Hare, Caspar (2010). “Take the Sugar”. In: *Analysis* 70.2, pp. 237–247.

Harsanyi, John C. (1955). “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility”. In: *Journal of Political Economy* 63.4, pp. 309–321.

Hong, Frank (forthcoming). “Know Your Way Out of St. Petersburg: An Exploration of "Knowledge-First" Decision Theory”. In: *Erkenntnis*, pp. 1–20.

Huemer, Michael (2016). *Approaching Infinity*. New York: Palgrave Macmillan.

Jech, Thomas (2006). *Set Theory: The Third Millennium Edition, Revised and Expanded*. Springer Monographs in Mathematics. Springer Berlin Heidelberg.

Jeffrey, Richard (1965). *The Logic of Decision*. University of Chicago Press.

Joyce, James M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Kaplan, Mark (1983). “Decision Theory as Philosophy”. In: *Philosophy of Science* 50.4, pp. 549–577.

Kosonen, Petra (2022). “Tiny Probabilities of Vast Value”. PhD thesis.

Lauwers, Luc (2016). “Why Decision Theory Remains Constructively Incomplete”. In: *Mind* 125.500, pp. 1033–1043.

Lauwers, Luc and Peter Vallentyne (2016). “Decision Theory Without Finite Standard Expected Value”. In: *Economics and Philosophy* 32.3, pp. 383–407.

— (2017). “A Tree Can Make a Difference”. In: *Journal of Philosophy* 114.1, pp. 33–42.

Lenman, James (2000). “Consequentialism and Cluelessness”. In: *Philosophy and Public Affairs* 29.4, pp. 342–370.

Levi, Isaac (1990). *Hard Choices: Decision Making Under Unresolved Conflict*. Cambridge University Press.

Lewis, David (1981). “Causal Decision Theory”. In: *Australasian Journal of Philosophy* 59.1, pp. 5–30.

McCarthy, David, Kalle Mikkola, and Teruji Thomas (2020). "Utilitarianism with and Without Expected Utility". In: *Journal of Mathematical Economics* 87, pp. 77–113.

McGee, Vann (1999). "An Airtight Dutch Book". In: *Analysis* 59.4, pp. 257–265.

Meacham, Christopher J. G. (2019). "Difference Minimizing Theory". In: *Ergo: An Open Access Journal of Philosophy* 6.

Nalebuff, Barry (1989). "Puzzles: The Other Person's Envelope is Always Greener". In: *The Journal of Economic Perspectives* 3.1, pp. 171–181.

Nover, Harris and Alan Hájek (2004). "Vexing Expectations". In: *Mind* 113.450, pp. 237–249.

Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press.

— (1997). "Equality and Priority". In: *Ratio* 10.3, pp. 202–221.

Quirk, James P and Rubin Saposnik (1962). "Admissibility and measurable utility functions". In: *The Review of Economic Studies* 29.2, pp. 140–146.

Russell, Jeffrey Sanford (2020). "Non-Archimedean Preferences Over Countable Lotteries". In: *Journal of Mathematical Economics* 88.May 2020, pp. 180–186.

— (2021). "On Two Arguments for Fanaticism". In: *GPI Working Papers* 17.

Russell, Jeffrey Sanford and Yoaav Isaacs (2021). "Infinite Prospects". In: *Philosophy and Phenomenological Research* 103.1, pp. 178–198.

Savage, Leonard (1954). *The Foundations of Statistics*. Wiley Publications in Statistics.

Schoenfield, Miriam (2014). "Decision Making in the Face of Parity". In: *Philosophical Perspectives* 28.1, pp. 263–277.

Seidenfeld, Teddy, Mark Schervish, and Joseph Kadane (2009). "Preference for Equivalent Random Variables: A Price for Unbounded Utilities". In: *Journal of Mathematical Economics* 45, pp. 329–340.

Smith, Martin (forthcoming). "Decision Theory and de Minimis Risk". In: *Erkenntnis*, pp. 1–24.

Smith, Nicholas J. J. (2014). "Is Evaluative Compositionality a Requirement of Rationality?" In: *Mind* 123.490, pp. 457–502.

Struik, D.J. (1953). *Lectures on Analytic and Projective Geometry*. Addison-Wesley mathematics series. Addison-Wesley Publishing Company.

Vallentyne, Peter and Shelly Kagan (1997). "Infinite Value and Finitely Additive Value Theory". In: *Journal of Philosophy* 94.1, pp. 5–26.

von Neumann, John and Oskar Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Wilkinson, Hayden (2022). "In Defense of Fanaticism". In: *Ethics* 132.2, pp. 445–477.

Chapter 2

Decision Theory Unbound

2.1 The Paradox

The *utility* of a possible situation is a measure of how much a chance of yielding that situation contributes to the overall value of a gamble. Choosing arbitrary situations to be at the zero point and unit of the utility scale, we say that a possible situation has utility 2 if a fifty-fifty gamble between that situation and the zero point is equally good as the unit point: fifty-fifty between 2 and 0 is equal to a sure 1. More generally, a situation is defined as having utility x, for x a number greater than 1, if a prospect that yields that situation with probability $1/x$ and yields utility 0 otherwise is equally good as a sure utility 1. If for every number some possible outcome has a utility at least as great as that number, we say that utility is *unbounded*.

If you read "good" above as "morally good", then whether utility is unbounded in that sense is a substantial ethical question. For example, one might think that it is good to create people whose lives will be wonderful, and so one might wonder whether a $1/n$ chance of creating n such people is equally good as creating exactly one such person for sure. For this to be so requires utility to be unbounded. If you read "good" in some other way, such as "good for me" or "good for the sake of winning this chess game", then whether utility is unbounded in the concommitant sense is a different question, but an equally substantial one in each case. Most of what follows will be true on many reasonable disambiguations, but where ambiguity might be problematic we will default to the moral reading.

The problem motivating this paper is that there are significant structural problems for the thesis of unbounded utility, which threaten to refute the thesis before the substantial ethics can even begin. One widely recognised problem is that, when utility is unbounded, the *expected utility* of a prospect might be undefined. In that case, standard decision theory, which compares prospects by comparing their expected utilities, falls silent. The canonical witness to this fact is the *St Petersburg prospect*, which yields utility 2 with probability $1/2$, utility 4 with probability $1/4$, and in general utility 2^n with probability $1/2^n$. The expected utility would be given by the sum

$$\frac{1}{2} \times 2 + \frac{1}{4} \times 4 + \cdots + \frac{1}{2^n} \times 2^n + \cdots = 1 + 1 + \cdots + 1 + \ldots$$

Since the sum does not converge, the expected utility is undefined. As a result, standard decision theory does not say how to value the St Petersburg prospect. In particular, it does not yield the correct verdict that it is worse than a prospect that with probability $1/2^n$ yields utility $2^n + 1$ instead of 2^n. This defect remains even when the expectation is taken in the extended real numbers, and hence can be given the value ∞. In that case, both of the prospects get expected utility ∞, so expected utility still cannot distinguish between them.[1]

But the problem is significantly worse than a mere gap in standard decision theory. The hypothesis of unbounded utility is flatly inconsistent with some very compelling principles governing the comparison of prospects, whether or not expected utilities are assigned to them. The question is not how to extend standard decision theory in this or that way, but rather whether a reasonable decision theory is compatible with unbounded utility at all. If not, then ethical principles that imply unbounded utility crash immediately.

The problem—a version of the Two-Envelope paradox—will be explained informally in the remainder of this section, and is precisified in Section 2.2. Section 2.3 presents and motivates a theory for reasoning in the presence of unbounded utility and establishes its consistency, thus guaranteeing that no further paradox arises. Having established the viability of this theory, the paper concludes (Section 2.4) by pointing out two of many interesting and underexplored avenues for further research on decision theory in the presence of unbounded utility.

The motivating paradox of the paper, first presented in Nalebuff 1989, can be illustrated as follows.[2] Suppose a coin is to be tossed (fairly and independently) infinitely many times. You have two options. The first is to take the *first-toss St Petersburg prospect*, whose outcome is determined by how the coin lands in accordance with Table 2.1. Choosing this prospect yields utility 2 if the coin lands tails right away, utility 4 if it lands heads then tails, 8 if it lands heads-heads-tails, and in general 2^n if it takes n tosses for it to land tails for the first time. Your second option is to take the *second-toss St Petersburg prospect*, described in Table 2.2.

[1] See Nover and Hájek 2004 for further discussion.

[2] See also Broome 1995, Arntzenius and McCarthy 1997, McGee 1999, Chalmers 2002, and Dietrich and List 2005. A similar puzzle is found in Seidenfeld, Schervish, and Kadane 2009.

Table 2.1: The first-toss St Petersburg prospect.

T	HT	...	H...HT	...
2	4	...	2^n	...

Table 2.2: The second-toss St Petersburg prospect ("/" separates the first coin toss from the rest).

T/T	H/T	T/HT	H/HT	...	T/H...HT	H/H...HT	...
2	2	4	4	...	2^n	2^n	...

Table 2.3: The first-toss prospect redescribed.

T/T	H/T	T/HT	H/HT	...	T/H...HT	H/H...HT	...
2	4	2	8	...	2	2^{n+1}	...

This prospect differs only in that the first coin toss is ignored. So whether the first toss lands heads or tails, the second-toss prospect yields utility 2 if the second toss lands tails, utility 4 if the second lands heads and the third lands tails, and in general utility 2^n if, starting from the second toss, it takes n tosses to get tails for the first time.[3] Which prospect, if either, is better? Here are arguments to two contradictory answers:

First argument: the two prospects are *equally good*. When evaluating a risky prospect, which coin tosses determine the outcome is not relevant. What matters are the probabilities of the various outcomes and the utility of those outcomes, and the two prospects are equal in these respects.

Second argument: the first-toss prospect is *strictly better* than the second-toss prospect. For suppose you learn the following: after the first toss, it took another n tosses for a toss to come up tails. That is, you learn something of the form: either T/H...HT or H/H...HT. This reduces your decision to that between the green cell in Table 2.3 and that in Table 2.2. And the green cell in Table 2.3 is better: halfway between 2 and 2^{n+1} is $2^n + 1$, so fifty-fifty between 2 and 2^{n+1} is an improvement on a sure 2^n (this is how utility is defined). But this is true for any n, so we can know in advance that the first will look a better gamble than the second. It seems then we should cut to the chase and judge the first better unconditionally.

These contradictory arguments constitute the paradox. The easy way out is to deny the possibility of the relevant prospects, an option usually taken by denying the unboundedness of utility.[4] Here, we will be taking the hard way out, which is to accept the possibility of the prospects but to block the reasoning that led to a

[3]If it matters, we can stipulate that both prospects yield utility 2 in the probability zero event that all tosses land heads.

[4]Savage 1954: 80-82, Russell and Isaacs 2021.

contradiction.[5] We will find that it is the second argument that must go. Although for any n the first-toss prospect is better than the second-toss prospect conditional on it taking n tosses to get tails, it is a fallacy to conclude from this that the first-toss prospect is better unconditionally (or so the proponent of unbounded utility must say). But this is only the beginning of the solution. The challenge is in demonstrating that, even without this intuitively tempting principle, there is a strong and systematic decision theory to be had that makes room for unbounded utility.

2.2 Formalizing the Paradox

2.2.1 Primitives

An *outcome* is intuitively a situation you might end up in, for which there is no further uncertainty about how good things will turn out if that outcome obtains. A *prospect* (also called a *gamble* or an *act*) is intuitively an option available to you which yields certain outcomes with certain probabilities. The object of decision theory is to rank the prospects in terms of the values of the outcomes they might yield and the probabilities with which they yield them.

To get clear on the paradox, it will be useful to employ a standard formalization of decision theory due to Savage (1954), which is rich enough to bring out the paradox, but idealized enough to obscure irrelevant complexities.

For concreteness, we can think of prospects as different ways of gambling on the result of the infinite sequence of coin tosses introduced in Section 2.1. That is, a prospect associates an outcome with each way the coin tosses might land, specifically the outcome that you will end up in if you take the prospect and the coin tosses land that way. Thus, the probability of a prospect p yielding an outcome o is given by the probability that the coin tosses land in one of the ways that p associates with o.

Let Ω be the set of maximally specific ways the coin tosses might land,[6] and $\mathbf{P}$ the usual probability function for fair and independent coin tosses (e.g., it gives a $1/4$ probability to the claim that the first toss lands heads and the third lands tails, and so on).[7] We take as primitive the following:

---

[5]Broome 1995 suggests this line, and it is endorsed by Arntzenius and McCarthy 1997, Dietrich and List 2005, and Meacham 2019, among others.

[6]Ω consists of propositions of the form

> Toss 1 lands heads and . . . and toss n lands tails and . . .

These are standardly represented by the set 2^ω of infinite 0-1-valued sequences. Any other state space with sufficiently rich probability structure would also serve for present purposes.

[7]With the usual measurable structure on 2^ω. $\mathbf{P}$ is the completion of the measure which gives

- A space of *outcomes* O (note: outcomes should not be conflated with ways the coin tosses might land, since outcomes are supposed to be maximally specific with regards to how good things turn out).
- A space of *prospects*, $\mathcal{P}$, which is a set of functions from Ω to O.[8]
- A relation $\succsim$ on prospects called *at least as good as*.

We also use $\succsim$ to compare outcomes with each other and with prospects, by identifying each outcome o with the prospect that yields o no matter what (i.e., the function that maps everything in Ω to o).

Finally, we define utility in the standard way, relative to an arbitrary choice of outcomes o_0 and o_1, with $o_0 \prec o_1$. o_0 and o_1 respectively have utilities 0 and 1 by stipulation; then define the utility function u so that a prospect with a $1/3$ chance of utility 3 and a $2/3$ chance of utility 0 is equally good as a sure utility 1, and similarly for other numbers.[9]

2.2.2 Assumptions

The paradox can be illustrated by deriving a contradiction from seven widely accepted decision-theoretic assumptions. Four of these are axioms from Savage's *Foundations of Statistics* (1954). The first two of these concern the very general structure of $\succsim$:

PREORDERING *At least as good as* is reflexive ($p \succsim p$) and transitive (if $p \succsim q \succsim r$, then $p \succsim r$).

---

probability $1/2^n$ to each set of sequences that contains all and only those which agree on some n positions.

[8] We assume that $\mathcal{P}$ contains only measurable functions, where O is given measurable structure compatible with $\succsim$ restricted to constant functions. We say that a prospect p *yields* o in the state $x \in \Omega$ if p maps x to o. Moreover, the *probability with which p yields an outcome in X* is given by

$$\mathbf{P}(\{\omega \in \Omega : p(\omega) \in X\}) \tag{2.1}$$

[9] In general:

- If $x \geq 1$, then for o to have utility x is for a $1/x$ chance of o and $1 - 1/x$ chance of o_0 to be equally good as a sure o_1.
- If $x \leq 0 < 1$, then for o to have utility x is for an x chance of o_1 and $1 - x$ chance of o_0 to be equally good as a sure o.
- If $x < 0$, then for o to have utility x is for a $1/(1-x)$ chance of o and a $1 - (1/(1-x))$ chance of o_1 to be equally good as a sure o_0.

TOTALITY For any prospects p and q, either $p \succsim q$ or $q \succsim p$.

The next one specifies that every possible correlation between coin-toss sequences and outcomes in $\mathcal{O}$ corresponds to a prospect in $\mathcal{P}$:

PROSPECT RICHNESS For any function[10] f from Ω to $\mathcal{O}$, there is a prospect in $\mathcal{P}$ that yields $f(s)$ when the coin tosses land in the sequence s.

The key axiom of Savage is the *Sure-Thing Principle*, which relies on a defined notion of *conditional* comparison of prospects. For a given subset of coin toss sequences E, let $p \upharpoonright E$ be the prospect that yields the same outcome as p when the way the coin tosses land is in E, but which yields utility 0 in every other situation. Write $p \succsim_E q$ for the comparison $p \upharpoonright E \succsim q \upharpoonright E$. Savage's Sure-Thing Principle says that to compare prospects that are equally good conditional on E, you can ignore what happens on E and compare them conditional on the complement of E $(\Omega \setminus E)$:

SURE-THING Let E be a subset of Ω with probability strictly between zero and one such that p and q are equally good conditional on E (i.e., $p \sim_E q$). Then $p \succsim q$ if and only if $p \succsim_{\Omega \setminus E} q$.

The next three principles go beyond Savage's axioms. Each has also been endorsed at various points in the literature, but together they give rise to the paradox. The first is the assumption of unbounded utility among outcomes in $\mathcal{O}$.

UNBOUNDED UTILITY For any real number x, there is an outcome in $\mathcal{O}$ that has utility x (§2.2.1).

Unbounded Utility implies that there are outcomes of utilities 2, 4, 8, and so on. Prospect richness says that putting these outcomes into Tables 2.1 through 2.3 gives you a corresponding prospect.[11]

The next assumption is that which undergirds the "first argument" of §2.1, which says that the coin a St Petersburg prospect starts from does not matter and so the two prospects are equally good:

[10] Strictly, we should restrict to *measurable* functions, where Ω is given its usual probability space structure and $\mathcal{O}$ is given measurable space structure compatible with the ordering induced by $\succsim$ on $\mathcal{O}$ by constant functions from Ω to $\mathcal{O}$.

[11] Prospect richness also guarantees that for each outcome o in $\mathcal{O}$ there is a prospect that yields o no matter how the coin tosses land. Thus we can compare outcomes in $\mathcal{O}$ with prospects and each other by using that sure prospect as a proxy (so, for example, $o \succsim p$ means that the sure o prospect is at least as good as p).

Stochastic Equivalence If for every outcome o, p is equally likely as q is to yield an outcome at least as good as o, then p is equally good as q.[12]

In effect, Stochastic Equivalence says that the comparison of prospects supervenes on the values of outcomes and the probabilities that the prospect yields them with. It can easily be confirmed that the first-toss and second-toss prospects of Section 2.1 both yield an outcome of utility greater than 2 with probability $1/2$, greater than 4 with probability $1/4$, and so on. So by Stochastic Equivalence they are equally good. Although Stochastic Equivalence goes beyond Savage's axioms, it is an entirely standard assumption for decision theory in settings where prospects can have infinitely many outcomes[13] and in cases where expected utilities are undefined on account of unbounded utility.[14]

The "second argument" in the paradox said that the first-toss prospect is strictly better than the second-toss prospect. The argument relied on the idea of comparing the two prospects conditional on a each member of a certain partition of how the coins might land. The principle needed to vindicate this reasoning is a generalization of Savage's Sure-Thing Principle (the name is from Russell and Isaacs 2021, in Dietrich and List 2005 a similar principle is called *Event-Wise Dominance*):

Countable Sure-Thing Let $\mathcal{E}$ be a partition of Ω into propositions of nonzero probability. Then if some prospect p is at least as good as q conditional on every member of $\mathcal{E}$, then p is at least as good as q. If in addition there is some member of $\mathcal{E}$ conditional on which p is strictly better than q (i.e., $p \succ_E q$), then p is strictly better than q.

Countable Sure-Thing implies that the first-toss St Petersburg prospect is strictly better than the second-toss prospect. We reason as follows. First, set $\mathcal{E}$ to be the following partition of Ω:

E_1 := the second coin toss lands tails

E_2 := the second coin toss lands heads and the third lands tails

$\vdots$

E_n := the second through $(n+1)^{\text{st}}$ coin tosses land heads and the $(n+2)^{\text{nd}}$ lands tails

$\vdots$

[12] In symbols:

$$\forall o(\mathbf{P}\{x \in \Omega : p(x) \succsim o\} = \mathbf{P}\{x \in \Omega : q(x) \succsim o\}) \to p \sim q$$

[13] See e.g., Blackwell and Girshick 1979 [1954]: 104, Joyce 1999: 26.

[14] Dietrich and List 2005, Easwaran 2014a, Meacham 2019.

E_n is the claim that if you take the first-toss St Petersburg prospect you will end up in the green cell in Table 2.3, and if you take the second-toss St Petersburg prospect you will end up in the green cell in Table 2.2. Reasoning using Savage's principles we may derive that the green cell in the first-toss prospect is better for any n, so $\mathcal{E}$ is a partition of propositions of nonzero probability such that conditional on any, the first-toss prospect is better. Countable Sure-Thing licenses the conclusion that the first-toss prospect is therefore better unconditionally. Thus we have derived the paradox from the aforementioned seven principles: Savage's axioms of Preordering, Totality, Prospect Richness, Sure-Thing, plus the three additional assumptions of Unbounded Utility, Stochastic Equivalence, and Countable Sure-Thing. Something has to give.

2.2.3 Weakening Totality and Stochastic Equivalence

Two of the most contentious principles in the derivation are Totality and Stochastic Equivalence. A different contradiction can be derived even when these principles are significantly weakened. The argument is slightly more complicated, but is dialectically significant because it sidesteps two frequent motivations for rejecting Totality and Stochastic Equivalence.

Most saliently, many authors are skeptical that these principles apply to prospects with infinitely many possible outcomes. Seidenfeld, Schervish, and Kadane (2009) and Lauwers and Vallentyne (2016; 2017) reject Stochastic Equivalence in such cases, and in particular they reject the equal value of the first- and second-coin St Petersburg prospects. Easwaran (2014a) Lauwers (2016), and Meacham (2019) believe that such prospects give rise to failures of Totality. Others—including Hare (2010), Bales, Cohen, and Handfield (2014), Schoenfield (2014)—have raised issues for Stochastic Equivalence in cases where Totality fails among the possible outcomes. So to reinforce the argument we shall employ weakenings of the two principles that avoid both complaints.

Let a *simple* prospect be one that can only yield finitely many different outcomes. And say that outcomes o and u are *comparable* if either $o \succsim u$ or $u \succsim o$. We will restrict both Totality and Stochastic Equivalence to simple prospects whose possible outcomes are all comparable to avoid both sorts of worry:

RESTRICTED TOTALITY If p and q are simple, and every pair of outcomes either might yield is comparable, then either $p \succsim q$ or $q \succsim p$.[15]

RESTRICTED STOCHASTIC EQUIVALENCE If p and q are simple, and every pair of

[15] Restricted Totality entails that every pair of outcomes in O is comparable—those who expect incomparability among outcomes may treat O as a special comparable subclass of all outcomes.

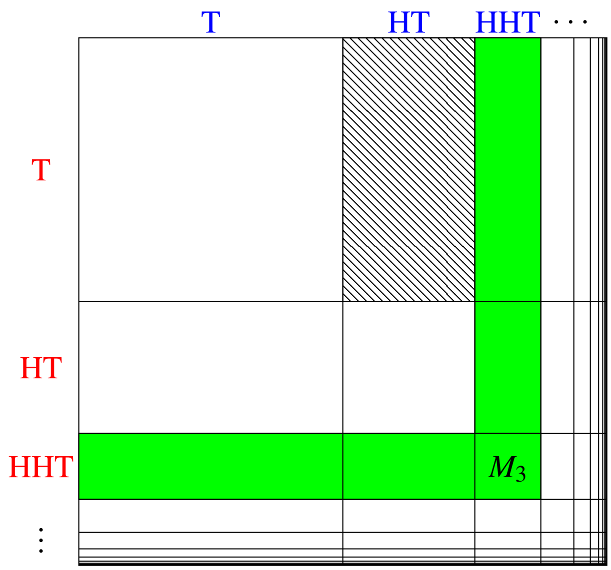


Figure 2.1: Geometric representation of state space with two infinite sequences of coin tosses. The space is partitioned into cells according to how many tosses it takes for the two sequences to get tails for the first time. For example, the hatched region represents the situation that the original sequence lands HT and the new sequence lands T. The proposition M_3 is highlighted in green. Note how the M_n will partition the space, and that each is symmetric with respect to the two sequences.

> outcomes either might yield is comparable, then if p and q yield the same outcomes with the same probabilities, then they are equally good.

To derive the paradox, we compare the first-toss and second-toss St Petersburg prospect with a further St Petersburg prospect that is determined by a disjoint infinite set of coin tosses. (To avoid positing these new coin tosses, we could have split the coin tosses into two infinite collections at the beginning.) Call these the *new* coin tosses. The *new* St Petersburg prospect yields utility 2 if the first new coin toss lands tails, 4 if the first lands heads and the second tails, and so on.

Recall that, by countable Sure-Thing, the first-toss prospect is better than the second-toss prospect (the argument to this effect is not altered by the restriction of Totality and Stochastic Equivalence). The new paradox will be derived by showing that the new St Petersburg prospect is nevertheless equally good as both of those.

To this end, we partition the space of ways the tosses could land, Ω, as follows: let M_n be the proposition that either the original tosses or the new tosses took n tosses to land tails for the first time, and the other took no more than n (see Figure 2.1). This sequence of propositions forms a partition of Ω into propositions of nonzero probability, so to show the first-toss and new St Petersburg prospects are equally good we can show that they are equally good conditional on any member of this partition. Crucially, each of the first-toss and new St Petersburg Prospects is simple conditional on any M_n. The figure makes clear how the situation is probabilistically symmetric, so by Restricted Stochastic Equivalence the first-toss and new St Petersburg prospects are equally good conditional on any M_n. Countable Sure-Thing then yields their unconditional equal goodness. The same would have gone if we had replaced the original coin tosses from the beginning with the original coin tosses besides the first coin, thereby establishing that second-toss prospect is also equally good as the new prospect.

2.3 Unbounded Utility Without Paradox

2.3.1 Costs of Rejecting Countable Sure-Thing

Given that neither Stochastic Equivalence nor Totality is the culprit, it is most natural to view the paradox as demonstrating the incompatibility of Unbounded Utility with Countable Sure-Thing. If so, then the viability of Unbounded Utility depends on whether Countable Sure-Thing can be rejected.

It must be stressed that Countable Sure-Thing formalizes an extremely tempting mode of reasoning in decision theory. This is the reasoning exemplified in Section 2.1, where we intuitively determine that the first-toss prospect is better than the second by dividing them each into cases (i.e., the green cell in Tables 2.2 and 2.3) and finding that first-toss prospect better conditional on any such case occurring. That is, Countable Sure-Thing gives voice to a sort of "anti-Dutch-book" principle: if you know you are about to learn something conditional on which the first-toss prospect looks better, then you should not need to learn it—you should be able to cut to the chase and judge it better unconditionally.

Nevertheless, mere intuitions are not sufficient to rule out Unbounded Utility. Failures of Countable Sure-Thing are surprising, but it is unclear whether they should be regarded as impossible, or simply as one more surprising consequence of reasoning with the infinite (in this case, prospects with infinitely many possible outcomes). McGee 1999, Dietrich and List 2005, and Meacham 2019 take the second line, and Arntzenius, Elga, and Hawthorne 2004 reinforce the case by drawing parallels with the present paradox and a range other puzzles in decision theory.[16] It should also be recognized that Unbounded Utility also has much going in its favour, such as a wide range of general ethical principles that entail it.[17]

[16]See Russell and Isaacs 2021 for a response on behalf of Countable Sure-Thing.

[17]Most notably, Unbounded Utility follows from standard Expected Total Utilitarianism, and therefore also from ethical views that agree with Expected Total Utilitarianism on a sufficiently

But there is a more pressing problem for those who accept Unbounded Utility, which is that decision theory in the presence of Unbounded Utility is much less well understood than decision theory where Unbounded Utility is eschewed in favour of Countable Sure-Thing. In the latter case, all of the other principles listed in Section 2.2 can be consistently endorsed, and Russell (2020) has demonstrated that in the presence of these principles, Countable Sure-Thing gives rise to a highly general representation theorem (generalizing a result of Blackwell and Girshick 1979: 105). The result is an elegant, powerful, and overall desirable decision theory.

By contrast, work on decision theory in the presence of Unbounded Utility has proceeded in a very piecemeal fashion. Too often, the literature focuses on concrete methods of adding numbers together as proposals for avoiding the problem of undefined expected utilities, obscuring the more pressing need for deciding which of the fundamental principles of decision theory, such as Stochastic Equivalence or Savage's Sure-Thing principle, can be accepted in light of the present paradox.[18] When fundamental principles are discussed, the results often seem dire. Seidenfeld, Schervish, and Kadane (2009), as well as Lauwers and Vallentyne (2016; 2017) take the very drastic step of rejecting Stochastic Equivalence, whereas Fine (2008) and Smith (2014) question even the assumption that a prospect guaranteed to yield a better outcome is better. Another very common (although plausibly less problematic) concession is to accept widespread failures of Totality (see, e.g., Colyvan 2008, Easwaran 2014a, Lauwers 2016, Meacham 2019).

Here is not the place to answer the question of which (if any) of Unbounded Utility and Countable Sure-Thing should ultimately be accepted. The goal of this paper is to address the second problem, which is answered by demonstrating that even in the presence of Unbounded Utility, the highly orthodox principles listed above, besides Countable Sure-Thing but including Totality and Stochastic Equivalence, are consistent. Therefore, decision-theoretic orthodoxy need not be rejected to make room for Unbounded Utility.

2.3.2 *The Theory* DTU

This section compiles these principles with an important strengthening of Stochastic Equivalence to *Stochastic Dominance*, and illustrates how the resulting theory can capture a great deal of good decision-theoretic reasoning. The resulting theory will be called DTU, (for Decision Theory with Unbounded utility).

---

wide range of case, e.g. in cases where deontological side-constraints are met or cases where all beneficiaries of an action are equally well-deserving. See also Beckstead and Thomas 2021, Russell 2021, Wilkinson 2022.

[18] Especially literature following Nover and Hájek 2004, including Colyvan 2008, Easwaran 2008, Gwiazda 2014, Easwaran 2014b, and Bartha 2016.

DTU extends the axioms stated in Section 2.2 with the exception of Countable Sure-Thing. These were the axioms of Savage, namely Preordering, Totality, Prospect Richness, and Sure-Thing, plus the two further axioms of Unbounded Utility and Stochastic Equivalence. DTU extends these axioms by strengthening Stochastic Equivalence to the also entirely standard principle of *Stochastic Dominance*:

Stochastic Dominance If for every outcome o, p is at least as likely as q is to yield an outcome better than o, then p is at least as good as q. If there is also some outcome o such that p is strictly more likely than q to yield an outcome better than o, then p is strictly better than q.[19]

Whereas Stochastic Equivalence says that the value of a prospect supervenes on the probability distribution over outcomes that prospect yields, stochastic dominance adds that shifting probability from worse outcomes to better ones is always an improvement. DTU will also include the simplifying assumption of *Archmidean Outcomes*, which says that every outcome in O has a well-defined utility (in effect ignoring the possibility of infinitely good or bad outcomes and of infinitesimally good or bad outcomes).

DTU is a strong theory, and one in which at least a very large proportion of good decision-theoretic reasoning can be carried out. Here, this will be illustrated by pointing out three consequences of the theory which make its overall structure clear.

The first consequence is the principle of *Statewise Dominance*:

Statewise dominance If it is certain that p would yield an outcome at least as good as q would, then $p \succsim q$. If there is also a nonzero probability that p yields a strictly better outcome than q, then $p \succ q$.[20]

Statewise Dominance is a desirable consequence because it is such an overwhelmingly plausible principle, but it has particular significance in the present context, where it has been suggested that the accepter of Unbounded Utility must give up

[19]Stochastic Dominance follows from Stochastic Equivalence by the mostly uncontroversial principle of Statewise Dominance, stated below. An early analysis of Stochastic Dominance is in Quirk and Saposnik 1962.

[20]We derive statewise dominance from stochastic dominance as follows. First, let p and q satisfy the antecedent of the first part of statewise dominance, and suppose for contradiction that p is worse than q. By stochastic dominance, it follows that for some o, q is more likely than p is to yield an outcome better than o. Then, conditional on the claim that q would yield an outcome better than o, there must be a nonzero probability that p would yield an outcome no better than o. Hence there is a nonzero probability that there is an outcome better than o that q yields and an outcome no better than o that p yields, so there is a nonzero probability that p yields an outcome worse than what q yields, contradicting the supposition.

dominance principles altogether in light of the various paradoxes. Arntzenius, Elga, and Hawthorne (2004) put this idea in a box:

> In infinite cases, rationality does not require one to choose one's dominant options

The box can be interpreted as disavowing principles like Countable Sure-Thing, or as disavowing even principles like Statewise or Stochastic Dominance. The consistency of DTU demonstrates, confirming the view of Dietrich and List (2005), that on the second interpretation Arntzenius and coauthors would be overreacting. The consistency also makes clearer how various dominance-like forms of reasoning must be distinguished to avoid paradox.[21]

The second important consequence of DTU is the fact that, since it includes Stochastic Dominance, the comparison of prospects can be simplified by instead comparing the probability distributions over outcomes that they give rise to. In fact, since we are supposing every outcome to have a utility and every utility to be had by some outcome, we can simplify matter further by comparing only the probability distribution over *utilities* (i.e., real numbers) that the two prospects give rise to. That is, the prospect p can be harmlessly conflated with the probability distribution μ that, for each set of real numbers X, assigns to X the probability that p yields an outcome whose utility is in X. Moreover, Prospect Richness guarantees that every real-valued probability distribution corresponds to some prospect, so the question of how to rank prospects reduces entirely to that of how to rank real-valued probability distributions.

The principles of DTU carry over smoothly to the comparison of probability distributions. We use greek letters μ, ν, ξ as variables for real-valued probability distributions and abuse notation by writing $\mu \succsim \nu$ for the claim that a prospect whose utility is distributed according to μ is at least as good as any prospect that yields utility according to the distribution ν. In this setting the Sure-Thing principle reduces to von Neumann and Morgenstern's (1944) principle of *Independence*:

INDEPENDENCE $\mu \succsim \nu$ if and only if, for any distribution ξ and number $x \in (0,1)$, $x\xi + (1-x)\mu \succsim x\xi + (1-x)\nu$ (the mixture of distributions $x\xi + (1-x)\mu$ is defined as the distribution satisfying

$$(x\xi + (1-x)\mu)(X) = x\xi(X) + (1-x)\mu(X)$$

for any measurable set X).

[21] It is worth noting that in the presence of the other assumptions, Stochastic Dominance follows

Stochastic Dominance can be rewritten in terms of distributions as follows:

STOCHASTIC DOMINANCE If for every real number x, μ assigns probability at least as great as ν does to the ray $[x, \infty)$ (i.e., if $\mu[x, \infty) \geq \nu[x, \infty)$), then $\mu \succsim \nu$. If there is also a real number y such thats μ assigns a strictly greater probability to $[y, \infty)$ than ν does ($\mu[y, \infty) > \nu[y, \infty)$), then $\mu > \nu$.

Both of these principles can now be visualized in terms of the *cumulative distribution functions*, or CDFs, of real-valued probability distributions. The CDF of μ is the function that assigns to each number x the probability that μ gives to the ray $[x, \infty)$. See Figures 2.2 and 2.3 for how the Independence and Stochastic Dominance conditions are visualized in terms of CDFs.

The final important consequence of DTU to discuss is the extent to which it vindicates standard decision-theoretic reasoning in the case of simple prospects (prospects which have only finitely many possible outcomes). By slightly varying von Neumann and Morgenstern's representation theorem (1944),[22] one can show that expected utility theory for simple prospects is simply a theorem of DTU. That is, we can recover the following principle:

SIMPLE EUT If p and q are simple prospects, then $p \succsim q$ if and only if the expected utility of p is at least as great as that of q.

Expected utility can also be straightforwardly visualized in terms of CDFs, as in Figure 2.4. DTU therefore explains why expectation-based reasoning is good in the finite cases without falling silent in the infinite cases, for the principles of DTU are fully general. In total, then, DTU is a strong and desirable theory. The question that remains is whether it contains a contradiction.

2.3.3 *Consistency of* DTU

No further contradiction arises in DTU without the addition of Countable Sure-Thing principle:[23]

Theorem 1. DTU *is consistent.*

In this sense we can be said to have solved the motivating paradox. The consis-

from Statewise Dominance plus Stochastic Equivalence.

[22]See Joyce 1999 for exposition.

[23]By 'consistent' I mean consistent in a suitably chosen formal system that includes as theorems at least all the mathematical and logical assumptions implicit in DTU (e.g., that Modus Ponens is valid and there are infinitely many real numbers). The exact details don't matter, so long as the chosen background formal system is provably consistent on its own.

Figure 2.2: Purple is the mixture of blue and red (by 0.5).

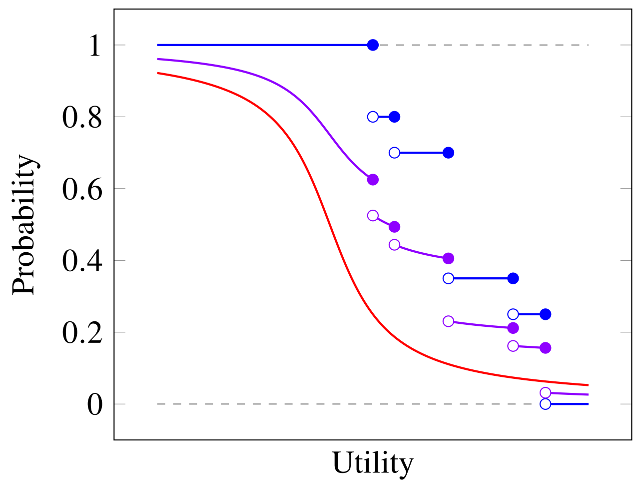


Figure 2.3: Blue dominates red.

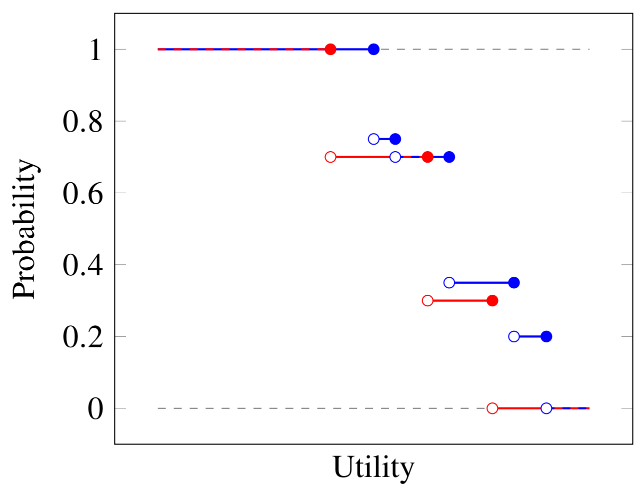
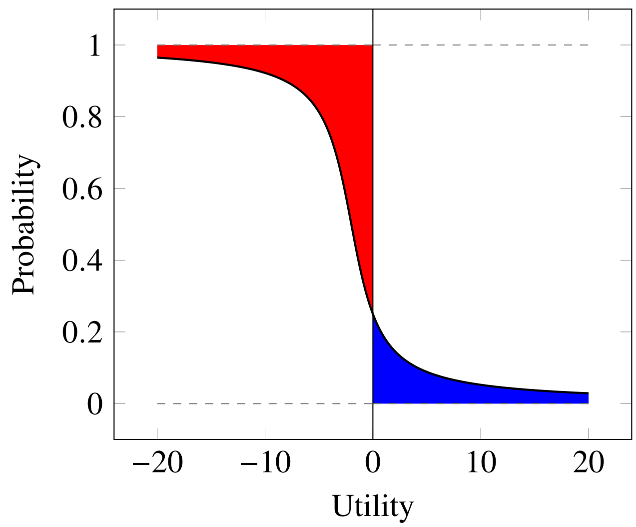


Figure 2.4: The expectation of the distribution is equal to the blue area minus the red area, if both areas are finite.



tency of DTU is established in Appendix 2.6. Here, the main ideas will be covered by establishing the consistency of a weaker theory, DU, which is DTU but with the axiom of Totality weakened to Restricted Totality (§2.2.3). DU weakens DTU by withholding on whether every pair of prospects is comparable. The proof of the consistency of DU is highly intuitive and entirely elementary; and it forms the core of the proof of Theorem 1, which complicates the argument by a standard application of non-principal ultrafilters. The proof strategy in either case can be attributed to Lauwers (2016), who uses it to establish the consistency of weak fragments of DU and DTU.[24]

It should be noted that the consistency of DTU over just DU is important even to those who reject the axiom of Totality. This is because authors who accept some or all of the axioms of DU—including Savage, as well as Joyce (1999), Easwaran (2014a) and many others who reject Totality—take the principles of DU to be not just true principles governing the structure of value under risk, but as principles constraining *all* coherent ways of preferring prospects. That is, they take (some or all of) the axioms of DU to be true when '$\succsim$' is replaced with a variable universally quantified over all "coherent preference orderings". On this picture, the consistency of Totality when added to DU is equivalent to the consistency of the following

[24]In particular, Lauwers can be credited with noticing that the preordering $\sqsupseteq$ defined in this section has properties corresponding to the axioms of Preordering and Independence, and that it can be extended to a total ordering with an ultrafilter as is done in Appendix 2.6. The author discovered the proof strategy independently.

principle:

Coherent Extensibility (i) For every pair of prospects p, q and coherent preference ordering $\succsim$ for which $q \not\succsim p$, there is a coherent preference ordering $\succsim'$ that extends $\succsim$ such that $p \succsim' q$.[25]

(ii) For any set of coherent preference orderings that extend one another, there is a coherent preference ordering that extends all of them.[26]

Coherent Extensibility is plausible even supposing Totality fails, for reasons presented in Kaplan 1983 and Hare 2010 (see also de Finetti 1972 §3.10.7 for analogous considerations in probability theory). Therefore, the consistency of DTU is significant even to those who reject Totality.

The remainder of this section establishes the following:

Theorem 2. DU *is consistent.*

The consistency will be shown by showing that the theory is true on an interpretation, by constructing an ordering $\sqsupseteq$ of prospects that has the properties DU says $\succsim$ has. What this amounts to is constructing a preordering of probability distributions satisfying Independence, Stochastic Dominance, and Simple EUT (§2.3.2), which we will also write $\sqsupseteq$.[27] See Figures 2.2, 2.3, and 2.4 for diagrams illustrating these conditions.

The construction of $\sqsupseteq$ is built on the observation that the ordering of distributions by their expected values satisfies all three constraints when we restrict attention to distributions whose expectation is well-defined (all of which are confirmed by inspecting the figures).[28]

---

[25]$\succsim_1$ *extends* $\succsim_2$ if for any p, q, $p \succsim_1 q$ if $p \succsim_2 q$, and $p \succsim_1 q$ only if $q \not\succsim_2 p$. Note that every ordering extends itself.

[26]Coherent Extensibility implies the existence of a total coherent preference ordering by Zorn's lemma, so the '$\succsim$' in DTU may be interpreted as that coherent preference ordering.

[27]Simple EUT immediately implies Unbounded Utility and Restricted Totality. Independence implies Sure-Thing. Prospect Richness is built into the model because all probability distributions are in the domain of $\sqsupseteq$.

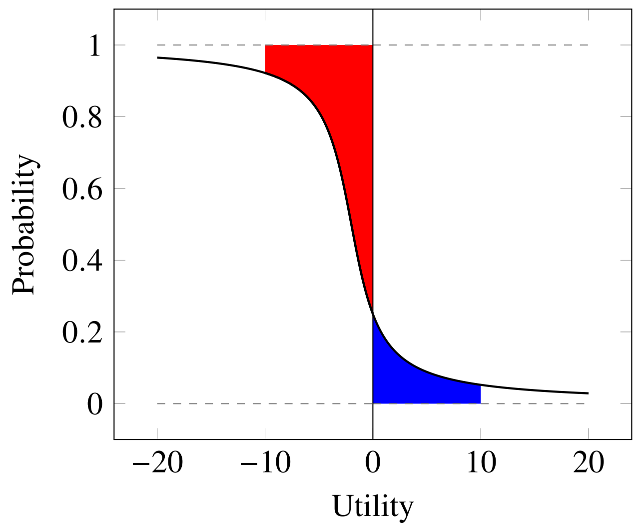
[28]Writing $\mathbf{E}\mu$ for the expectation of μ, we observe the following:

- Whenever Simple EUT requires that $\mu \sqsupseteq \nu$, by definition the expectations $\mathbf{E}\mu$ and $\mathbf{E}\nu$ are well-defined and $\mathbf{E}\mu \geq \mathbf{E}\nu$.
- The expectation of a mixture of prospects with well-defined expectations is the mixture of their expectations:

$$\mathbf{E}(x\xi + (1-x)\mu) = x\mathbf{E}\xi + (1-x)\mathbf{E}\mu$$

Hence $\mathbf{E}(x\xi + (1-x)\mu)$ is at least as great as $\mathbf{E}(x\xi + (1-x)\nu)$ if and only if $\mathbf{E}\mu$ is at least as

Figure 2.5: The expectation truncated to 10 of the distribution (i.e., $\mathbf{E}_t\mu$ for $t = 10$) is equal to the blue area minus the red area. Notice that both areas will be finite for any t.



The only problem is that the expectation of a distribution is not always defined. This, we have already observed (§2.1), is the case for the distribution of the St Petersburg prospect. For the construction of $\sqsupseteq$, what we will do instead is compare prospects by the sequence of their *truncated expectations*. For a positive real number t, write $\mathbf{E}_t\mu$ for the expectation of the distribution that is just like μ but that puts all the probability of a value bigger than t onto t, and all the probability of a value less than $-t$ onto $-t$. This distribution agrees with μ on subsets of $(-t, t)$, but gives probabilities $\mu([t, \infty))$ and $\mu((-\infty, -t])$ respectively to the utilities t and $-t$.[29] See Figure 2.5 for a geometrical representation.

When the expectation of a distribution μ is well-defined, it is given by the limit of the truncated expectations as t goes to infinity:

$$\mathbf{E}\mu = \lim_{t\to\infty} \mathbf{E}_t\mu$$

great as $\mathbf{E}\nu$, when the expectations exist.

- If μ stochastically dominates ν and both have well-defined expectations, $\mathbf{E}\mu$ is at least as great as $\mathbf{E}\nu$, and $\mathbf{E}\mu$ is strictly greater if it strictly stochastically dominates.

29

$$\mathbf{E}_t\mu := \int_{-t}^{t} z \,\mathrm{d}\,\mu(z)$$

Crucially, truncated expectations are all well-defined for any real-valued probability distribution. Even if the limit is not well-defined, each truncated expectation is. Thus, *every* pair of probability distributions can be compared by comparing their truncated expectations, and this is how we will build the required preordering.

$\mu \sqsupseteq \nu$ will be defined as: as t increases, eventually $\mathbf{E}_t\mu$ is always at least as great as $\mathbf{E}_t\nu$. More precisely:

$$\mu \sqsupseteq \nu := (\exists s > 0)(\forall t > s)(\mathbf{E}_t\mu \geq \mathbf{E}_t\nu)$$

$\sqsupseteq$ is clearly reflexive, and is transitive because if $\mathbf{E}_t\mu \geq \mathbf{E}_t\nu$ for all t greater than s_1 and $\mathbf{E}_t\mu \geq \mathbf{E}_t\xi$ for all t greater than s_2, then $\mathbf{E}_t\mu \geq \mathbf{E}_t\xi$ for all t greater than both s_1 and s_2.

We then use the properties of expectations to show that $\sqsupseteq$ always satisfies Simple EUT, Independence, and Stochastic Dominance:

Simple EUT: If μ is simple, then there is a biggest utility, s_μ, that gets a nonzero probability. Then for t bigger than s_μ, $\mathbf{E}_t\mu = \mathbf{E}\mu$. So in particular, if $\mathbf{E}\mu \geq \mathbf{E}\nu$ then for some s (whichever is larger out of s_μ or s_ν), $\mathbf{E}_t\mu \geq \mathbf{E}_t\nu$ for all $t > s$.

Independence: The truncated expectation of a mixture is the mixture of truncated expectations:

$$\mathbf{E}_t(x\xi + (1-x)\mu) = x\mathbf{E}_t\xi + (1-x)\mathbf{E}_t\mu$$

And by basic algebra we have $\mathbf{E}_t\mu \geq \mathbf{E}_t\nu$ if and only if

$$x\mathbf{E}_t\xi + (1-x)\mathbf{E}_t\mu \geq x\mathbf{E}_t\xi + (1-x)\mathbf{E}_t\nu$$

So $\mu \sqsupseteq \nu$ if and only if $x\xi + (1-x)\mu \sqsupseteq x\xi + (1-x)\nu$.

Stochastic Dominance: If μ stochastically dominates ν, then the truncation of μ to t also stochastically dominates the truncation of ν to t. Hence $\mathbf{E}_t\mu \geq \mathbf{E}_t\nu$ for any t so $\mu \sqsupseteq \nu$. If in addition, for some number s, μ gives a strictly higher probability than ν to (s,∞), then $\mathbf{E}_t\mu > \mathbf{E}_t\nu$ for all $t > s$. So $\mu \sqsupseteq \nu$.

DU is therefore consistent.

2.4 Further Directions

DTU is a strong and plausible theory, so is a good starting point for the investigation of unbounded utility. Nevertheless, DTU leaves many questions undecided. The model construction techniques for DTU and DU are fairly general, and slight variations in the models show how various comparisons of prospects that might seem obvious are in fact independent of the theory. One diagnosis is that DTU should be extended so as to imply these comparisons. Another is that, in light of the failure of Countable Sure-Thing, we should be a little less trusting of our intuitions when unbounded utility is on the table, so comparisons that go beyond DTU should, sometimes surprisingly, be regarded as suspicious. In this section, we consider two possible extensions of DTU as case studies. These case studies illustrate how there is still a great deal of work, both formal and philosophical, to be done in the investigation of unbounded utility.

2.4.1 Expected Utility Theory

Consider the principle that lies at the center of decision theory:

Expected Utility Theory (EUT) If p and q both have an expected utility, then $p \succsim q$ if and only if the expected utility of p is at least as great as the expected utility of q.

EUT is silent about prospects, like the St Petersburg prospect, which lack an expected utility. Still, the existence of prospects beyond its remit does not contradict EUT as stated above; they simply show it is not the end of the story. And in almost every contemporary treatment of unbounded utility, EUT is assumed to be true. For example, Colyvan 2008, Easwaran 2008, Seidenfeld, Schervish, and Kadane 2009, Gwiazda 2014, Easwaran 2014b, Bartha 2016, Lauwers and Vallentyne 2016, and Meacham 2019 all presuppose EUT, and take their task to be to *extend* EUT so as to be able to evaluate prospects without an expected utility (Easwaran 2014a and N. J. J. Smith 2014 are notable exceptions).

There are some dissenters from EUT, for example Buchak 2013 and N. J. J. Smith 2014. But these theorists also depart radically from orthodox decision theory, rejecting the Sure-Thing principle and *ipso facto* the Independence axiom of DTU. However, even if we take the orthodox approach to decision theory with DTU, EUT builds in a novel assumption. That is, Expected Utility Theory is *not* a consequence of DTU, and hence must be supported by some additional argument.

Rather, DTU vindicates only *Simple EUT* (§2.3.2), which is the restriction of EUT to simple prospects—those with only finitely many possible outcomes. The proof that DTU is compatible with the failure of EUT is in Appendix 2.6 (Theorem 3). An easier exercise is to show that the model just constructed for DU is one where EUT fails. For this, consider a prospect that yields utility n with probability $1/2^n$.

This prospect has expected utility given by the sum:

$$\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \cdots + \frac{1}{2^n} \times n + \cdots = 2$$

By contrast, its truncated expected utility is always strictly less than 2, so this prospect is strictly worse than 2 in the model.

While EUT is not implied by DTU, nor is its negation:

Theorem 3. *EUT is independent of* DTU.[30]

Whether it should be accepted is therefore an independent philosophical question. I submit that it should not be. The generalization from Simple EUT to EUT proper is almost exactly analogous to the generalization from Sure-Thing to Countable Sure-Thing, and we have seen that this is exactly the sort of generalization that unbounded utility should make us suspicious of.

To draw out the analogy, consider a prospect p that yields utility y with probability x (among other possibilities). The Sure-Thing principle says that p can be compared with any other prospect as follows: first conditional on p yielding y, then conditional on p not yielding y. Thus, the value contributed by the x chance of utility y is independent of whatever other outcomes p might yield, and so in some sense can simply be added to the total. The definition of utility is chosen so that the utility contribution of an x chance of utility y is the product xy, so the value of p can be thought of as xy *plus* the value of p that is contributed by the non-y possibilities.

$$xy + \ldots$$

Suppose p is simple, so it has finitely many possible utilities, y_1, through y_n, which it yields with some probabilities x_1 through x_n. By repeating the above reasoning n times, the Sure-Thing principle says that the value of p is given by the sum:

$$x_1 y_1 + \cdots + x_n y_n$$

Which is exactly the verdict of Simple EUT.

To derive EUT proper from this sort of reasoning, we need to extend it to non-simple prospects, such as a prospect q that yields utilities in the infinite sequence $y_1, y_2, \ldots$ with probabilities $x_1, x_2, \ldots$. The Sure-Thing principle says that the value of q is given by the value of the contribution of an x_n chance of y_n plus the value of q conditional on not getting utility y_n, for any n. But since there are infinitely many different possible outcomes, it just does not follow that the value of q is determined

[30] Proof in Appendix 2.6.

by adding these contributions together all at once with the sum

$$x_1y_1 + \cdots + x_ny_n + \ldots$$

which is what EUT demands.

To get this in the same way as we got Simple EUT from Sure-Thing, we would need a principle that allows us to compute the value of a prospect by dividing it into *infinitely* many pieces, and evaluating those pieces all at once. And the most obvious such principle would be *Countable* Sure-Thing. Since we have just been forced to reject Countable Sure-Thing, EUT should also look very dubious, and certainly should not be considered as being on a par with Simple EUT.

2.4.2 Symmetries of Value

We now turn to another possible extension of DTU. The idea is that the *at least as good* ordering of prospects should be invariant under certain operations on the prospects. For example, if you add a fixed utility all possible outcomes of two prospects, that shouldn't affect which is better of the two. Similarly, if you multiply the utility of all of their possible outcomes by a fixed positive amount. Writing $\alpha \cdot p + \beta$ for a prospect which yields utility $\alpha x + \beta$ whenever p would yield utility x, we formalize this idea with the following two principle:

Shift Invariance For any real number β, $p \succsim q$ if and only $p + \beta \succsim q + \beta$.

Scale Invariance For any positive real number α, $p \succsim q$ if and only if $\alpha \cdot p \succsim \alpha \cdot q$.

This idea has come up a few times in the literature, such as in Hájek's (2014, pp. 490-491) discussion of Gwiazda 2014, and by Easwaran:[31]

> [. . .] utility is only defined up to a shift and a stretch—there is no well-defined 0 and no well-defined unit in which utility is measured. Thus, it ought to be the case that adding a constant to the utility of every outcome of a gamble affects the overall value of the gamble by adding the same constant, and similarly for multiplying the utilities by a positive constant.[32] Easwaran 2014b, p. 490.

[31] More recently, Wilkinson 2022 has marshalled a principle closely related to Scale Invariance (called *Scale Independence*) in defence of Unbounded Utility.

[32] It should be noted that the strengthening of Scale Invariance suggested in a naïve reading of this passage is straightforwardly inconsistent with DTU (using the same argument as in Section 2.1):

Value Scaling A fifty-fifty gamble between $2 \cdot p$ and 0 is equally good as p.

Another natural idea in the same family is that replacing utility x with utility $-x$ throughout two prospects reverses their comparison.

Reflection Anti-Invariance $p \succsim q$ if and only if $(-1) \cdot q \succsim (-1) \cdot p$.

These principles are plausible, and have already been put to work in various places. But mere plausibility isn't enough. Without some guarantee that they do not lead to a further paradox, they cannot be accepted. The model theory just laid out above and in the appendix allows us to begin the investigation of these principles. There are two extremely salient questions to ask of each principle. First: does it *follow* from DTU? Second: is it *consistent* with DTU, or some other natural theory like DU?

On the first question, the answer is surprisingly negative in all three cases. That is, none of shift or Scale Invariance, or Reflection Anti-Invariance, can be derived in DTU (see Appendix 2.6). On the other hand, the answer is positive in each case on the question of consistency. In fact, it is consistent with DTU that all three principles be true together (this is established when we return to these principles in Chapter 4).

These three principles therefore constitute a promising way of strictly extending DTU. The question of whether to accept them once again becomes a philosophical one, and a good example of how decision theory in the presence of unbounded utility remains an interesting topic for further research.

References

Arntzenius, Frank, Adam Elga, and John Hawthorne (2004). "Bayesianism, Infinite Decisions, and Binding". In: *Mind* 113.450, pp. 251–283.

Arntzenius, Frank and David McCarthy (1997). "The Two Envelope Paradox and Infinite Expectations". In: *Analysis* 57.1, pp. 42–50.

Bales, Adam, Daniel Cohen, and Toby Handfield (2014). "Decision Theory for Agents with Incomplete Preferences". In: *Australasian Journal of Philosophy* 92.3, pp. 453–70.

Bartha, Paul (2016). "Making Do Without Expectations". In: *Mind* 125.499, pp. 799–827.

Beckstead, Nick and Teruji Thomas (2021). "A Paradox for Tiny Probabilities and Enormous Values". In: *GPI Working Papers* 7.

Blackwell, David A and Meyer A Girshick (1954). *Theory of games and statistical decisions*. Wiley.

— (1979). *Theory of games and statistical decisions*. Courier Corporation. *Theory of games and statistical decisions*. Wiley, 1954.

Broome, John (1995). "The Two-Envelope Paradox". In: *Analysis* 55.1, pp. 6–11.

Buchak, Lara (2013). *Risk and Rationality*. Oxford University Press.

Chalmers, David (2002). "The St. Petersburg Two-Envelope Paradox". In: *Analysis* 62.2, pp. 155–157.

Colyvan, Mark (2008). “Relative Expectation Theory”. In: *Journal of Philosophy* 105.1, pp. 37–44.
de Finetti, Bruno (1972). *Probability, Induction, and Statistics*. New York: John Wiley.
Dietrich, Franz and Christian List (2005). “The Two-Envelope Paradox: An Axiomatic Approach”. In: *Mind* 114.454, pp. 239–248.
Easwaran, Kenny (2008). “Strong and Weak Expectations”. In: *Mind* 117.467, pp. 633–641.
— (2014a). “Decision Theory Without Representation Theorems”. In: *Philosophers' Imprint* 14.
— (2014b). “Principal Values and Weak Expectations”. In: *Mind* 123.490, pp. 517–531.
Fine, Terrence (2008). “Evaluating the Pasadena, Altadena, and St Petersburg Gambles”. In: *Mind* 117.467, pp. 613–632.
Gwiazda, Jeremy (2014). “Orderly Expectations”. In: *Mind* 123.490, pp. 503–516.
Hájek, Alan (2014). “Unexpected Expectations”. In: *Mind* 123.490, pp. 533–567.
Hare, Caspar (2010). “Take the Sugar”. In: *Analysis* 70.2, pp. 237–247.
Jech, Thomas (2006). *Set Theory: The Third Millennium Edition, Revised and Expanded*. Springer Monographs in Mathematics. Springer Berlin Heidelberg.
Joyce, James M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.
Kaplan, Mark (1983). “Decision Theory as Philosophy”. In: *Philosophy of Science* 50.4, pp. 549–577.
Lauwers, Luc (2016). “Why Decision Theory Remains Constructively Incomplete”. In: *Mind* 125.500, pp. 1033–1043.
Lauwers, Luc and Peter Vallentyne (2016). “Decision Theory Without Finite Standard Expected Value”. In: *Economics and Philosophy* 32.3, pp. 383–407.
— (2017). “A Tree Can Make a Difference”. In: *Journal of Philosophy* 114.1, pp. 33–42.
McGee, Vann (1999). “An Airtight Dutch Book”. In: *Analysis* 59.4, pp. 257–265.
Meacham, Christopher J. G. (2019). “Difference Minimizing Theory”. In: *Ergo: An Open Access Journal of Philosophy* 6.
Nalebuff, Barry (1989). “Puzzles: The Other Person's Envelope is Always Greener”. In: *The Journal of Economic Perspectives* 3.1, pp. 171–181.
Nover, Harris and Alan Hájek (2004). “Vexing Expectations”. In: *Mind* 113.450, pp. 237–249.
Quirk, James P and Rubin Saposnik (1962). “Admissibility and measurable utility functions”. In: *The Review of Economic Studies* 29.2, pp. 140–146.
Russell, Jeffrey Sanford (2020). “Non-Archimedean Preferences Over Countable Lotteries”. In: *Journal of Mathematical Economics* 88.May 2020, pp. 180–186.
— (2021). “On Two Arguments for Fanaticism”. In: *GPI Working Papers* 17.
Russell, Jeffrey Sanford and Yoaav Isaacs (2021). “Infinite Prospects”. In: *Philosophy and Phenomenological Research* 103.1, pp. 178–198.
Savage, Leonard (1954). *The Foundations of Statistics*. Wiley Publications in Statistics.
Schoenfield, Miriam (2014). “Decision Making in the Face of Parity”. In: *Philosophical Perspectives* 28.1, pp. 263–277.

Seidenfeld, Teddy, Mark Schervish, and Joseph Kadane (2009). "Preference for Equivalent Random Variables: A Price for Unbounded Utilities". In: *Journal of Mathematical Economics* 45, pp. 329–340.

Smith, Nicholas J. J. (2014). "Is Evaluative Compositionality a Requirement of Rationality?" In: *Mind* 123.490, pp. 457–502.

von Neumann, John and Oskar Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Wilkinson, Hayden (2022). "In Defense of Fanaticism". In: *Ethics* 132.2, pp. 445–477.

2.5 Appendix: Axioms of DTU

PREORDERING *At least as good as* is reflexive ($p \succsim p$) and transitive (if $p \succsim q \succsim r$, then $p \succsim r$).

TOTALITY For any prospects p and q, either $p \succsim q$ or $q \succsim p$.

SURE-THING Let E be a subset of Ω with probability strictly between zero and one such that p and q are equally good conditional on E (i.e., $p \sim_E q$). Then $p \succsim q$ if and only if $p \succsim_{\Omega \setminus E} q$.

PROSPECT RICHNESS For any function[33] f from Ω to $\mathcal{O}$, there is a prospect in $\mathcal{P}$ that yields $f(s)$ when the coin tosses land in the sequence s.

UNBOUNDED UTILITY For any real number x, there is an outcome in$\mathcal{O}$ that has utility x (§2.2.1).

ARCHIMEDEAN OUTCOMES Every outcome in $\mathcal{O}$ has a utility.

STOCHASTIC DOMINANCE If for every outcome o, p is at least as likely as q is to yield an outcome better than o, then p is at least as good as q. If there is also some outcome o such that p is strictly more likely than q to yield an outcome better than o, then p is strictly better than q.

DU replaces Totality with:

RESTRICTED TOTALITY If p and q are simple, and every pair of outcomes either might yield is comparable, then either $p \succsim q$ or $q \succsim p$.

Also of note are the following consequence of DU.

[33] Strictly, we should restrict to *measurable* functions, where Ω is given its usual probability space structure and $\mathcal{O}$ is given measurable space structure compatible with the ordering induced by $\succsim$ on constant functions from Ω to $\mathcal{O}$.

INDEPENDENCE $\mu \succsim \nu$ if and only if, for any distribution ξ and number $x \in [0,1]$, $x\xi + (1-x)\mu \succsim x\xi + (1-x)\nu$.

SIMPLE EUT If p and q are simple prospects, then $p \succsim q$ if and only if the expected utility of p is at least as great as that of q.

2.6 Appendix: Proof of Consistency of DTU

We now prove the main result (Theorem 1) and some other theorems. We proceed by extending the proof of Section 2.3 with an ultrafilter. It is also possible to give a more abstract proof by adapting the technique of Fine 2008 to accomodate Stochastic Dominance and applying Zorn's lemma directly, however the concrete models generated by the truncated expectation ordering are particularly natural (they are, for example, anticipated by Easwaran 2014b) and straightforward, and are sufficiently modular to achieve all desired independence and consistency results.

Definition 1 (Cobounded ultrafilter)**.** A *cobounded ultrafilter* on $[0,\infty)$ is a set of subsets of $[0,\infty)$, $\mathcal{U}$, such that:

- If $X \subseteq Y$ and $X \in \mathcal{U}$, then $Y \in \mathcal{U}$.
- If $X \in \mathcal{U}$ and $Y \in \mathcal{U}$, then $X \cap Y \in \mathcal{U}$.
- For every $X \subseteq [0,\infty)$, either $X \in \mathcal{U}$ or $[0,\infty) \setminus X \in \mathcal{U}$.
- $\mathcal{U}$ contains no bounded sets.

Lemma 4 (Ultrafilter lemma)**.** *There is a cobounded ultrafilter.*

Proof. Standard (e.g., Jech 2006, Theorem 7.5). □

In what follows, let $\mathcal{U}$ be a cobounded ultrafilter that is held fixed throughout. Roughly speaking, we will extend $\sqsupseteq$ by adding $p \sqsupseteq q$ when $\mathbf{E}_t(p)$ exceeds $\mathbf{E}_t(q)$ on a set in $\mathcal{U}$ (some strict comparisons become nonstrict). The following standard definition will make it easier to show that the resultant ordering has the desired properties:

Definition 2. For functions f and g from $[0,\infty)$ to $\mathbf{R}$, say $f \sim^*_{\mathcal{U}} g$ if f and g are the same on a member of $\mathcal{U}$. Let the set of *hyperreals* be the set of equivalence classes of functions from $[0,\infty)$ to $\mathbf{R}$ modulo $\sim^*_{\mathcal{U}}$. For such a function f, let $[f]$ be the corresponding hyperreal number.

Lemma 5. *The hyperreals form an ordered field under pointwise addition, pointwise multiplication, and the total ordering $\geq$, defined so that $[f] \geq [g]$ if and only if the set of points x for which $f(x) \geq g(x)$ is in $\mathcal{U}$.*

Proof. Standard (see also Lemma 6). $\square$

Theorem 1. DTU *is consistent.*

Proof. We will modify the model defined for Theorem 2 by replacing $\sqsupseteq$ with $\sqsupseteq^*$, defined as:

$$\mu \sqsupseteq^* \nu := [\mathbf{E}_t(\mu)] \geq [\mathbf{E}_t(\nu)]$$

Now we check that each of the desired principles holds in the model. Simple EUT holds for the same reasoning as before. It remains to check that $\sqsupseteq^*$ is a total preordering that satisfies Independence and Stochastic Dominance.

Totality and Preordering. Lemma 5.

Independence. For the same reasons as in the proof of Theorem 2, we have:

$$\mathbf{E}_t(x\xi + (1-x)\mu) = x\mathbf{E}_t\xi + (1-x)\mathbf{E}_t\mu$$

And by Lemma 5 we have

$$[x\mathbf{E}_t\xi + (1-x)\mathbf{E}_t\mu] = x[\mathbf{E}_t\xi] + (1-x)[\mathbf{E}_t\mu]$$

Now, $\mu \sqsupseteq^* \nu$ is equivalent to $[\mathbf{E}_t\mu] \geq [\mathbf{E}_t\nu]$, and so is also equivalent to

$$[\mathbf{E}_t(x\xi + (1-x)\mu)] \geq [\mathbf{E}_t(x\xi + (1-x)\nu)],$$

which is equivalent to $x\xi + (1-x)\mu \sqsupseteq^* x\xi + (1-x)\nu$.

Stochastic Dominance. For Theorem 2 we showed that if μ dominates ν, then $\mathbf{E}_t\mu \geq \mathbf{E}_t\nu$ for all t in $[0,\infty)$. And $[0,\infty) \in \mathcal{U}$, so $\mu \sqsupseteq^* \nu$. And if μ strictly dominates then for some $s > 0$, $\mathbf{E}_t\mu > \mathbf{E}_t\nu$ for all $t > s$, and the set of points greater than s is in $\mathcal{U}$ because it is cobounded, so $\mu \sqsupseteq^* \nu$. $\square$

Theorem 3. *EUT is independent of* DTU.[34]

This theorem requires some more definitions.

Definition 3 (Continuous Hyperreal)**.** For functions $f : [0,\infty) \to \mathbf{R}$ and $g : [0,\infty) \to \mathbf{R}$, say that $f \sim^c_{\mathcal{U}} g$ if, for any $\varepsilon > 0$, the set of points x for which $|f(x) - g(x)| < \varepsilon$ is a member of $\mathcal{U}$. Then, let the *continuous hyperreals* be the

[34]Proof in Appendix 2.6.

set of functions modulo $\sim^c_{\mathcal{U}}$. Let $[f]^c$ be the continuous hyperreal of which f is a member. The reals can be identified with the equivalence classes of constant functions $[y \mapsto x]^c$.

Lemma 6. *The continuous hyperreals form an ordered ring under pointwise addition, pointwise multiplication, and the total ordering $\leq$, defined so that $[f]^c \leq [g]^c$ if and only if for every positive real number ε, the set of points x for which $f(x)$ is no greater than $g(x) + \varepsilon$ is a member of $\mathcal{U}$.*

Proof. First we must check that the pointwise operations are well-defined on the continuous hyperreals: that if $[f] = [g]$ and $[h] = [i]$, then $[f] + [h] = [g] + [i]$ and $[f] \cdot [h] = [g] \cdot [i]$. To see that this is so, fix some ε and let X and Y respectively be the sets on which f and h and g and i are within $\varepsilon/2$ of each other in the case of addition, and within $\sqrt{\varepsilon}$ in the case of multiplication. Then, since $X \cap Y$ is a member of $\mathcal{U}$, the operations work the same on different members of the same equivalence class. That the continuous hyperreals form a ring follows from the fact that the pointwise operations make the underlying set of functions into a ring.

That $\leq$ is reflexive is obvious, and that $\leq$ is transitive follows from the fact that if $X \in \mathcal{U}$ is the set on which $f + \varepsilon/2$ is at least as great as g, and $Y \in \mathcal{U}$ is the set on which $g + \varepsilon/2$ is as great as h, then $f + \varepsilon$ is at least as great as h on $X \cap Y$, which is a member of $\mathcal{U}$. Totality follows from the maximality property of ultrafilters. The ordered ring properties follow from the properties of real addition and multiplication. □

Proof of Theorem 3. To show that EUT does not follow from DTU, notice that the ordering in Theorem 1 extends the ordering in Theorem 2, so also violates EUT.

To show that EUT is consistent, we use the same proof as Theorem 1 but with $\sqsupseteq^*$ replaced with $\sqsupseteq^c$, defined as:

$$\mu \sqsupseteq^c \nu := [\mathbf{E}_t\mu]^c \geq [\mathbf{E}_t\nu]^c$$

We then prove that the resulting model satisfies the desired principles. Things go the same as before with one exception, which is Stochastic Dominance. For this, it is slightly harder to show that when μ strictly dominates ν it is strictly better, because in general $\mathbf{E}_t\mu$ can always strictly exceed $\mathbf{E}_t\nu$, but by a vanishing amount, in which case $[\mathbf{E}_t\mu]^c = [\mathbf{E}_t\nu]^c$. However in the case where μ strictly dominates ν, there is some utility s such that μ is strictly more likely than ν to yield an outcome at least as good as s. It follows that for $t > s$, there is some $\delta > 0$ such that $\mathbf{E}_t\mu > \mathbf{E}_t\nu + \delta$, so in fact $[\mathbf{E}_t\mu]^c > [\mathbf{E}_t\nu]^c$ as required.

It remains only to show that EUT holds in the model. This corresponds to the condition that when μ and ν have expectations, $\mu \sqsupseteq^c \nu$ if and only if $\mathbf{E}\mu \geq \mathbf{E}\nu$. So suppose that μ does have an expectation. Then

$$\mathbf{E}\mu := \lim_{x\to\infty} \mathbf{E}_t\mu$$

Which is to say that, for all $\varepsilon > 0$, there is an $s > 0$ such that for $t > s$, $|\mathbf{E}_t\mu - \mathbf{E}\mu| < \varepsilon$. Therefore, $[\mathbf{E}_t\mu]^c = \mathbf{E}\mu$, so in general if ν also has an expected utility, then $\mu \sqsupseteq^c \nu$ if and only if $\mathbf{E}\mu \geq \mathbf{E}\nu$. □

Remark 1. *For any unbounded subset $X \subseteq [0, \infty)$, there is a cobounded ultrafilter that contains X.*

Proof. Jech 2006, Theorem 7.5. □

Corollary 7. *Thus, if $\mathbf{E}_t\mu \geq \mathbf{E}_t\nu$ on an unbounded set, it is consistent with* DTU *that μ is at least as good as ν.*

Corollary 8. *Since $\mathbf{E}_t\mu \geq \mathbf{E}_t\nu$ on an unbounded set if and only if $\neg(\mu \sqsupseteq \nu)$, it follows that if μ is not at least as good as ν in the model constructed for Theorem 2, it is consistent in* DTU *that ν is at least as good as μ.*

Remark 2. *In the above proofs, $\mathbf{E}_t\mu$ can be replaced with $\mathbf{E}_t^{fg}\mu$, defined as*

$$\int_{-g(t)}^{f(t)} z\, d\mu(z)$$

for any two nonnegative, increasing, and unbounded functions f and g, and the resulting ordering of prospects still satisfies the relevant combination of principles.

Corollary 9. *Let the symmetric St Petersburg prospect be that which yields utilities 2 and -2 each with probability $1/4$, 4 and -4 with probabilities $1/8$, and in general utilities 2^n and -2^n each with probabilities $1/2^{(n+1)}$ (see Table 2.4). By setting f and g appropriately, it is possible to build models for Theorem 3 where:*

- *The symmetric St Petersburg prospect is better than every outcome.*
- *The symmetric St Petersburg prospect is worse than every outcome.*
- *For any particular outcome x, the symmetric St Petersburg prospect is equally good as x. (If x isn't 0, then Reflection Anti-Invariance is false in the model.)*

Theorem 10. *Reflection Anti-Invariance holds in the models constructed for* DTU*,* DU*, and* DTU + *EUT.*

$1/4$	$1/4$	$1/8$	$1/8$	$\dots$	$\dots$	$2^{-(n+1)}$	$2^{-(n+1)}$	$\dots$	$\dots$
2	-2	4	-4	$\dots$	$\dots$	2^n	-2^n	$\dots$	$\dots$

Table 2.4: The symmetric St Petersburg prospect.

Theorem 11. *Shift Invariance holds in the models constructed for* DU *and* DTU + *EUT.*

Proof. For Shift Invariance in the model of DTU+EUT, write $\mu_{+\beta}$ for the distribution that gives $X + \beta$ (i.e., $\{x + \beta : x \in X\}$) the same probability as $\mu(X)$. We have

$$\begin{aligned}
\mathbf{E}_t\mu_{+\beta} &= \int_{-t}^{t} z \,\mathrm{d}\,\mu_{+\beta}(z) \\
&= \int_{-t-\beta}^{t-\beta} z + \beta \,\mathrm{d}\,\mu(z) \\
&= \int_{-t}^{t} z + \beta \,\mathrm{d}\,\mu(z) + \int_{-t-\beta}^{-t} z + \beta \,\mathrm{d}\,\mu(z) - \int_{t-\beta}^{t} z + \beta \,\mathrm{d}\,\mu(z) \\
&= \mathbf{E}_t\mu + \int_{-t}^{t} \beta \,\mathrm{d}\,\mu(z) + \left(\int_{-t-\beta}^{-t} z \,\mathrm{d}\,\mu(z) - \int_{t-\beta}^{t} z \,\mathrm{d}\,\mu(z) \right)
\end{aligned}$$

The second term in the last line tends to β, since μ is a probability distribution, and third term tends to 0, because it is bounded in absolute value above by $t\beta\mu[t-\beta, t)$. Hence $[\mathbf{E}_t\mu_{+\beta}]^c = [\mathbf{E}_t\mu]^c + \beta$. □

Theorem 12. *Scale Invariance is true in the model constructed for* DU.

Theorem 13. *Scale Invariance fails in all models of* DTU *constructed above.*

Proof. Write μ_2 for the distribution satisfying:

$$\mu_2(X) = \mu(X \div 2)$$

For all measurable X. μ_2 is the distribution corresponding to multiplying each possible utility by 2.

Notice that $\mathbf{E}_t\mu_2 = \mathbf{E}_{2t}\mu$. Hence for a counterexample to Scale Invariance in the model it will suffice to find μ and ν such that for some δ, the sets

$$\{t \in \mathbf{R}^+ : \mathbf{E}_t\mu \geq \mathbf{E}_t\nu + \delta\}$$

and

$$\{t \in \mathbf{R}^+ : \mathbf{E}_{2t}\nu \geq \mathbf{E}_{2t}\mu + \delta\}$$

are both in $\mathcal{U}$.

To this end, notice also that a function $f(t) : \mathbf{R}^+ \to \mathbf{R}^+$ is $\mathbf{E}_t\mu$ for some μ when f is nondecreasing, continuous, differentiable almost everywhere, with a nonincreasing derivative bounded above by 1, which tends towards 0 in the limit. So it suffices to find f and g with this property so that the set on which $f(t) \geq g(t) + \delta$ and the set on which $g(2t) \geq f(2t) + \delta$ are both in $\mathcal{U}$.

Set $f(t)$ and $g(t)$ as follows:

$$f(t) := \sqrt{t+1}$$
$$g(t) := \begin{cases} \sqrt{t+1} & 0 \leq t < 1 \\ \sqrt{t+1} + \frac{\sin(\pi \log_2 t)}{1000} & t \geq 1 \end{cases}$$

We have for $t \geq 1$,

$$\begin{aligned} g(t) - f(t) &= \frac{\sin\left(\pi \log_2 t\right)}{1000} \\ g(2t) - f(2t) &= \frac{\sin\left(\pi \log_2 2t\right)}{1000} \\ &= -\frac{\sin\left(\pi \log_2 t\right)}{1000} \end{aligned}$$

So for any $\delta > 0$, the set of points $t \geq 1$ for which $f(t) \geq g(t) + \delta$ and $g(2t) \geq f(2t) + \delta$ are the same, and are both unbounded.

Next we must show that $f(t)$ and $g(t)$ are indeed $\mathbf{E}_t\mu$ and $\mathbf{E}_t\nu$ for some distributions μ and ν. For this, notice that both are continuous, and both are smooth everywhere besides 1, with derivatives and second derivatives as follows:

$$\begin{aligned} f'(t) &= \frac{1}{2\sqrt{t+1}} \\ f''(t) &= \frac{-1}{4\sqrt[3]{t+1}} \\ g'(t) &= \frac{1}{2\sqrt{t+1}} + \frac{\pi \cos\left(\pi \log_2 t\right)}{1000(\ln 2)t} \\ g''(t) &= \frac{-1}{4\sqrt[3]{t+1}} + \frac{-\pi^2(\ln 2)\sin\left(\pi \log_2 t\right) - \pi \cos\left(\pi \log_2 t\right)}{1000(\ln 2)t^2} \end{aligned}$$

f' is bounded above by 1, tends towards 0, and is positive. f'' is negative so f' is always decreasing, so $f(t)$ is indeed $\mathbf{E}_t\mu$ for some μ. For g, notice that for $t \geq 1$ the numerator of the added term decays like $1/t$ for g' and like $1/t^2$ for g'', which is much more rapidly than $1/\sqrt{t}$ or $1/\sqrt[3]{t}$, so $g(t)$ is $\mathbf{E}_t\nu$ for some ν.

What we have shown is that for any $\delta > 0$, $f(t) \geq g(t) + \delta$ if and only if $g(2t) \geq f(2t) + \delta$. Hence if the set of such t is in $\mathcal{U}$, we have a counterexample to Scale Invariance. Suppose otherwise, then for every ε the set of t such that $|f(t) - g(t)| < \varepsilon$ is in $\mathcal{U}$. For a given ε, this set is given by

$$\{t \in \mathbf{R}^+ : |\sin(\pi \log_2 t)| < 1000\varepsilon\}$$

Then, we may repeat the above reasoning, now comparing $f(t)$ with $h(t)$, defined as follows:

$$h(t) := \begin{cases} \sqrt{t+1} & 0 \leq t < 2 \\ \sqrt{t+1} + \frac{\cos(\pi \log_2 t)}{1000} & t \geq 2 \end{cases}$$

And by the same argument we must conclude the following set to be in $\mathcal{U}$ if there is no counterexample to Scale Invariance:

$$\{t \in \mathbf{R}^+ : |\cos(\pi \log_2 t)| < 1000\varepsilon\}$$

But sin and cos are out of phase: sin takes a value near 0 when cos takes an extremal value, and vice-versa. So the two sets are disjoint (for small values of ε). So in any case there is a counterexample to Scale Independence in the model. □

Chapter 3

Adding Lotteries

3.1 Introduction

Orthodoxy in decision theory proceeds by assigning real-valued *utilities* to outcomes and by comparing risky prospects by the *expected utility* of the outcomes those prospects might yield. For example, a fair coin toss between outcomes of utilities 2 and 8 has an expected utility given by $0.5 \times 2 + 0.5 \times 8 = 5$, so is equally good as a certain outcome of utility 5.

Supposing there are outcomes of arbitrarily great utility (positive or negative), not every prospect has an expected utility. The canonical witness to this fact is the *St Petersburg prospect*, which for each number $n \geq 1$ yields utility 2^n with probability $1/2^n$. Its expected utility would be given by the sum

$$\sum_{n=1}^{\infty} \frac{2^n}{2^n}$$

but the sum diverges to infinity. The question arises of how to compare prospects, like the St Petersburg prospect, that lack expected utilities. Recent literature has witnessed an explosion of decision theories that are purported to give more satisfying results where expected utilities are not present, while respecting the comparison by expected utility when expected utilities do exist.[1]

This paper considers an especially intriguing cluster of views developed by Colyvan 2008, Seidenfeld, Schervish, and Kadane 2009, Colyvan and Hájek 2016, Bartha 2016, Lauwers 2016, and Meacham 2019. The very rough idea these authors share is that a pair of prospects can be compared by looking at the difference in

[1]Some contributions to this literature not discussed here include Easwaran 2008, N. J. J. Smith 2014, Gwiazda 2014, Easwaran 2014b, Hájek 2014.

utility that the prospects might yield in various situations.

Colyvan's *Relative Expectation Theory* is probably the best known of these. It says that to compare prospects X and Y, take the expected *difference* in utility that you would get from taking X over Y. If in expectation the difference is positive, then X is better. For example, if X would give utility 10 if a coin lands heads and 0 if it lands tails, and Y would give utility 6 on heads and 2 on tails, then in expectation the improvement of X on Y is by 1 point of utility. 1 is positive, so X is better, according to Colyvan.

	Heads	Tails
X	10	0
Y	6	2
$X - Y$	4	-2

Notice that Colyvan's theory does not require X or Y to have expected utilities individually to yield a comparison between them; which is how the theory advances on standard Expected Utility Theory in cases of undefined expected utility.

Seidenfeld, Schervish, and Kadane (SSK) purify the idea further with a principle we shall call *Addition Invariance*.[2] Instead of appealing to the expected difference between X and Y in their comparison, Addition Invariance says that X and Y can be compared by treating their difference, $X - Y$, as a prospect in itself. X is better than Y if and only if the prospect $X - Y$ is better than a sure utility 0. Supposing that a prospect with higher expected utility is always better, Addition Invariance agrees with Colyvan's Relative Expectation Theory, whenever Relative Expectation Theory delivers a verdict. However, just as X and Y might lack expected utilities, so might $X - Y$. In that case, Relative Expectation Theory falls silent, whereas Addition Invariance can still be applied.

This paper investigates the consequences of Addition Invariance for decision theory, and lays out some considerations for assessing whether the principle is true. Addition Invariance is shown to have a wide range of welcome consequences, bringing order to decision theory in cases where expected utilities are undefined. This fact, when combined with the naturalness of Addition Invariance, suggests that Addition Invariance can unify and explain significant parts of our decision-theoretic knowledge, and the principle becomes more tempting as a result.

However, SSK have shown that Addition Invariance conflicts with another fundamental principle of decision theory: *Stochastic Equivalence*, which says that the only thing that matters in the evaluation of a prospect is the probabilities of the different outcomes you could get if you took it. The incompatibility with Stochastic

[2] SSK isolate a weaker principle which they call "Coherent Indifference"; we shall see that the motivation they give for Coherent Indifference also strongly suggests Addition Invariance.

Equivalence is a major cost for theories with Addition Invariance, but it is a cost that SSK and others have accepted. Here, a more thorough assessment of the costs and benefits of Addition Invariance and Stochastic Equivalence tips the scales against Addition Invariance.

3.2 Formalism

For the investigation of Addition Invariance we adopt a Savage-style framework for decision theory, which treats the prospects as measurable functions from states to outcomes. The probability space of states, $\mathcal{S} = \langle \Omega, \mathcal{E}, \mathbf{P} \rangle$, is assumed to to be isomorphic to Lebesgue measure on the unit interval. It will be helpful to think of the state-space as determined by an infinite sequence of fair and independent coin tosses, with states being maximally specific ways for the tosses to land, and $\mathbf{P}$ the usual probability measure for fair and independent coin tosses.

The space of outcomes is identified with the measurable space of real numbers, where the number x is identified with the outcome of utility x. Prospects are identified with measurable functions from Ω to $\mathbf{R}$, or real-valued random variables on $\mathcal{S}$. We conventionally use upper-case roman letters (e.g., X, Y, Z) to range over random variables.

As is well-known, the Savage framework of states and outcomes bakes in the idealizing assumption that there is a set of possibilites called "states" that determine how each prospect would go if it were taken, such that there are prospects to be taken corresponding to arbitrary measurable functions from states to outcomes. We have also significantly idealized by assuming the utilities of outcomes to span exactly the real numbers, so that there are no gaps in utility and also no infinite or infinitesimal levels of value. These assumptions are helpful for stating and reasoning with Addition Invariance, and are mostly harmless in that it is very plausible that there is some restricted domain of prospects where they are true. Section 3.4 investigates Addition Invariance in a more realistic setting.

$\succsim$, a binary relation between prospects, will be the *at least as good as* relation of decision theory. To investigate SSK's proposed addition to decision theory we shall take for granted the following axioms:

Statewise Dominance If $X(s) \geq Y(s)$ for all outcomes s in some set of probability 1, then $X \succsim Y$. If, in addition, $X(s) > Y(s)$ for all s in some set of nonzero probability, then $X \succ Y$.[3]

[3]SSK only assume a weaker dominance principle:

Coherent Strict Preference If for some $\varepsilon > 0$ $X(s) \geq Y(s) + \varepsilon$ for all outcomes s, then $X \succ Y$.

Preordering $\succsim$ is a preordering. That is, for all prospects X, Y, Z, $X \succsim X$ (reflexivity), and $X \succsim Y \succsim Z$ implies $X \succsim Z$ (transitivity).[4]

Simple Prospect Totality If X and Y could each yield one of only finitely may outcomes (i.e., if the range of both random variables is finite), then $X \succsim Y$ or $Y \succsim X$.

For the final principle, let a *utility-defining* prospect be one that, for some number $x \geq 1$, yields utility x with probability $1/x$ and 0 with probability $1 - 1/x$. The name is chosen because in settings where utility is not taken as primitive but is instead defined relative to an arbitrary zero point and unit of utility, an outcome's having utility x with $x \geq 1$ is usually defined as being such that a $1/x$ chance of that outcome and a $1 - 1/x$ chance of a the zero point is equally good as the unit. So we take for granted:

Minimal Expected Utility Theory If X is utility-defining, then $X \sim \mathbf{1}$ ($\mathbf{1}$ is the identically 1 random variable).

SSK's novel axiom is what they call *Coherent Indifference*:

Coherent Indifference If $X \sim Y$ then $X - Y \sim \mathbf{0}$ ($-$ is statewise subtraction and and $\mathbf{0}$ is the identically 0 random variable).

Here, we shall focus primarily on a principle that strengthens the conditional in Coherent Indifference to a biconditional, and which replaces the equal value relation $\sim$ with $\succsim$.

Addition Invariance $X \succsim Y$ if and only if $X - Y \succsim \mathbf{0}$.[5]

Equivalently:

Addition Invariance (alt. formulation) $X \succsim Y$ if and only if $X + Z \succsim Y + Z$.

The motivation SSK give for Coherent Indifference is the following:

> [Coherent Indifference] expresses the idea that when variables X and Y are indifferent there is no value in selling one for gaining the other. Such a trade is judged indifferent with the status-quo. (p. 330)

[4]SSK instead take the strict relation $\prec$, *worse than*, as primitive, and they assume it to be a total strict preordering. Interpreting $\succsim$ as $\not\prec$ yields a *total* preordering—we assume a mere preordering here.

[5]Thus formulated, it may seem that Addition Invariance illicitly assumes that the zero point on the scale of utility has special significance. In fact the opposite is the case, as discussed in Section 3.4.2.

This sort of motivation also extends to Addition Invariance, which plausibly adds that if there *is* value for selling one for the other, then the one you should sell is worse than the one you should buy. Addition Invariance is a particularly natural extension: we are thinking of prospects as a vector space of functions under statewise addition of utilities, and Addition Invariance says exactly that the ordering $\succsim$ is compatible with this vector space structure. Whereas Coherent Indifference already gives rise to the impossibility results SSK are interested in, Addition Invariance is very plausibly the principle that decision theorists like SSK should accept as a fundamental axiom.

Let the *base theory* be the conjunction of Preordering, Simple Prospect Totality, Minimal Expected Utility Theory, and Statewise Dominance. We will investigate Addition Invariance by analysing and comparing the effect of adding Addition Invariance or its competitors to the base theory.

3.3 Addition Invariant Decision Theory

In the base theory, Addition Invariance has many desirable consequences, which otherwise would have to be taken as primitive assumptions. Addition Invariance therefore promises to play an important unifying role in axiomatic decision theory. This section illustrates how some widely accepted but nevertheless occasionally controversial principles follow from Addition Invariance.

3.3.1 The Sure-Thing Principle

First we shall consider Savage's *Sure-Thing Principle*. For an event $E \in \mathcal{E}$, we write $X \upharpoonright E$ for the *restriction of* X *to* E, which yields the same result as X on each state in E, and otherwise yields utility 0.

Sure-Thing Where E is an event with $0 < \mathbf{P}(E) < 1$, if $X \upharpoonright E \sim Y \upharpoonright E$, then $X \succsim Y$ if and only if $X \upharpoonright E^c \succsim Y \upharpoonright E^c$.

Theorem 14. *Sure-Thing follows from Addition Invariance.*

Proof sketch. $X = X \upharpoonright E + X \upharpoonright E^c$ and $Y = Y \upharpoonright E + Y \upharpoonright E^c$. □

The Sure-Thing principle plays an important role in decision theory. To take an example emphasised by Hájek and Smithson (2012), suppose that you are to choose between ordering pizza or chinese food for dinner. Both options are risky, in the sense that there will be nontrivial probabilities for both options regarding how good the pizza or chinese food is, how quickly it will be delivered, and so on. There may also be a nonzero probability that a meteorite will end life on Earth before the food arrives, conditional on which both options are equally good. The Sure-Thing principle says that the comparison between the two options is only sensitive to what their payoffs would be in cases where they would yield different payoffs. Possibilities where they are equally good—such as if the meteorite wipes us all out before the food arrives—may be ignored for the sake of the comparison.

The Sure-Thing principle is especially theoretically important where the prospects being compared have no well-defined expected utility. Suppose that in the previous example, conditional on the meteorite striking Earth the two gambles yield utilities with the probabilities in the St Petersburg prospect (perhaps because the meteorite sends us all to Heaven for a number of days determined by some divine coin-tosses). In this case, then no matter how unlikely the meteorite is to strike Earth, both prospects lack an expected utility. We need something like the Sure-Thing principle to correctly tell us that we may decide on pizza or Chinese food based on which we prefer, given that neither affects the probability or severity of the meteorite. With Addition Invariance, the Sure-Thing principle is derived, rather than taken as an axiom as in Savage's work.

3.3.2 Relative Expectation Theory

Savage (following von Neumann and Morgenstern) put the Sure-Thing principle to work in deriving a *representation theorem*; a theorem to the effect that there is a way of assigning utilities to outcomes so that prospects may be compared by their expected utilities. Addition Invariance performs that work in the base theory, which recall contains assumes the very weak *Minimal* Expected Utility Theory. To be more precise, the base theory plus Addition Invariance yields the restriction of Colyvan's Relative Expectation Theory to pairs of prospects whose difference can take one of only finitely many values:

SIMPLE RELATIVE EXPECTATION THEORY $X - Y$ can take only finitely many values (i.e., if its range is finite), then $X \succsim Y$ if and only if $\mathbf{E}(X - Y) \geq 0$.

Theorem 15. *Simple Relative Expectation Theory follows from Addition Invariance.*

Proof sketch. We show that if a prospect X can take finitely many values, then $X \succsim 0$ if and only if $\mathbf{E}X \geq 0$, whence the result will follow by Addition Invariance. The proof is by induction on the number of possible outcomes of X. The claim follows from Dominance when X has only one nonzero outcome. When X has two or more nonzero outcomes x and y with probabilities p_x and p_y, we may replace y with zero by adding a compensating amount of $p_y y / p_x$ to x, so the result follows by induction. □

Simple Relative Expectation Theory is importantly weaker than the full Relative Expecation Theory that follows in SSK's system and that is propounded by Colyvan.

The latter theory can be regarded as a limiting case of the former. The full Relative Expectation Theory does not follow from Addition Invariance, and instead requires an additional principle asserting the continuity of value with respect to the topology of convergence in expectation:

L^1 CONTINUITY If $\mathbf{E}(X_i - X)$ goes to 0 as i goes to infinity, then if $Y \succ X$ there is a number N such that for $i > N$, $Y \succ X_i$.

Theorem 16. *L^1 Continuity and Relative Expectation Theory are equivalent, given Addition Invariance.*

3.3.3 Affine Symmetries

The third class of desirable principles that follow from Addition Invariance consists of three principles of symmetry:

SHIFT INVARIANCE For any real number c, $X \succsim Y$ if and only if $X + \mathbf{c} \succsim Y + \mathbf{c}$.

RATIONAL SCALE INVARIANCE For a a positive rational number, $X \succsim Y$ if and only if $aX \succsim aY$.

REFLECTION ANTI-INVARIANCE $X \succsim Y$ if and only if $-Y \succsim -X$.

The principles assert the invariance of the at least as good as ordering under the operations of increasing the utility of all the possible payoffs by the same amount c, and of multiplying the utility of all possible payoffs by a positive rational number a (the first principle is a trivial consequence of Addition Invariance; the second requires Totality).

Shift Invariance and Rational Scale Invariance are not often discussed, because they add nothing in cases where expected utilities are well-defined. However, when expected utilities are not defined, the principles are no longer so obvious, and very similar principles have been put to work in this setting by Alan Hájek (2014), Kenny Easwaran (2014b) and myself in Chapter 4. There, the principles are primitive assumptions to be justified by philosophical argument. For someone who accepts Addition Invariance, no further argument is required, for they are already theorems. (A tempting generalization of Rational Scale Invariance permits a to be positive irrational as well; this stronger scale invariance principle does not seem to follow from Addition Invariance.)

3.3.4 The Strength of Addition Invariance: Some Independence Results

We have seen four desirable consequences of Addition Invariance in the very weak base theory: the Sure-Thing principle, Simple Relative Expectation Theory (which when augmented with L^1 Continuity yields Colyvan's Relative Expectation Theory), as well as the Shift and Rational Scale Invariance principles. Addition Invariance unifies these principles in a satisfying way. This section reports a theorem that emphasises the potential importance of Addition Invariance in justifying these principles. The result says that if Addition Invariance is denied, no two of the principles entail the other two. Therefore, at least three of the four principles must be taken as primitive if they are to be included in the overall theory.

Theorem 17. *(a) Sure-Thing is independent of Relative Expectation Theory plus Shift and Scale Invariance.*

(b) Simple Relative Expectation Theory is independent of Shift and Scale Invariance, and independent of Sure-Thing plus Rational Scale Invariance.

(c) Shift Invariance is independent of Sure-Thing plus Rational Scale Invariance.

(d) Rational Scale Invariance is independent of Sure-Thing plus Relative Expectation Theory plus Shift Invariance.

(Note that Simple Relative Expectation Theory follows from Sure-Thing plus Shift Invariance, and Shift Invariance follows from Simple Relative Expectation Theory.)

The preceding results suggest Addition Invariance as an initially plausible and powerfully unifying addition to the base theory. It is an optional addition, since it does not follow from any combination of other principles mentioned thus far:

Theorem 18. *Addition Invariance is independent of Sure-Thing plus Relative Expectation Theory plus Shift and Rational Scale Invariance.*

3.4 Addition Invariance With Less Idealization

The Savage-style framework used so far, wherein prospects are identified with functions from some state space Ω to real numbers, presents Addition Invariance as an extremely natural principle. In assuming a fixed state space and the identification of outcomes with real numbers, the framework builds in two significant idealizations that obscure some difficulties for formulating Addition Invariance. This section makes these idealizations explicit and assesses the prospects for Addition Invariance when they are avoided.

3.4.1 Statewise Operations

Savage's formalism builds in a set of *states* which determine the outcome of any possible prospect. It is often useful to imagine prospects as gambles on the result of some chancy process, like a sequence of coin tosses. Since in this case the outcome of any prospect is determined by the result of the coin tosses, states may be identified with maximally specific propositions regarding how the coin tosses land. Consequently, the statewise difference of prospects is easy to compute: $X - Y$ is the gamble which pays utility $x - y$ when the coins land such that X would pay utility x and Y would pay utility y.

In reality, the consequences of our risky actions are not determined by any sequence of coin tosses. Uncertainty in the consequences of our actions can come from uncertainty about the weather, about people's moods, about the trajectory of an asteroid, or anything else. This being the case, it is difficult to identify any one random process that determines the outcomes of all prospects, whose maximally specific results could be called states. But without such a set of states Addition Invariance cannot be applied, because it requires prospects to be compared by subtracting them state-by-state. The question arises of whether Addition Invariance is such a natural principle after all, or if it only seems natural as an artefact of the idealizations built into Savage's formalism.

This question can be answered by formulating a decision theory which does not rely on a primitive notion of state. The purpose of states in Savage's formalism is to keep track of what outcome taking a prospect *would* result in. But although this indexation of counterfactual consequences by states is handy, we need not take it for granted. Instead, we may take *would* as primitive instead, or more precisely a binary sentential operator '$\Box\!\!\rightarrow$' that formalizes the counterfactual conditional relevant to decision theory and a unary sentential operator '$\Box$' which formalizes necessity. On this formalization, prospects and outcomes are types of proposition—a prospect is a proposition that intuitively says you take a certain action, and an outcome a proposition which says exactly how good things are.

Savage's axioms may now be reformulated without appeal to states. Assuming a utility function U which maps outcomes to real numbers, a state-free version of Statewise Dominance may be formulated as follows: where A and B are prospects, we say that A (weakly) dominates B if

$$\Box\forall o, o' \in \mathcal{O}.((A \Box\!\!\rightarrow o \land B \Box\!\!\rightarrow o') \rightarrow U(o) \geq U(o')),$$

which is to say, it is necessary that the outcome A would result in is at least as good as the outcome B would result in. (Redefining strict dominance, and stating other principles like Sure-Thing, is similarly easy but requires the introduction of a propositional probability function.)

Similarly, Addition Invariance may be stated by redefining a sum of prospects A and B to be any prospect C satisfying

$$\Box\forall o, o', o'' \in \mathcal{O}.((A \mathrel{\Box\!\!\rightarrow} o \wedge B \mathrel{\Box\!\!\rightarrow} o' \wedge C \mathrel{\Box\!\!\rightarrow} o'') \rightarrow (U(o) + U(o') = U(o'')))$$

Assuming a sum of any two prospects exists, Addition Invariance may now be asserted using the new definition of addition instead of any notion of state.

This reformulation shows that Addition Invariance does not necessarily trade on any idealizations built into Savage's notion of state. However, the reformulation also reveals an important assumption elided in SSK's use of Savage's formalism, namely that the addition or subtraction of two prospects always exists.

When prospects are arbitrary real-valued random variables on the space of states, the existence of $X + Y$ or $X - Y$ is a trivial, but in the state-free formulation special assumptions will be needed to ensure that for any two prospects have a sum or difference. What exactly these assumptions would have to be depends on how the concept of prospect is cashed out in the state-free formalism.

One natural conception of a prospect is a proposition that you by your actions can make true; these are propositions such as *you go swimming*, *you go shopping*, and presumably not things like *Caesar is assassinated* or *everyone is happy forevermore*. On this conception it seems very implausible that there would be a sum or difference of any two prospects. For example, you might be able to go swimming and able to go shopping, but not able to do anything which necessarily results in an outcome which adds the utility that would result from swimming to the utility that would result from shopping. In general, if there is anything you can do which has a chance of nonzero utility, adding that prospect to itself over and over would require you to be able to do infinitely many things, with more and more extreme possible outcomes. No human has that sort of power.

A more permissive conception of prospect has it that any proposition whatsoever is a prospect. You may have no control over whether Caesar is assassinated, but it is a prospect nevertheless on this conception, so decision theory will say whether it is better for Caesar to be assassinated than for you to go swimming now. Still it is impossible, however, for every pair of prospects to have a sum and a difference, assuming the following two axioms of counterfactual logic (on top of S4 for $\Box$):

Actual Outcome Necessarily, if p is true, then the outcome that would result if p were true is the outcome that is true.

$$\Box\forall o \in \mathcal{O}.\forall p.((o \wedge p) \rightarrow (p \mathrel{\Box\!\!\rightarrow} o)).$$

Compossible Outcome If p is not compatible with an outcome o, then o would not be the result of p.

$$\Box(\Box(p \to \neg o) \to \neg(p \mathrel{\Box\!\!\to} o)).$$

To see why, consider the tautologous proposition $\top$. By Actual Outcome, whichever outcome will in fact result is the one that would result if $\top$ were true. So if there is such a prospect as $\top + \top$, it must be a prospect which, if the utility of the outcome that will in fact result is x, would yield $x + x$ were it true.

$$\Box\forall o, o' \in \mathcal{O}.(o \to ((\top + \top) \mathrel{\Box\!\!\to} o') \to (2U(o) = U(o'))).$$

By Actual Outcome, $\top + \top$ can then only be compatible with outcomes of utility 0, so by Compossible Outcome it is necessary that $\top + \top$ would result in utility 0. But by definition, this means that $\top$ would necessarily result in utility 0 as well, but since $\top$ is compatible with anything possible, Actual Outcome will imply that every outcome has utility 0, which is absurd.

To ensure the existence of sums and differences of prospects, then, the class of prospects must lie between the class of all propositions and the class of propositions that we ourselves can make true. It is unclear what exactly the relevant restriction on propositions would be, or why such a restriction would be relevant to decision theory.

A more promising route, perhaps, is to let go of the requirement that every pair of prospects has a sum or difference. We may weaken Addition Invariance to say only that *if* C can be added to both A and B, then A is at least as good as B if and only if the sum of A and C is at least as good as the sum of B and C. The cost of this weakening is that we may no longer derive the important principles of decision theory, such as the Sure-Thing principle or Relative Expectation Theory, that were used to motivate Addition Invariance in Section 3.3. There may still be large swathes of prospects where sums and differences do exist, where we will be able to derive restricted versions of Sure-Thing and the other principles. This is important work for the weakened Addition Invariance, but it seems likely that moving to the more realistic state-free setting makes the principle less desirable.

3.4.2 Addition of Outcomes

Outcomes have so far been identified with real-number utilities. The notions of statewise addition and subtraction of prospects presupposes that we may add or subtract outcomes as real numbers. But an outcome is not a number, it is a possible way for the world to be; a possible history of the universe that is maximally specific regarding how good things turn out. It is unclear whether it makes any sense to add or subtract such things. This raises the question of whether the notion of addition in Addition Invariance is in good standing, independently of complaints about statewise operations mentioned in the previous subsection.

Standardly, numerical utilities are introduced by definition. Two outcomes are arbitrarily selected as the zero point and unit of utility, and other levels of utility are defined in terms of these reference points. For example, an outcome is said to have utility 2 if fifty-fifty between that outcome and the zero point is equally good as the unit outcome. Since the zero point and unit are arbitrary, a notion of addition between outcomes defined in terms of utility is also only defined relative to a choice of zero point and unit. It is therefore desirable that any choice of zero point and unit should work equally well from the point of view of Addition Invariance, or else the principle will only be true on a special and nonstandard precisification of the definition of utility, if at all. Fortunately, this is the case: suppose we shift and rescale the utility scale by the affine transformation $ax + b$, with $a > 0$. On this transformation, what we would mean by "x" is $ax + b$, what we would mean by "$-$" is the function that maps utility $ax + b$ and utility $ay + b$ to utility $a(x - y) + b$, and so on. Therefore, what we would mean by the sentence stating Addition Invariance is the principle

Addition Invariance$_{ax+b}$ $X \succsim Y$ if and only if $a(X - Y) + \mathbf{b} \succsim \mathbf{b}$.

Addition Invariance$_{ax+b}$ is indeed equivalent to Addition Invariance (for the same reasons as in Section 3.3.3), so the principle is insensitive to how the zero point and unit of utility are chosen.

On the assumptions required to assign utilities to outcomes, Addition Invariance is well-defined. However, one might still complain about these assumptions, for example by positing outcomes that are infinitely good or bad or infinitesimally good or bad. For example, an Expected Total Utilitarian would say that, relative to a zero point of nobody existing and a unit of one happy person existing, n happy people existing has utility n. It would be very natural for them to suppose that infinitely many happy people (and nothing else) would be better than any finite number, in which case we would have an outcome without a (real number) utility. If there are outcomes without utilities, we are once again in no position to help ourselves to the notion of adding outcomes together. Addition Invariance may be interpreted as applying only in the special case where all outcomes involved have utilities, but this modification is both *ad hoc* while also neutering the strength promised by Addition Invariance. We would no longer get the full Sure-Thing principle as a consequence, only a surrogate that works in the aforementioned special case.

In settings where real number utilities fail, there may be alternative definitions that give a broader interpretation of the sum of outcomes. For example, it has been suggested that infinitesimal or infinite outcomes might be assigned hyperreal utilities. If so, then Addition Invariance might be strengthened by interpreting '+' as hyperreal addition in a setting where prospects are modelled as functions from states to hyperreal numbers. There may however be different sensible ways of assigning hyperreal numbers to outcomes for which Addition Invariance, so interpreted, would make inequivalent claims. There is no longer any guarantee (without further information about the new definition of utility) that it picks out a unique assignment of hyperreals to outcomes modulo an affine transformation, so we may not reason as before to establish the well-definedness of Addition Invariance.

A more contentious but still potentially valuable way of cashing out the addition in addition invariance would appeal not to a notion of utility achieved via the representation of simple prospects by their utility, but via the representation of the intrinsic ethical features of outcomes. Consider a Parfitian utilitarianism, according to which the thing that matters is the amount of happiness and sadness in the world, where these are considered as scalar fields on $\mathbf{R}^4$ (any σ-finite measure space will do in place of $\mathbf{R}^4$). A spacetime point, we may imagine, carries with it an intensity of happiness or sadness represented as a real number, and the amount of happiness in a region of finite volume is given by integrating the intensity of happiness or sadness in that region (positive numbers indicating a preponderance of happiness and negative numbers sadness). Outcomes, on this view, may be identified as measurable functions from the set of spacetime to real numbers. We will assume outcomes to be locally integrable, which means that a finite area has a finite total amount of happiness or sadness within it.

A natural set of axioms on this view would take the base theory minus Minimal Expected Utility Theory (which presumes the notion of utility), and which adds a few utilitarian-friendly principles:

Local Finite Possibility Every locally integrable real function on spacetime is an outcome.

Total Utilitarianism If f and g are integrable outcomes, then $\int f \geq \int g$ if and only if $f \succsim g$.

Minimal Expected Totalism For $x \geq 1$, a gamble with a $1/x$ chance of x happiness in the unit cube and 0 everywhere else, and a $1 - 1/x$ chance of 0 everywhere, is equally good as a sure 1 unit of happiness in the unit cube.

Spacetime Point Pareto If f dominates g pointwise, then $f \succsim g$.

(Since the outcomes have such a complex structure, it may also be desirable to drop Simple Prospect Totality for the reasons presented in Amanda Askell's dissertation (2018), in which case Total Utilitarianism on its own gives enough comparability for the theory to get off the ground.)

In this case, the notion of addition of prospects can be made sense of as pointwise addition of functions, in which case we can once again consider the effect of a version of Addition Invariance:

Utilitarian Addition Invariance $X \succsim Y$ if and only if $X - Y \succsim \mathbf{0}$, where now subtraction is statewise pointwise subtraction, and $\mathbf{0}$ is the function that maps every state to the function that maps every point to 0.

This is a setting where outcomes are *not* representable by real-valued utilities, because outcomes with an infinite volume of happiness and no sadness are infinitely better than those with a merely finite volume of it. Nevertheless, their addition is perfectly sensible, and Utilitarian Addition Invariance is a natural principle to consider that also yields important consequences in much the same manner as Addition Invariance did in the simpler finite-utility setting. We also have:

Theorem 19. *The theory consisting of Statewise Dominance, Preordering, Totality, Local Finite Possibility, Total Utilitarianism, Minimal Expected Totalism, Spacetime Point Pareto, Totality, and Utilitarian Addition Invariance is consistent.*

Proof. Let $\langle S_i \rangle_{i=1}^{\infty}$ be an expanding system of spheres whose union covers $\mathbf{R}^4$, and let the outcomes be locally integrable functions from $\mathbf{R}^4$ to $\mathbf{R}$. For a given prospect X, let X^i be the real-valued random variable that takes the value x in any state where X yields an outcome whose integral on S_i is x.

Let $\sqsupseteq_{\mathbf{R}}$ be an ordering of real-valued random variables that satisfies Statewise Dominance, Totality, Expected Utility Theory, and Addition Invariance (see Section 3.6). And let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbf{N}$.

Where $\boldsymbol{\varepsilon}$ is an outcome with ε units of happiness in S_1 and 0 units elsewhere, define

$$X \sqsupseteq Y := \forall \varepsilon > 0. \left\{ i \in \mathbf{N} : X^i + \boldsymbol{\varepsilon} \sqsupseteq_{\mathbf{R}} Y^i \right\} \in \mathcal{U}.$$

This preordering of prospects satisfies the relevant principles by the arguments in Chapter 2, Appendix 2.6. □

The theorem shows that even in a setting without a numerical representation of utility, principles like Addition Invariance can make sense.

3.5 Stochastic Equivalence

The principle of *Stochastic Equivalence* says that choiceworthiness of a prospect supervenes on the probability distribution over utilities that taking that prospect gives rise to.

STOCHASTIC EQUIVALENCE If X and Y have the same distribution, then $X \sim Y$.

Theorem 20 (SSK)**.** *Stochastic Equivalence is inconsistent with Addition Invariance*

The proof of SSK's theorem is highly instructive, so shall be included here.

Proof. Suppose that a nickel is to be (fairly, independently) tossed until it lands heads for the first time. Then a dime will be tossed. P, P_{H}, and P_{T} are prospects defined as follows:

P You receive utility 2^n, where n is the number of tosses of the nickel it takes for it to land heads for the first time (the bold letter indicates how the dime lands).

$\mathrm{H}\mathbf{H}$	$\mathrm{H}\mathbf{T}$	$\mathrm{TH}\mathbf{H}$	$\mathrm{TH}\mathbf{T}$	...	...	$\mathrm{T}^{n-1}\mathrm{H}\mathbf{H}$	$\mathrm{T}^{n-1}\mathrm{H}\mathbf{T}$	...	...
2	2	4	4	...	...	2^n	2^n	...	...

P_{H} If the dime lands tails, then you get utility 2. If it lands heads, you receive utility 2^{n+1} where n is the number of tosses it takes for the nickel to land heads for the first time.

$\mathrm{H}\mathbf{H}$	$\mathrm{H}\mathbf{T}$	$\mathrm{TH}\mathbf{H}$	$\mathrm{TH}\mathbf{T}$	...	...	$\mathrm{T}^{n-1}\mathrm{H}\mathbf{H}$	$\mathrm{T}^{n-1}\mathrm{H}\mathbf{T}$	...	...
4	2	8	2	...	...	2^{n+1}	2	...	...

P_{T} If the dime lands heads, then you get utility 2. If it lands tails, you receive utility 2^{n+1} where n is the number of tosses it takes for the nickel to land heads for the first time.

$\mathrm{H}\mathbf{H}$	$\mathrm{H}\mathbf{T}$	$\mathrm{TH}\mathbf{H}$	$\mathrm{TH}\mathbf{T}$	...	...	$\mathrm{T}^{n-1}\mathrm{H}\mathbf{H}$	$\mathrm{T}^{n-1}\mathrm{H}\mathbf{T}$	...	...
2	4	2	8	...	...	2	2^{n+1}	...	...

All three prospects are stochastically equivalent. So if Stochastic Equivalence holds, Addition Invariance will say $P - P_{\mathrm{H}}$ and $P_{\mathrm{T}} - P$ are both equally good as $\mathbf{0}$, so are equally good as each other by transitivity of $\sim$. Therefore $2P - P_{\mathrm{H}} - P_{\mathrm{T}}$ is equally good as $\mathbf{0}$. But $2P - P_{\mathrm{H}} - P_{\mathrm{T}}$ is the constantly -2 prospect, which is worse than $\mathbf{0}$ by Dominance. □

Notice that in this inconsistency SSK's principle of Coherent Indifference, which says only that $X - Y \sim \mathbf{0}$ if $X \sim Y$ (but not necessarily vice-versa), is sufficient in place of Addition Invariance. The only other nonlogical principles required are the transitivity of $\sim$, Statewise Dominance (or SSK's slightly weaker principle of Coherent Strict Preference), and Stochastic Equivalence. Since transitivity and Statewise Dominance seem impossible to give up, either Addition Invariance or Stochastic Equivalence must go.

3.5.1 Weakening Stochastic Equivalence and Addition Invariance

The inconsistency with Stochastic Equivalence is a serious problem. Before we move to litigating the choice between the two principles, it is worth assessing some weakenings of the two principles that are compatible.

Meacham (2019) has suggested a way of weakening either principle that renders it compatible with the other, at the cost of requiring widespread incomparability between prospects for the other principle to then hold. Let us say $X \approx Y$ if X is neither better nor worse than Y: if $X \nprec Y$ and $Y \nprec X$. Being neither better nor worse is different from being equally good if and only if there can be incomparability among prospects. The two weakenings Meacham suggests are:

STOCHASTIC $\approx$-EQUIVALENCE If X and Y have the same distribution, then $X \approx Y$.

ADDITION $\nprec$-INVARIANCE $X \nprec Y$ if and only if $X - Y \nprec 0$

Theorem 21. *Addition Invariance is consistent with Stochastic $\approx$-Equivalence, and Stochastic Equivalence is consistent with Addition $\nprec$-Invariance (consistency remains even given Relative Expectation Theory).*

Proof. Section 3.6 sketches the proof of the first part. For the second part, note that the model with $\sqsupseteq$ in Chapter 2satisfies Addition $\nprec$-Invariance. □

Meacham prefers to reject Addition Invariance to retain Stochastic Equivalence (a "Moorean fact", he calls it), but accepts Addition $\nprec$-Invariance.

How satisfying these weakenings are depends on how plausible it is for there to be incomparability among prospects, on how much work these principles can do without their stronger counterparts, and on whether they can be plausibly motivated without appealing to arguments that would also justify their strengthening to $\sim$ and $\succsim$ respectively. It should be noted that weakening Addition Invariance in this way will still yield Simple Relative Expectation Theory, since we have assumed there is no incomparability among simple prospects. However the Sure-Thing principle will no longer be derivable without additional assumptions.

Another strategy is to restrict either principle to simple prospects or to prospects which have well-defined expected utilities, where they will no longer be applicable to the prospects P, P_{H}, and P_{T} used to generate the contradiction:

SIMPLE STOCHASTIC EQUIVALENCE If X and Y have the same distribution and are simple, then $X \sim Y$.

SIMPLE ADDITION INVARIANCE If $X - Y$ is simple, then $X \succsim Y$ if and only if $X - Y \succsim \mathbf{0}$.

RELATIVE EXPECTATION THEORY If $X - Y$ has an expected utility, then $X \succsim Y$ if and only if $X - Y \succsim \mathbf{0}$.

These weakenings are sufficiently drastic that it is unlikely that there is any point in taking these as primitive axioms, because given either Addition Invariance or Stochastic Equivalence, and mild assumptions, the "simple" restriction of the other principle follows.

Theorem 22. *Simple Stochastic Equivalence follows from the Sure-Thing principle (and thus from Addition Invariance).*

Theorem 23. *Simple Addition Invariance follows from the Sure-Thing principle, Shift Invariance, and Stochastic Equivalence. Relative Expectation Theory follows from these assumptions plus L^1 Continuity.*

Proof. Chapter 4, Theorem 31 and Corollary 32. □

3.5.2 Problems of Cluelessness

The application of decision theory is difficult. In general, we do not know with perfect precision either the probabilities of the various outcomes our action might yield, or their utilities. This is the problem of *Cluelessness*, named by Lenman (2000) who emphasised it as a problem for consequentialism. Principles like Addition Invariance and Stochastic Equivalence may have practical application, in that they might be appealed to in order to gain knowledge of the best course of action. This section examines the usefulness of the two principles in this practical sense with the aim of deciding between them on this basis.

A similar idea is frequently found in the justification of other decision-theoretic principles in the literature. One of Savage's (1954: §5.5) justifications for his axioms is that they permit the seamless transition between the more realistic "grand-world" and more tractable "small-world" descriptions of decision problems.

The application of Stochastic Equivalence is that it allows us to ignore the payoff structure of prospects, and focus only on the properties of their probability distributions over utility. The cost of accepting Addition Invariance, then, is that it brings the payoff structure of prospects back into the picture. However, Addition Invariance is consistent with Expected Utility Theory, so we might suppose that the proponent of Addition Invariance may ignore everything but the probability when comparing prospects with finite expected utilities. Any differential application of Stochastic Equivalence must therefore derive from pairs of prospects X and Y such that $X - Y$ does not have an expected utility.

It is natural, then, to try to find applications for Stochastic Equivalence by gambles of infinite expected utility based on independent random processes, such as two St Petersburg prospects based on different sequences of coin tosses. Stochastic Equivalence will say that two such prospects are equally good, whereas Addition Invariance will stay silent. However, it is important to notice that mere *approximate* stochastic equivalence counts for nothing when applying Stochastic Equivalence. Knowing only that the two sequences of coin tosses are approximately fair and independent is not sufficient for concluding with the equal value of the two prospects, because, supposing each toss in each sequence to be independent and equally probable to yield heads, any nonzero improvement on the probability of heads will make for an infinite improvement in expectation. Therefore, any uncertainty about whether one of the sequences is the slightest bit more biased than the other reintroduces uncertainty about which prospect is better. It is more contentious to assume that we very often find ourselves in a position to know that two random processes are *exactly* stochastically equivalent.

Whereas it is contentious to think we often know that two random processes are exactly stochastically equivalent, it is downright implausible to think we are often in a position where we can say much at all about the subtraction of two prospects. Suppose that you have two options: a St Petersburg prospect based on fair and independent coin tosses, and that same prospect except with the first toss replaced with the toss of a spare coin, which would also be fair and probabilistically independent. Suppose that in fact you take the original prospect, so the spare coin is never tossed. Stochastic Equivalence says that the two prospects are equally good; the proponent of Addition Invariance must also make a judgement about how taking the alternative prospect would have counterfactually affected the coin tosses. One hypothesis is that, had you taken the second option, the coins after the first would have landed the same way. In that case, there are two natural hypotheses about how the spare coin would have landed:

(i) The spare coin would definitely have landed the same way as the original coin actually does (perhaps because it would have been tossed by the same person, who would have tossed it in the same way). In that case, the subtraction of the first and second prospects is identically 0, so they are equally good.

(ii) How the spare coin would have landed is probabilistically independent of how the original coin actually does (perhaps because your choosing to switch coins would have had an effect on the toss, or perhaps because the toss is truly indeterministic). In that case, the subtraction of the first and second prospects is not identically 0 and lacks an expected utility.

The decision between (i) and (ii) is rather obscure, both philosophically (Goodsell 2022) and empirically. It can only be made worse when the case is less idealized. This obscurity is present even supposing the tosses to be exactly fair and independent.

So it seems going for Addition Invariance instead of Stochastic Equivalence makes the problem of cluelessness significantly worse.

Addition Invariance helps in cases where the subtraction of two prospects $X - Y$ is easier to compare with $\mathbf{0}$ than X and Y were originally. Often, such an application of Addition Invariance might be recoverable even in a theory with Stochastic Equivalence. For example, if $X - Y$ dominates $\mathbf{0}$, or has a higher expected utility than $\mathbf{0}$, then a theory that includes dominance or Relative Expected Utility Theory will entail the comparison $X \succsim Y$. There is a sense in which this is always possible. Say that a theory Γ *logically implies* a comparison between real random variables X and Y if X and Y satisfy '$\succsim$' on any interpretation of '$\succsim$' on which Γ is true. Two theories can logically imply the same comparisons without being logically equivalent, and in particular:

Theorem 24. *The base theory plus Addition Invariance logically implies the same comparisons as the base theory plus Sure-Thing principle and Simple Relative Expectation Theory.*

Proof sketch. There is a minimal ordering $\sqsupseteq$ which satisfies Sure-Thing and Relative Expectation Theory, and $\sqsupseteq$ also satisfies Addition Invariance. Therefore, no comparison implied by Addition Invariance is not also implied by Sure-Thing and Relative Expectation Theory. Conversely, Addition Invariance implies those two principles so no comparison implied by Sure-Thing plus Relative Expectation Theory is not implied by Addition Invariance. □

Theorem 25. *Theorem 24 holds when Addition Invariance is strengthened by the addition of Totality, or when Addition Invariance and the theory with Sure-Thing and Relative Expectation Theory are both strengthened by L^1 Continuity.*

Proof sketch. Section 3.6. □

From the practical perspective, then, the additional strength of Addition Invariance over Sure-Thing and Relative Expectation Theory is not so significant.

On its alternative formulation, Addition Invariance says $X \succsim Y$ if and only if $X + Z \succsim Y + Z$, for any Z. On this formulation we can think of Z as an indeterminate "status quo" and of X and Y as possible changes to the status quo. Addition Invariance then says that we don't need to know what the status quo is before deciding between the two options, because the comparison would be the same no matter the status quo. $X + Z$ is the prospect where in each state you add the utility of X to the utility Z, so on this way of thinking about things we must assume that they payoffs of X and Y are independent of the status quo, in the sense that an outcome of x for X contributes an equal amount of utility no matter what the outcome of Z is. If we think of Z as something beyond the control of the agent, this requires that X and

Y add a good or bad within the control of the agent that is independent of what is beyond the agent's control. An Average Utilitarian will not find this assumption plausibly satisfied in many cases, since according to them whether a local action is good or bad depends on a global feature, namely the average welfare. By contrast, a Total Utilitarian will think that, when X and Y are prospects about what happens within an agent's cone of causal influence, and Z is a prospect about what happens outside of that cone, then the contributions of X or Y and Z can be added together.

Addition Invariance can for a Total Utilitarian be viewed as a strong principle of Separability: how local interventions X and Y compare is invariant under changes in what is going on outside of your cone of causal influence. This principle of separability is *prima facie* desirable and would have important implications for ethics, but is inconsistent with Stochastic Equivalence (given mild assumptions about how big the world can be) for the same reason as SSK's theorem. Addition Invariance therefore gains some desirability for Total Utilitarians. It simplifies the comparison of ethical prospects by allowing us to ignore what is not affected, but complicates it by not allowing us to ignore counterfactual structure in favour of probabilistic structure. Weighing the costs and benefits is difficult, but an important factor in weighing the acceptability of Addition Invariance.

3.5.3 Decisions and News

There are two ways in which an action can be ethically evaluated on the balance of probabilities. It can be evaluated on the basis of its choiceworthiness, a feature that has to do with how likely good would come about if that action were taken. It can also be assessed in the same way as every other happening. How good would the news be that the action was taken? These two modes of assessment are generally thought to come apart in cases where a potential action is correlated with bad news but does not cause it. For example, if you think that being blessed at birth causes you to both live piously and go to heaven, but that living piously is costly and does not have any causal effect on going to heaven, then living piously is a poor decision such that it would be good news for you to take that option, since it would indicate that you are destined for heaven. Call these two evaluative notions *decision-value* and *news-value*.[6]

The principle of Addition Invariance trades on the distinction between decision-value and news-value. From the news-value perspective, a prospect is identified with the proposition that you take the prospect, and is evaluated as any other proposition. The states in Savage-style causal decision theory are possibilities which are modally

[6] *Evidential* decision theorists, most famously Jeffrey (1965), identify the two. *Causal* decision theorists, including Gibbard and Harper (1978), Lewis (1981), and (1999), distinguish them.

independent of the available options, so when we think of "options" not as actions but as propositions that one might learn, any notion of state would have to be modally independent of anything one might learn. The news-value approach to decision theory does not seem to require the existence of any such things as states, as indicated by classical theories of news-value due to Jeffrey (1965) and Bolker (1967). We may suppose that all the only uncertainty in the world comes from some coin tosses, so that one knows how good things would be conditional on how the tosses turn out, but not how they will turn out. In this setting, there seems a perfectly good notion of news-value given by computing the expected utility conditional on any proposition about the coin-tosses, but no notion of state by which two propositions about coin-tosses can be compared statewise. In addition, when we reformulate Addition Invariance in the state-free way as in Section 3.4.1, statewise operations are replaced with counterfactual properties of propositions. But not only do counterfactuals have little to do with the news-theoretic evaluation of propositions, the impossibility result in that section suggests that there could not be any reasonable conditional that gives rise to a notion of addition or subtraction between propositions suitable to news-value.

It is widely accepted that decision-value and news-value are distinct, so it is legitimate for the proponent of Addition Invariance to rely on that distinction. However, it is generally thought that the two notions bear many structural similarities to each other. For example, Savage's (1954) theory of decision-value and Jeffrey's (1965) and Ethan Bolker's (1967) theories of news-value give rise to distinct interpretations of the structural axioms of von Neumann and Morgenstern, and thereby to expected utility representations of both decision-value and news-value. It is reasonable to suppose that broad structural principles should be mirrored between both notions. To accept Addition Invariance for decision-value but no analogue for news-value is to posit an important structural feature one has that the other lacks. Moreover, since Addition Invariance has no obvious analogue for news-value, a Stochastic Equivalence for news-value seems to stand unopposed. And to accept Addition Invariance for decision-value and Stochastic Equivalence for news-value would be to posit an especially surprising clash in the structure of the two notions.

Principles that link decision-value and news-value further problematize Addition Invariance. The cases where decision-value and news-value are widely accepted to come apart are cases where an action's being performed is correlated with an outcome that performing that action would not likely result in. Therefore, one would expect decision-value and news-value comparisons to coincide in cases where there are no such spurious correlations.

For a measurable set of real numbers X, let p_X be the proposition that the actual world will have a utility in X. Say that an action is *unspurious* if the probability of decision-theoretic counterfactuals with your doing A as the antecedent aligns with probability conditional on your doing A:

$$\mathbf{P}(A \mathrel{\Box\!\!\rightarrow} p_X) = \mathbf{P}(p_X|A), \tag{3.1}$$

for every measurable set X of real numbers.

Bracketing off the standardly recognized cases where decision-value and news-value come apart suggests that they do not come apart for unspurious prospects:

Unspurious Evidential Decision Theory If A and B are unspurious actions, then A is at least as good as B in decision-value if and only if A is at least as good as B in news-value.

Unspurious actions generally arise when the prospects can be thought of as bets on a random process, such that your bets are probabilistically independent of that process, as well as certainly counterfactually independent of it.[7] In particular, bets on an infinite sequence of coin tosses are typically thought to be unspurious, in which case the prospects P, P_{H}, and P_{T}, are unspurious. Thus, if we accept Addition Invariance for decision-value and Unspurious Evidential Decision Theory, we must also deny Stochastic Equivalence for news-value. There must be, that is, propositions that are not equally good news, despite every measurable set of outcomes being equally likely conditional on either. To emphasise the counterintuitiveness of this result, we can treat the agent in the decision problem as themselves governed by a random process. The utility of the true possible world can be thought of as a random variable, which we hope to be higher rather than lower (i.e., we hope to be in a better situation rather than a worse one). Suppose P and P_{H} are not equally good in news-value (otherwise substitute P_{T} for one of them), and suppose that the agent is equally likely to choose P as she is to choose P_{H}, with there being no correlation between her choice and the coins. Clearly, no matter what she chooses there are equal probabilities of any set of outcomes, so her choice is probabilistically independent of the utility random variable. But despite being probabilistically independent of how good things will turn out, the news would *not* be equally good no matter what we learn about her choice.

This sort of situation is hard to stomach. It is compounded by the fact that, in the news-value case, there seems to be no strong principle like Addition Invariance that can take the place of Stochastic Equivalence. It seems then that the best option, given Addition Invariance, would be to reject Unspurious Evidential Decision Theory. Again, this seems to be a serious cost.

[7] Stating this claim precisely and proving it requires some setup that will be omitted here.

3.6 Metatheory

This section establishes the consistency of Addition Invariance with Totality, as well as certain other interesting metatheoretical results such as Theorem 25.

The space of prospects, in the present setting, is the vector space of real random variables $\mathcal{V}$. We shall consider models where logical and mathematical constants are given their intended interpretation, but where '$\succsim$' may be interpreted deviantly. These correspond to linear preorderings of the vector space of real random variables ('linear' here always means compatible with vector operations, rather than total).

Theorem 26. *The base theory and Addition Invariance are true on all and only interpretations of '$\succsim$' that are linear preorderings of $\mathcal{V}$ that agree with the dominance and simple expected utility preorderings.*

Theorem 27. *A linear preordering of $\mathcal{V}$ satisfies L^1 Continuity if and only if it extends the expected utility preordering.*

An *extension* of a preordering is a preordering that includes at least as many comparisons but which does not collapse any strict comparisons: $\sqsupseteq'$ extends $\sqsupseteq$ iff $\sqsupseteq \subseteq \sqsupseteq'$ and $\sqsupset \subseteq \sqsupset'$. Since linearity is the only constraint required of the extension, besides that it be an extension of the smaller preorderings, the following lemma will be helpful in the metatheory of Addition Invariance:

Lemma 28. *Let $\sqsupseteq$ be any linear preordering. If $x \not\sqsupseteq y$ and $y \not\sqsupseteq x$, then there is a linear extension of $\sqsupseteq$, $\sqsupseteq'$, with $x \sqsupseteq' y$, and there is a linear extension $\sqsupseteq''$ with $y \sqsupseteq'' x$.*

Totality requires there to be a total linear extension of the dominance and simple expected utility preorderings (delete 'simple' when assuming L^1 Continuity). By the lemma, the existence of total linear extensions is guaranteed by Zorn's Lemma.

These theorems together imply that the base theory is consistent with Addition Invariance with or without Totality. However, they also immediately give rise to Theorem 25, since they show that any comparison consistent with the minimal constraints on a preordering of $\mathcal{V}$ can be consistently included in a total extension of the minimal preordering. It follows also that the counterexample to Stochastic Equivalence due to SSK's Theorem (Theorem 20) is nonconstructive: it is consistent to add any of $P \sim P_{\mathrm{H}}$, $P \sim P_{\mathrm{T}}$, or $P_{\mathrm{T}} \sim P_{\mathrm{H}}$, but not any two of these comparisons. More generally, any particular comparison of two prospects given by Stochastic Equivalence is consistent with Addition Invariance, since the dominance and expected utility preorderings never contradict Stochastic Equivalence. This yields the first part of Theorem 21.

References

Askell, Amanda (2018). "Pareto Principles in Infinite Ethics". PhD thesis. New York University.

Bartha, Paul (2016). "Making Do Without Expectations". In: *Mind* 125.499, pp. 799–827.

Bolker, Ethan D. (1967). "A Simultaneous Axiomatization of Utility and Subjective Probability". In: *Philosophy of Science* 34.4, pp. 333–340.

Colyvan, Mark (2008). "Relative Expectation Theory". In: *Journal of Philosophy* 105.1, pp. 37–44.

Colyvan, Mark and Alan Hájek (2016). "Making Ado Without Expectations". In: *Mind* 125.499, pp. 829–857.

Easwaran, Kenny (2008). "Strong and Weak Expectations". In: *Mind* 117.467, pp. 633–641.

— (2014b). "Principal Values and Weak Expectations". In: *Mind* 123.490, pp. 517–531.

Gibbard, Allan and William L. Harper (1978). "Counterfactuals and Two Kinds of Expected Utility". In: *Foundations and Applications of Decision Theory*. Ed. by A. Hooker, J. J. Leach, and E. F. McClennen. D. Reidel, pp. 125–162.

Goodsell, Zachary (2022). "Tossing Morgenbesser's Coin". In: *Analysis* 82.2, pp. 214–221.

Gwiazda, Jeremy (2014). "Orderly Expectations". In: *Mind* 123.490, pp. 503–516.

Hájek, Alan (2014). "Unexpected Expectations". In: *Mind* 123.490, pp. 533–567.

Hájek, Alan and Michael Smithson (2012). "Rationality and Indeterminate Probabilities". In: *Synthese* 187.1, pp. 33–48.

Jeffrey, Richard (1965). *The Logic of Decision*. University of Chicago Press.

Joyce, James M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Lauwers, Luc (2016). "Why Decision Theory Remains Constructively Incomplete". In: *Mind* 125.500, pp. 1033–1043.

Lenman, James (2000). "Consequentialism and Cluelessness". In: *Philosophy and Public Affairs* 29.4, pp. 342–370.

Lewis, David (1981). "Causal Decision Theory". In: *Australasian Journal of Philosophy* 59.1, pp. 5–30.

Meacham, Christopher J. G. (2019). "Difference Minimizing Theory". In: *Ergo: An Open Access Journal of Philosophy* 6.

Savage, Leonard (1954). *The Foundations of Statistics*. Wiley Publications in Statistics.

Seidenfeld, Teddy, Mark Schervish, and Joseph Kadane (2009). "Preference for Equivalent Random Variables: A Price for Unbounded Utilities". In: *Journal of Mathematical Economics* 45, pp. 329–340.

Smith, Nicholas J. J. (2014). "Is Evaluative Compositionality a Requirement of Rationality?" In: *Mind* 123.490, pp. 457–502.

Chapter 4

Symmetries of Value

4.1 The Affine Symmetry Principles

This paper investigates two natural principles of decision theory, together called the *affine symmetry principles*, which posit how certain symmetries in the value of the outcomes of risky prospects induce corresponding symmetries in the comparison of the prospects themselves. The principles, despite initially seeming obvious, are shown to have striking consequences for decision theory in cases where the standard expected utility of a prospect is undefined.

The possible outcomes of prospects have an interval scale structure, corresponding to how much a chance of them contributes to the overall value of a prospect. For example, we say that outcome o_2 is halfway between o_1 and o_3 if getting o_2 for sure is equally good as a a fifty-fifty gamble between o_1 and o_3, whereas o_2 is a third of the way from o_1 to o_3 if getting o_2 for sure is equally good as a $2/3$ chance of o_1 and a $1/3$ chance of o_3. This structure can be graphically represented by placing the outcomes on a line as in Figure 4.1, so that the ordering of the outcomes is represented by the left-to-right ordering of points on the line, and the ratios of differences between outcomes are represented by the ratios of distances between their corresponding points.

Figure 4.1: Interval scale structure of outcomes represented by placing outcomes called "Death", "Coal", "Cake", and "Heaven" on the line. The diagram represents Heaven as the best, Death as the worst, certainty of Coal as being equally good as 3/4 chance of Death and a 1/4 chance of Heaven, certainty of Cake as being equally good as 2/3 chance of Coal and 1/3 chance of Heaven, and so on.

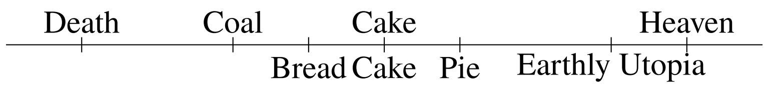


Figure 4.2: In this figure with three additional outcomes ("Bread", "Pie", "Earthly Utopia") are identified on the line, so that the outcomes with labels above the line exhibit the same interval scale structure as the outcomes with labels below the line. Both above and below the line, the second worst outcome is a quarter of the way from the worst to the best, the third worst is half of the way from the worst to the best, and so on.

Typical approaches to decision theory—including the widely accepted Expected Utility Theory—constrain the overall ranking of prospects in terms of the interval scale structure of outcomes.[1] For any such approach, the comparison of two prospects depends only on the interval scale structure exhibited by the outcomes those prospects might yield. If Figure 4.2 is accurate about the interval scale structure of prospects, this means that any verdict these approaches yield about gambles between the outcomes Death, Coal, Cake, and Heaven would be the same if Death were replaced with Bread, Coal with Cake, Cake with Pie, and Heaven with Earthly Utopia.

A very natural idea, which is the topic of this paper, is that the comparison between any two prospects does not depend on any feature of outcomes besides their interval scale structure. This is the principle of *Positive Affine Invariance*. The other idea investigated in this paper is closely related. The principle of *Negative Affine Anti-Invariance* says that if the ordering of prospects by *better than* and the ordering of prospects by *worse than* are determined by the interval scale structure of the outcomes in the same way, which is to say that reversing the interval scale structure while keeping all ratios of differences the same should flip the comparison of any two prospects. Together these are called the *affine symmetry principles*.

To state the affine symmetry principles more carefully, let an *affine permutation*

[1]Isaac Levi (1990) argued that rational preference over outcomes does not in general exhibit this interval scale structure, and instead exhibits structure corresponding to a set of "permissible" interval scale structures that intuitively correspond to different ways of evaluating prospects which the agent has not ruled out as incorrect (permissible ways of evaluating are indexed by utility functions; see page 80 of Levi 1990). Nevertheless, Levi's theory of preference presupposes the evaluation of prospects with respect to a particular permissible way of evaluating, for which there is always a unique interval scale structure of outcomes as outlined above. This paper is intended to complement Levi's approach by further examining the evaluation of prospects with respect to any particular permissible interval scale structure, especially in cases where the standard approach of comparing by expected utility does not yield an answer. "Better than", "equally good as" and so on as they occur here can be interpreted as relative to an arbitrary permissible evaluation of prospects, which is assumed to be held fixed throughout.

Table 4.1: The alternating St Petersburg prospect

$1/2$	$1/4$	...	$1/2^n$	...
-2	4	...	$(-2)^n$	...

of outcomes be a permuation of outcomes π that preserves the ratios of differences between outcomes: so that $\pi(o_2)$ is the same proportion of the way from $\pi(o_1)$ to $\pi(o_3)$ as o_2 is from o_1 to o_3. An affine permutation is said to be *positive* if it preserves the ordering of outcomes: $\pi(o_1)$ is at least as good as $\pi(o_2)$ just in case o_1 is at least as good as o_2; otherwise it *reverses* the ordering of outcomes, in which case the affine permutation is said to be *negative.*

For an arbitrary prospect μ let $[\pi]\mu$ be the prospect that yields $\pi(o)$ whenever μ would yield o. The affine symmetry principles may now be stated as follows:

Positive Affine Invariance For any positive affine permutation of outcomes π, and for any prospects μ and ν, $\mu \succsim \nu$ if and only if $[\pi]\mu \succsim [\pi]\nu$.

Negative Affine Invariance For any negative affine permutation of outcomes π, and for any prospects μ and ν, $\mu \succsim \nu$ if and only if $[\pi]\mu \precsim [\pi]\nu$.

Which is to say that the comparison of prospects μ and ν is insensitive to anything besides the interval scales structure of the outcomes they might yield.

The standard view in decision theory, Expected Utility Theory, includes theorems very close to the affine symmetry principles. A *utility function* is a function from outcomes to real numbers U that correctly represents their interval scale structure, so that o_1 is at least as good as o_2 if and only if $U(o_1) \geq U(o_2)$, as well as so that o_2 is a proportion of x of the way from o_1 to o_3 if and only if

$$U(o_2) = (1 - x)U(o_1) + xU(o_3).$$

Expected Utility Theory says that, for any utility function U, a prospect at least as great expected utility is at least as good. For any utility function U, an affine permutation of outcomes performs some affine transformation of real numbers $ax+b$ on the utility of those outcomes, so the ordering of prospects by their expected utility is invariant under positive affine permutations of outcomes, and is reversed by negative affine permutations of outcomes.

However, Expected Utility Theory does not quite establish the affine symmetry principles because not every prospect has an expected utility. A good example, and one we will return to in Section 4.5, is the *alternating St Petersburg prospect*, which for some utility function yields an outcome with utility $(-2)^n$ with probability $1/2^n$, for every positive integer n (see Table 4.1). The expected utility of this prospect would be given by the sum

$$-2 \times \frac{1}{2} + 4 \times \frac{1}{4} - 8 \times \frac{1}{8} + \cdots = -1 + 1 - 1 + \ldots,$$

which alternates between -1 and 0 and never converges. The affine symmetry principles apply even to prospects that lack expected utilities; saying, for example, that replacing all the outcomes in alternating St Petersburg with outcomes that are twice as far from the outcome called "Cake" should preserve which is better out of that prospect and certainty of Cake.

The affine symmetry principles are the topic of this paper. We begin in Section 4.2 by contrasting them with closely related but importantly different assumptions in decision theory. Section 4.3 lays out an axiomatic framework for decision theory that forms the backdrop within which the affine symmetry principles will be studied, following the system DTU of Chapter 2. Section 4.5 investigates the consequences of the affine symmetry principles in this setting, including both general structural consequences, such as the vindication of Mark Colyvan's (2008) Relative Expectation Theory, as well as consequences about the comparisons of particular prospects. Most strikingly, they imply that the alternating St Petersburg prospect is exactly equally good as a sure utility $-1/2$, and that the "Pasadena" prospect of Nover and Hájek 2004 is equally good as utility $\ln 2$, among other comparisons. Model theory for DTU plus the affine symmetry principles is developed in Section 4.6, establishing among other things the consistency of the affine symmetry principles. The affine symmetry principles, then, form a well-motivated, powerful, and consistent way of extending decision theory in cases where expected utility is undefined.

We will assume throughout that outcomes are representable by real number utilities. This assumption rules out infinite ratios of differences between outcomes. There is no philosophical reason behind such a restriction, but it simplifies the mathematics significantly, so it is worthwhile to restrict our attention here to some domain of outcomes for which all ratios of differences are finite, so that the real number utility representation is possible.

4.2 Distinguishing the Affine Symmetry Principles

Lore in decision theory has it that utility is only 'defined' (or 'unique') up to a positive affine transformation. What this means is that for any utility function U, any positive affine transformation of U (i.e., $aU + b$ for a positive real number a and real number b) represents the interval scale structure of outcomes equally well. This being the case, it can be hard to see why the affine symmetries, or at least Positive Affine Invariance, could add anything to standard decision theory. But there is

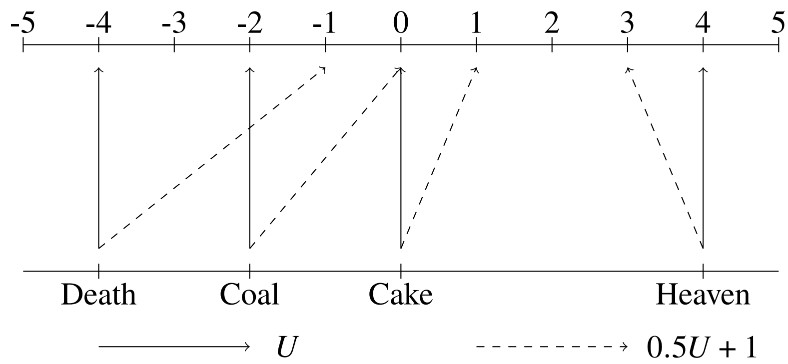


Figure 4.3: U and $0.5U + 1$ are utility functions, so are equally good representations of the interval scale structure of outcomes.

an important difference: the lore is a well-established theorem of decision theory, whereas the affine symmetry principles are substantial additional axioms. Being clear on the difference is required for understanding the affine symmetry principles.

The difference between the lore and the affine symmetry principles can be illustrated by recognizing the difference between an affine transformation of a utility function and an affine permutation of outcomes. A utility function is a function from outcomes to numbers that represent those outcomes. An affine transformation of the utility function is another function from outcomes to numbers, so that different numbers are chosen to represent the same outcomes—see Figure 4.3. We may think of the difference as a change in units, such as representing the freezing point of water by the number 0 (as in degrees Celsius) or the number 32 (as in degrees Fahrenheit). When it is observed that utility is 'defined' up to a positive affine transformation, what is being observed is that standardly, the role of 'the utility function' is to numerically represent the interval scale structure of outcomes, so that any positive affine transformation of the utility function would do an equally good job in that role. By contrast, an affine permutation of outcomes is a function from outcomes to outcomes: outcomes themselves replace outcomes on an affine permutation of outcomes, and numbers need not enter the picture—see Figure 4.4. The theorem is that the interval scale structure of outcomes is faithfully represented by any positive affine transformation of a utility function. The principle of Positive Affine Invariance adds that nothing besides the interval scale structure of outcomes enters into the comparison of prospects.[2]

The distinction is somewhat blurred in the literature, especially in connection with unbounded utility. In the following passage, Easwaran argues from the obser-

[2]This distinction is closely related to the distinction between passive and active transformations in geometry—see page 84 of Struik 1953.

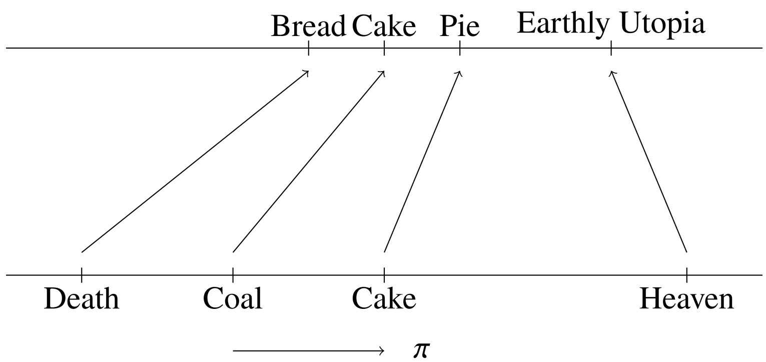


Figure 4.4: π is a positive affine transformation, because it preserves the ordering and ratios of differences between prospects (in relation to Figure 4.3, $\pi = U^{-1} \circ (0.5U + 1)$). Positive Affine Invariance says that substituting Death for Bread, Coal for Cake, etc. in two prospects preserves their comparison.

vation that utility is only defined up to a positive affine transformation to a principle much like Positive Affine Invariance:

> Hájek (2013, pp. 9–10) [Hájek 2014: 541-42] notes that utility is only defined up to a shift and a stretch—there is no well-defined 0 and no well-defined unit in which utility is measured. Thus, it ought to be the case that adding a constant to the utility of every outcome of a gamble affects the overall value of the gamble by adding the same constant, and similarly for multiplying the utilities by a positive constant. (Easwaran 2014b: 523)

The subtle point is that the principle being argued for is one that also refers to the utility function, so it would seem that affine permutations of outcomes never enter the picture. But Easwaran is not arguing for a triviality. What follows 'it ought to be the case that' is not deductively entailed by what precedes 'thus'. Rather, the picture that Easwaran is working with is that a typical decision theory specifies an ordering of (some) prospects by an operation that converts a utility function into a real-valued *value function* on prospects, so that when two prospects are given values in this sense, the one with the greater value is better. Symbolically, an Easwaran-style theory posits a two-place value function $V(\cdot, \cdot)$, which takes a utility function and a prospect and outputs a number. Expected Utility Theory is a theory of this sort, where $V(U, \mu)$ is the expectation of U on μ, or $\mathbf{E}_{\mu}U$, as are the proposals investigated in Easwaran's paper. Easwaran's constraint amounts to forbidding such theories where the ordering of prospects achieved at the end is sensitive to which utility function is plugged in at the beginning, which is to say that for any prospects

μ and ν, any utility function U, and positive real numbers a and real number b

$$V(U,\mu) \geq V(U,\nu) \leftrightarrow V(aU+b,\mu) \geq V(aU+b,\nu).$$ [3]

Easwaran's constraint is motivated by the same idea we have used to motivate the affine symmetry principles: a theory which violated the constraint would be sensitive to features of prospects other than probability and the interval scale structure of their possible outcomes; such a theory would posit an as-yet unrecognized feature of outcomes to be relevant in the evaluation of prospects, and it seems very odd to accept such a theory. But it is not something that is established just by noting that "utility is only defined up to a shift and a stretch."

Adopting Easwaran's constraint does not require accepting Positive Affine Invariance. Rather, the constraint forbids accepting theories which posit specific comparisons that contradict Positive Affine Invariance, such as a theory which gives value $-1/2$ to the alternating St Petersburg prospect but 3 to double the alternating St Petersburg prospect (this would be a counterexample to Positive Affine Invariance because double $-1/2$ is -1, so double the alternating St Petersburg would have to be no better than utility -1). Whereas Positive Affine Invariance says that theories violating Easwaran's constraint are false, Easwaran's constraint only forbids us from accepting them, which could be for any number of reasons (for example, it might be because we can't know any such theory to be true). Nevertheless, Positive Affine Invariance would be a very natural explanation of why adopting Easwaran's constraint is a good idea.

4.3 Axiomatic Decision Theory

The affine symmetry principles will be investigated in the context of the theory DTU of Chapter 2, with three simplifying modifications. The first of these is the identification of prospects with probability distributions over outcomes—in effect identifying the prospect with the probability distribution over outcomes that would arise from taking that prospect. This identification is guaranteed to be harmless by DTU. The second simplifying assumption is the identification of equally good outcomes—allowing us to suppose that there is exactly one outcome of each utility. The final simplification is that a utility function U is taken as primitive, and we assume that simple prospects (prospects which can only yield one of finitely many

[3] Equivalently, given the Independence principle (Section 4.3) and if V extends expected utility:

$$V(aU+b,\mu) = aV(U,\mu) + b$$

outcomes) are ordered by their expected utilities. Again, this simplification is permitted by the axioms of DTU, which imply the existence of such a function.

The resulting simplified theory has two primitives: a measurable space $\mathcal{O}$ of outcomes, and the *at least as good* ordering of probability distributions over $\mathcal{O}$, $\succsim$. We use 'prospect' and 'probability distribution over $\mathcal{O}$' interchangeably, and use greek letters μ, ξ, ν to range over prospects. We also write outcomes as the arguments of $\succsim$, such as in $u \succ o$, to mean $\delta_u \succ \delta_o$, where δ_u and δ_o are the prospects with 100% chance of u and of o respectively.

Total Preordering $\succsim$ is reflexive ($\mu \succsim \mu$), transitive (if $\mu \succsim \xi \succsim \nu$, then $\mu \succsim \nu$), and total ($\mu \succsim \nu$ or $\nu \succsim \mu$).[4]

Independence $\mu \succsim \nu$ if and only if for any prospect ξ and number x between 0 and 1, $x\mu + (1-x)\xi \succsim x\nu + (1-x)\xi$.

Unbounded Utility U is a measurable bijection from outcomes to real numbers.

Simple Expected Utility Theory If μ and ν are simple, then $\mu \succsim \nu$ if and only if $\mathbf{E}_\mu U \succsim \mathbf{E}_\nu U$.

Stochastic Dominance If for every outcome o, the probability on μ of an outcome better than o ($\mu(\{u : u \succ o\})$) is at least the probability on ν of an outcome at least as good as o, then $\mu \succsim \nu$. If there is also some outcome o for which the probability on μ of an outcome better than o is strictly greater than the probability on ν of an better than o, then $\mu \succ \nu$.

To reason with affine permutations of outcomes, we adopt the notation of writing $aU + b$ for the permutation which maps an outcome of utility x to the outcome of utility $ax + b$. More accurately, this permutation would be denoted $U^{-1} \circ (aU + b)$; the benefit of the compact notation is that it allows us to write $[aU + b]$ for the function which modifies a prospect by replacing each outcome by the permuted outcome; which equivalently replaces any chance that prospect has of utility x with the same chance of utility $ax + b$. We may now state the affine symmetry principles as follows:[5]

[4]The totality of $\succsim$ is somewhat controversial, despite being an axiom of both von Neumann and Morgenstern and Savage. It would be rejected by Isaac Levi if "$\succsim$" is understood as preference for some rational agent. However, as mentioned in footnote 1, "$\succsim$" is better understood in the context of Levi's (1990) approach as *being at least as good according to some particular permissible way of evaluating prospects*, which is very plausibly a total preordering.

[5]Work of Seidenfeld, Schervish, and Kadane (2009) suggests a strengthening of the affine symmetry principles that contradicts Stochastic Dominance. Treating prospects as utility-valued random variables on a common state space instead of as probability distributions over outcomes, the princi-

Positive Affine Invariance When $a > 0$, $\mu \succsim \nu$ if and only if $[aU + b]\mu \succsim [aU + b]\nu$.

Negative Affine Anti-Invariance When $a > 0$, $\mu \succsim \nu$ if and only if $[-aU + b]\mu \precsim [-aU + b]\nu$.

In fact, only Negative Affine Anti-Invariance is strictly needed, since it straightforwardly implies Positive Affine Invariance in DTU. However, for the purpose of understanding proofs, it is helpful to replace Negative Affine Anti-Invariance with the weaker Reflection Anti-Invariance:

Reflection Anti-Invariance $\mu \succsim \nu$ if and only if $[-U]\mu \precsim [-U]\nu$.

Positive Affine Invariance is independent of Reflection Anti-Invariance, but together these principles entail and are entailed by Negative Affine-Invariance (Reflection Invariance does not have an obvious interpretation in terms of the interval scale structure of outcomes, but is easy to use in settings with a primitive utility function). DTU + Sym will be DTU plus the affine symmetry principles.

As noted in Chapter 2, DTU does not quite imply Expected Utility Theory, being limited to the weaker *Simple* Expected Utility Theory, which only says anything about simple prospects. A very natural addition to DTU that fills this gap is the principle of L^1 Continuity (Chapter 3), which says that the preordering $\succsim$ is continuous with respect to convergence in expectation:

L^1 Continuity Let μ_i be a sequence of prospects that are the distributions of some outcome-valued random variables X_i on $[0, 1]$ with Lebesgue measure, and suppose X_i converge in expected utility to some X which is distributed according to the prospect μ. Then if $\mu_i \succsim \nu$ for each i, then $\mu \succsim \nu$.

Adding L^1 Continuity to DTU brings out some further interesting consequences of the affine symmetry principles, which will be noted in the following. DTU + Sym + L^1 Continuity will be the result of adding L^1 Continuity to DTU + Sym.

---

ple their work suggests extends Positive Affine Invariance by saying that adding a random variable, instead of just a constant, preserves the ordering:

Random Variable Affine Invariance For any random variables X, Y, Z and real number $a > 0$, $aX + Z \succsim aY + Z$ if and only if $X \succsim Y$.

This principle entails both affine symmetry principles, but Seidenfeld, Schervish, and Kadane show that Random Variable Affine Invariance, and in fact a weaker principle they call *Coherent Indifference* is inconsistent with Stochastic Dominance. Chapter 3contains a detailed discussion of this alternative approach. Notice, however, that the Random Variable Affine Invariance principle is not motivated by the idea that the interval scale structure of outcomes determines the ordering of prospects.

4.4 Theoretical Symmetry in Extensions of DTU

Standard approaches to decision theory, including DTU, keep track of the interval scale structure of outcomes by taking as primitive a utility function ('U', in the case of DTU). This provides a coordinate-system for the outcomes, which makes it easier to state and prove claims about the primary topic of decision theory, which is the ranking of the prospects $\succsim$. However, the utility function builds in more information than just the interval scale structure of outcomes: there is no zero point or unit in the interval scale structure, only the ordering of outcomes and the ratios of their differences.

Nevertheless, the comparisons given by DTU and other standard theories are sensitive only to the interval scale structure of outcomes. This can be observed by restating the theory in a coordinate-free way. Typically, a theory that takes a utility function as primitive can be stated in a coordinate-free way by replacing the name of the utility function with a variable in each axiom, and saying that the axiom holds for every utility function. In the case of DTU, we may replace the two axioms that mention the utility function U, Unbounded Utility and Simple Expected Utility Theory, with the following quantified axiom to get *coordinate-free* DTU:

REPRESENTABILITY There is a utility function: a measurable function from outcomes to real numbers f such that if μ and ν are simple, then $\mu \succsim \nu$ if and only if $\mathbf{E}_\mu f \succsim \mathbf{E}_\nu f$.

COORDINATE-FREE UNBOUNDED UTILITY Every utility function is a bijection from outcomes to real numbers.

DTU is evidently a conservative extension of coordinate-free DTU; all DTU adds is that U is the utility function already posited by the Representability axiom.

The same holds for other theories proposed in the literature. Consider, for example, the addition of L^1 Continuity to DTU. A coordinate-free version of this theory would add to coordinate-free DTU the following axiom:

COORDINATE-FREE L^1 CONTINUITY For every utility function U, L^1 Continuity holds, which is to say that $\succsim$ is continuous with respect to convergence in expectation of U for any utility function U.

It is easy to see that the coordinate-free principle is equivalent to the original. This means, once again, that DTU + L^1 Continuity appeals only to the interval scale structure of outcomes, and not to the additional information provided by the utility function.

The same goes for many other proposals in the literature, including Kenny Easwaran's *Weak Expectation Theory* (2008) or the stronger *Principal Value Theory*

(2014b),[6] as well as Mark Colyvan's Relative Expectation Theory (2008). This holds also for more radical departures from orthodox decision theory, such as the "Coherent Indifference" theory of Seidenfeld, Schervish, and Kadane (2009) which denies the identification of prospects with probability distributions. Even highly radical theories which complicate the definition of the interval scale structure of outcomes, such as the risk-averse theory of Lara Buchak (2013: 238) or the low-probability-ignoring view of Nick Smith (2014), end up positing an interval scale structure to outcomes determining the value prospects in terms of that structure, rather than in terms of a particular utility function.

This is to say: despite the superficial appeal to a utility function that is typical in stating these theories, which builds in more information than the interval scale structure of outcomes there is a general consensus in the literature that theories should appeal only to the interval scale structure of outcomes in constraining the ordering of prospects. The affine symmetries stand to explain why this is a good idea, by claiming that the probabilities and of outcomes and how they relate in terms of interval scale structure is sufficient to determine the comparison of any two prospects.

4.5 Consequences of Affine Symmetry

We now turn to the question of how adding the affine symmetry principles affects decision theory in DTU. The most important theorem is that the affine symmetry principles do not already follow from DTU. Therefore, there is no deductive argument from the observation that utility is only defined up to a positive affine transformation to either principle:

Theorem 29. *Neither affine symmetry principle is a theorem of* DTU*. The same is true when* DTU *extended by any or all of L^1 Continuity, Relative Expectation Theory, Weak Expectation Theory, or Principal Value Theory.*

Proof. Chapter 2, see Remark 2 for a countermodel to Reflection Anti-Invariance and Theorem 13 for a countermodel to Positive Affine Invariance. □

We turn then to the surprisingly subtle question of what the affine symmetry principles do add to DTU. Notice that if the affine symmetry principles are consistent, then they do not entail that a doubling of utility yields a doubling of overall value, in the sense that $\mu \sim 0.5\delta_0 + 0.5[2U]\mu$ where δ_x is the prospect that yields utility x with 100% probability; this identity of value provably fails when μ is a St Petersburg

[6] This is established in Easwaran 2014b; Easwaran also conjectures that the theory of Gwiazda 2014 has the same property.

prospect. By contrast, they do imply that adding to a fixed amount of utility to each outcome adds to the overall value of a prospect in the following sense that fifty-fifty between some prospect μ units and utility b, and fifty-fifty between that prospect μ plus b units of utility ($[U+b]\mu$) and utility 0, are equally good. More generally:

Theorem 30 (DTU + Sym)**.**

$$x\left[U+\frac{b}{x}\right]\mu+(1-x)\nu \sim x\mu+(1-x)\left[U+\frac{b}{1-x}\right]\nu \tag{4.1}$$

Proof. By affine symmetry we have for any μ and b,

$$\frac{1}{2}\left[U+\frac{b}{2}\right]\mu+\frac{1}{2}\left[-U-\frac{b}{2}\right]\mu \sim \delta_0.$$

Shifting utility up by $b/2$ yields

$$\frac{1}{2}\left[U+b\right]+\frac{1}{2}[-U]\mu \sim \delta_{b/2}$$

(4.1) is then equivalent to

$$\begin{aligned}
\delta_0 &\sim \frac{x}{2}\left[U+\frac{b}{x}\right]\mu+\frac{1-x}{2}\nu+\frac{x}{2}[-U]\mu+\frac{1-x}{2}\left[-U-\frac{b}{1-x}\right]\nu \\
&\sim x\left(\frac{1}{2}\left[U+\frac{b}{x}\right]\mu+\frac{1}{2}[-U]\mu\right)+(1-x)[-U]\left(\frac{1}{2}\nu+\frac{1}{2}[U+\frac{b}{1-x}]\nu\right) \\
&\sim x\left(\frac{1}{2}\left[U+\frac{b}{x}\right]\mu+\frac{1}{2}[-U]\mu\right)+(1-x)[-U]\left(\frac{1}{2}\nu+\frac{1}{2}\left[U+\frac{b}{1-x}\right]\nu\right) \\
&\sim x\delta_{\frac{b}{2x}}+(1-x)\delta_{\frac{b}{2(1-x)}}.
\end{aligned}$$

□

(Theorem 30 can fail in DTU without the affine symmetries; as witnessed in the model of DTU without L^1 Continuity in Chapter 2.)

This result can be extended, in the presence L^1 Continuity, to vindicate the central principle of Colyvan 2008, *Relative Expectation Theory*. Stating Relative Expectation Theory in the present setting, where prospects are identified with probability distributions, is a little awkward. Here, Colyvan's principle is stated by relating the prospects to be compared with random variables on $[0,1]$ that are distributed according to those prospects.[7]

[7]The principle is more natural in a setting where prospects are identified with random variables

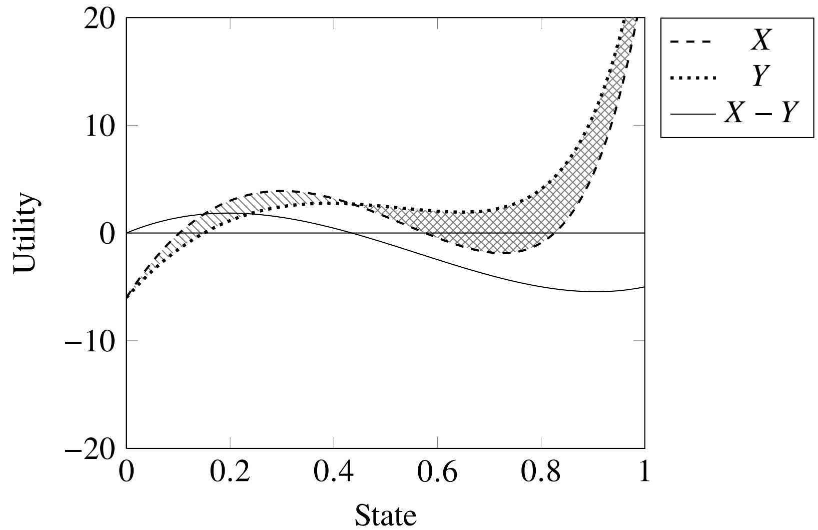


Figure 4.5: Graphs of the utility of outcome-valued random variables on $[0, 1]$ X and Y, and their difference $X - Y$. Relative Expectation Theory says that if $X - Y$ has an expected utility, then the distribution of X is at least as good as the distribution of Y just in case $X - Y$ has a nonnegative expected utility. Equivalently: if the hatched and crosshatched areas both have finite area, then the distribution of X is better than the distribution of Y just in case the northwest-hatched shape has an area at least as great as the crosshatched shape. Simple Relative Expectation Theory says the same in the case where $X - Y$ is simple (which is not so in the figure).

Relative Expectation Theory Let X and Y be any outcome-valued random variables on $[0, 1]$ with Lebesgue measure that are distributed according to μ and ν respectively. Let $X - Y$ be the outcome-valued random variable on $[0, 1]$ that gives yields utility $x - y$ when X yields utility x and Y yields utility y. Then if $X - Y$ has an expected utility, $X \succsim Y$ if and only if the expected utility of $X - Y$ is greater than or equal to 0. See Figure 4.5.

Like how Expected Utility Theory should be distinguished from Simple Expected Utility Theory when L^1 Continuity is not assumed, we also distinguish a version of Relative Expectation Theory that applies when the difference of the relevant random variable is simple:

Simple Relative Expectation Theory Let X and Y be outcome-valued random variables on $[0, 1]$ with Lebesgue measure that are distributed according to μ and ν respectively. Let $X - Y$ be the outcome-valued random variable on

over a space of "states", in the style of Savage 1954.

> $[0, 1]$ that gives yields outcome $x - y$ when X yields x and Y yields y. Then if $X - Y$ is simple (i.e., has finite codomain), then $X \succsim Y$ if and only if the expected utility of $X - Y$ is greater than or equal to 0.

Both principles are quite the mouthful, and may not seem particularly natural in this setting. This does not matter, since both can be derived from very natural assumptions:

Theorem 31 (DTU + Sym)**.** *Simple Relative Expectation Theory is true.*

Proof. First note that $\mu \succsim \nu$ is equivalent to $0.5\mu + 0.5\delta_0 \succsim 0.5\nu + 0.5\delta_0$. Let X_μ, X_ν, and X_ξ be distributed according to $0.5\mu + \delta_0$, $0.5\nu + 0.5\delta_0$, and $0.5\xi + 0.5\delta_0$ respectively such that $U \circ X_\mu = U \circ X_\nu + U \circ X_\xi$. Let o be one of the outcomes of ξ, and let X^1_μ modify X_μ by subtracting o from all those outcomes where X_ξ yields o, and adding the same probability of o to the region where X_μ yields utility 0. By Theorem 30, the distribution of X^1_μ is equally good as μ. Repeating this process for the n possible outcomes of ξ, we find the distribution of X^n_μ is equally good as μ, and the distribution of X^n_μ is $0.5\nu + 0.5\xi$, whence the result follows by Independence and Simple Expected Utility Theory. □

Corollary 32 (DTU + Sym)**.** *Relative Expectation Theory is equivalent to L^1 Continuity.*

We will say that μ differs from ν by ξ when for some X and Y distributed according to μ and ν, $X - Y$ is distributed according to ξ.

Notice that by Reflection Anti-Invariance, a symmetric prospect—a prospect μ that is equal to $[-U]\mu$—is equally good as zero utility. From Relative Expectation Theory it also follows that if μ and $[-U]\mu$ differ by a prospect of finite expectation ξ, then μ differs by the expectation of ξ from a symmetric prospect, so μ is equally good as the expected utility of ξ. From this observation, we may derive a useful strengthening of Expected Utility Theory that does not seem to have appeared in the literature. For a prospect μ, let F be the *cumulative distribution function* of μ, which maps each real number x to the probability that μ yields an outcome with utility at least x. When μ has an expected utility, it is given by the integral

$$\mathbf{E}_\mu U = \int_0^\infty F(z)\,\mathrm{d}z - \int_{-\infty}^0 1 - F(z)\,\mathrm{d}z$$

Set $F^+(z) = F(z)$, and set $F^-(z) := 1 - F(-z)$. Then when the expected utility of μ exists, we also have

$$\int_0^\infty F(z)\,\mathrm{d}z - \int_{-\infty}^0 1 - F(z)\,\mathrm{d}z = \int_0^\infty F^+(z) - F^-(z)\,\mathrm{d}z$$

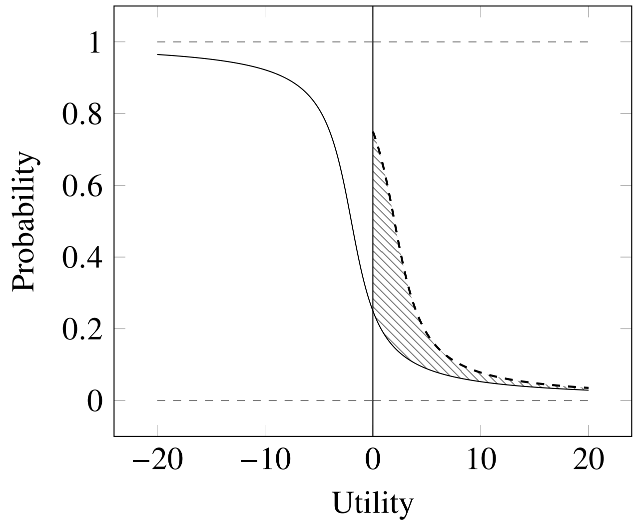


Figure 4.6: If the cumulative distribution function over utility for a prospect is the solid curve, then the folded expected utility of that prospect is equal to negative one times the hatched area, if that area is finite.

but the right-hand-side integral might exist even when the expectation does not. For example, if μ is symmetric but lacks an expectation, then $F^+ = F^-$, so the right-hand-side integral exists and is zero. Let the integral of $F^+ - F^-$ from 0 to infinity be the *folded expected utility* of μ.

Folded Expectation Theory If μ and ν have folded expected utilities (see Figure 4.6), then $\mu \succsim \nu$ if and only if the folded expected utility of μ is at least as great as that of ν.

Theorem 33 (DTU + Sym)**.** *Folded Expectation Theory follows from L^1 Continuity.*

Proof. Let μ be a prospect with cumulative distribution function F. Consider the random variables X^+, defined to agree with $(F^+(z))^{-1}$ almost everywhere where $F(z)$ is positive, and to be identically zero otherwise, X^-, which is the same but for F^-, and their difference $X^+ - X^-$ (see Figure 4.7). If μ has a folded expected utility, then the distribution of $X^+ - X^-$ has an expected utility equal to the folded expected utility of μ, so μ is equally good as its folded expected utility by Theorem 32. □

Folded Expectation Theory should be distinguished from some related proposals in the literature. It is (on its own) incomparable in strength with both Easwaran's Weak Expectation Theory (Easwaran 2008) and the stronger Principal Value Theory (Easwaran 2014b). It is not weaker since unlike these two proposals, Folded

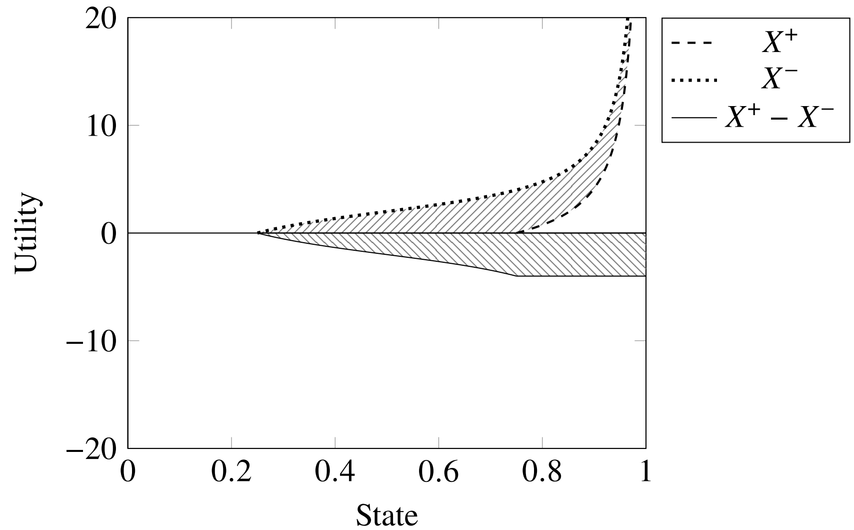


Figure 4.7: Graphs of the random variables X^+, X^-, and $X^+ - X^-$, where F is the cumulative distribution function from Figure 4.6. Notice that the two hatched shapes have equal area, and equal area to the hatched shape in Figure 4.6.

Table 4.2: The Pasadena prospect.

$1/2$	$1/4$	...	$1/2^n$	...
2	$-4/2$	...	$-(-2)^n/n$	...

Expectation Theory evaluates every symmetric prospect as equally good as utility 0; Easwaran's two principles sometimes fail to assign values to such prospects. It is not stronger than either principle for a more complicated reason. A prospect μ has a weak expectation when the following two conditions are met. First,

$$\int_0^t F^+(z) - F^-(z)\,\mathrm{d}\,z$$

must have a limit as t goes to infinity; which is the *principal value* of the integral. Second, μ must have 'thin tails': the probability that μ yields an outcome better than utility t or worse than $-t$ must decay with $o(t^{-1})$ as t goes to infinity. In this case the *weak expected utility* of the prospect is given by the principal value of the integral. These conditions are met when μ is the *Pasadena* prospect, which for each positive integer n yields utility $-(-2)^n$ with probability $1/2^n$ (see Table 4.2). Easwaran shows the Pasadena prospect to have a weak expected utility of $\ln 2$. By contrast, the prospect does not have a folded expected utility.

Proposition 34. *The Pasadena prospect does not have a folded expected utility.*

Proof. Let F be the cumulative distribution function for the Pasadena prospect. F^+ and F^- are the following piecewise constant functions:

$$F^+(z) = \begin{cases} 2/3 & 0 \le z < 2 \\ 2/3 - 1/2 & 2 \le z < 8/3 \\ 2/3 - 1/2 - 1/8 & 8/3 \le z < 32/5 \\ \dots & \dots \\ 2/3 - \sum_{k=1}^{n} \frac{1}{2^{2k-1}} & \frac{2^{2n-1}}{2n-1} \le z < \frac{2^{2n+1}}{2n+1} \\ \dots & \dots \end{cases}$$

$$F^-(z) = \begin{cases} 1/3 & 0 \le z < 2 \\ 1/3 - 1/4 & 2 \le z < 16/4 \\ 1/3 - 1/4 - 1/16 & 16/4 \le z < 64/6 \\ \dots & \dots \\ 2/3 - \sum_{k=1}^{n} \frac{1}{2^{2k}} & \frac{2^{2n}}{2n} \le z < \frac{2^{2n+2}}{2n+2} \\ \dots & \dots \end{cases}$$

$F^+ - F^-$ is computed to be the following:

$$F^+(z) - F^-(z) = \begin{cases} 1/3 & 0 \le z < 2 \\ 1/3 - 1/2 + 1/4 & 2 \le z < 8/3 \\ 1/3 - 1/2 + 1/4 - 1/8 & 8/3 \le z < 16/4 \\ 1/3 - 1/2 + 1/4 - 1/8 + 1/16 & 16/3 \le z < 32/5 \\ \dots & \dots \\ 1/3 + \sum_{k=1}^{n} \frac{1}{(-2)^k} & \frac{2^n}{n} \le z < \frac{2^{n+1}}{n+1}, n \ge 2 \\ \dots & \dots \end{cases},$$

which pleasingly simplifies to

$$F^+(z) - F^-(z) = \begin{cases} 1/3 & 0 \le z < 2 \\ (-2)^{-n}/3 & \frac{2^n}{n} \le z < \frac{2^{n+1}}{n+1}, n \ge 2 \end{cases}.$$

Notice that $|F^+(z) - F(z)|$ decays with $1/(3z)$, so $F^+(z) - F^-(z)$ is not integrable, meaning that the Pasadena prospect lacks a folded expected utility. □

For a nontrivial application of Folded Expectation Theory, we may evaluate

Table 4.3: The Arroyo prospect.

$1/2$	$1/6$	$\ldots$	$1/(n^2+n)$	$\ldots$
2	-3	$\ldots$	$(-1)^{n+1}(n+1)$	$\ldots$

Table 4.4: The Arroyo prospect + 0.5 units of utility.

$1/2$	$1/6$	$\ldots$	$1/(n^2+n)$	$\ldots$
2.5	-2.5	$\ldots$	$(-1)^{n+1}(n+1.5)$	$\ldots$

Bartha's (2016) *Arroyo* prospect, which for each positive n yields utility $-(-1)^n(n+1)$ with probability $1/(n^2+n)$. Like the Pasadena prospect, the Arroyo prospect's expected utility would be given by the divergent sum

$$1 - 1/2 + 1/3 + \ldots,$$

which conditionally converges to $\ln 2$. Bartha demonstrates that the Arroyo prospect does not have a weak expected utility. But it does have a folded expected utility of $\ln 2$:

Theorem 35 (DTU + Sym + L^1 Continuity)**.** *The Arroyo prospect is equally good as utility* $\ln 2$*.*

Proof. Let A be the Arroyo prospect, and let F be the cumulative distribution function for $[U{+}0.5]A$ (Table 4.4). Then $F^+(z) - F^-(z)$ is the cumulative distribution function of a prospect which yields utility $2n + 0.5$ with probability

$$\frac{1}{2n(2n-1)} - \frac{1}{2n(2n+1)},$$

and such a prospect has expected utility $0.5 + \ln 2$. □

DTU + Sym, with or without L^1 Continuity, also goes beyond Folded Expectation Theory in many ways. We shall find that the Pasadena prospect, unevaluable by Folded Expectation Theory, must receive utility $\ln 2$ on DTU + Sym + L^1 Continuity. To achieve this result, we will first give a simpler example of some the necessary techniques by evaluating the alternating St Petersburg prospect from Section 4.1 (Table 4.1), which yields utility $(-2)^n$ with probability $1/2^n$ for each positive integer n. The alternating St Petersburg paradox does not have folded expected utility, or a weak expectation, or a principal value (the relevant integral oscillates between -1 and 0). Nevertheless, we find:

Theorem 36 (DTU + Sym)**.** *The alternating St Petersburg prospect has utility* $-1/2$*.*

Proof. Let P be the alternating St Petersburg prospect. We have the identity

$$P = 0.5\delta_{-2} + 0.5[-2U]P.$$

Suppose $P \succ \delta_{-1/2}$. By the affine symmetry principles, it follows that $[-2U]P \prec \delta_1$. By Simple EUT and Independence, we have

$$0.5\delta_{-2} + 0.5[-2U]P \prec 0.5\delta_{-2} + 0.5\delta_1 \sim \delta_{-1/2}$$

so $P \prec \delta_{-1/2}$, a contradiction. An analogous contradiction arises from supposing $P \succ \delta_{-1/2}$, so by Totality we have $P \sim \delta_{-1/2}$. □

This result can be generalized to any prospect that is self-similar in the sense of being related to itself by a convex combination with an arbitrary prospect ν and itself under a negative affine transformation:

Theorem 37 (DTU + Sym)**.** *Suppose $\mu \sim (1-x)\nu + x[-aU+b]\mu$, with $a > 0$ and $1 > x > 0$. Then*

$$\mu \sim \sum_{n=0}^{\infty}(1-x)x^n \left[-a^n U + \frac{(-a)^n bx}{1-x}\right]\nu.$$

Proof. Define

$$E\mu := (\mu \sim (1-x)\nu + x[-aU+b]\mu),$$

notice first that by Theorem 30, $E\mu$ is equivalent to

$$\mu \sim (1-x)\left[U + \frac{bx}{1-x}\right]\nu + x[-aU]\mu.$$

We now show that solutions to E are equally good; which is to say that if $E\mu$ and $E(\mu')$ then $\mu \sim \mu'$. Suppose for contradiction that $E\mu$ and $E\mu'$ and $\mu \succ \mu'$. Then, by the affine symmetries and independence,

$$(1-x)\left[U + \frac{bx}{1-x}\right]\nu + x[-aU]\mu \prec (1-x)\left[U + \frac{bx}{1-x}\right]\nu + x[-aU]\mu',$$

so $\mu \prec \mu'$, a contradiction.

It remains only to show that $E\mu_E$, where

$$\mu_E := \sum_{n=0}^{\infty}(1-x)x^n\left[-a^nU + \frac{(-a)^n bx}{1-x}\right]\nu.$$

for this, it suffices to expand the definition as follows:

$$\begin{aligned}\mu_E &= (1-x)\left[U + \frac{bx}{1-x}\right]\nu + \sum_{n=1}^{\infty}(1-x)x^n\left[-a^nU + \frac{(-a)^n bx}{1-x}\right]\nu \\ &= (1-x)\left[U + \frac{bx}{1-x}\right]\nu + x[-aU]\mu_E,\end{aligned}$$

so $E\mu_E$ follows by reflexivity of $\sim$. □

Theorem 37 yields shows that a self-similarity of value of the form

$$\mu \sim (1-x)\nu + x[aU+b]\mu$$

has a unique solution in value when $a < 0$; which must therefore be the value of the unique prospect which solves the self-similarity equation of prospects

$$\mu = (1-x)\nu + x[aU+b]\mu$$

It is evident that the uniqueness of solution to the equation of value no longer holds when $a > 0$. For let $a = 1/x$, and $\nu = \delta_{(-xb)/(1-x)}$, and the resulting equation of value,

$$\mu \sim (1-x)\delta_{\frac{-xb}{1-x}} + x\left[\frac{1}{x}U + b\right]\mu$$

is solved for any simple μ.

Let us now return to the Pasadena prospect from Table earlier (4.2, originally discussed by Nover and Hájek 2004). The sum that would give its expected utility,

$$\sum_{n=1}^{\infty}\frac{1}{2^n}\frac{-(-2)^n}{n}$$

diverges, but the sequence of its partial sums in the given order,

$$1 - \frac{1}{2} + \frac{1}{3} - \dots$$

Table 4.5: The Highland Park prospect, H.

$1/2$	$1/4$	$1/8$	$1/16$	...	$1/2^{2n-1}$	$1/2^{2n}$	...
2	-4	$8/3$	$-16/3$	...	$2^{2n-1}/(2n-1)$	$-2^{2n}/(2n-1)$	...

conditionally converges to $\ln 2$. It is natural to suppose that the Pasadena prospect really is equally good as utility $\ln 2$, and this judgement has been supported by a variety of principles in the literature, including Weak Expectation Theory. Hájek (2014: 550), identifies five additional proposals as entailing this judgement, an observation he has called a 'hexangulation' argument for the Pasadena prospect being equally good as utility $\ln 2$. As previously mentioned, Hájek independently comes close to endorsing the affine symmetry principles (at the very least he seems willing to apply Easwaran's constraint—see Section 4.2—to rule out any theory which gives rise to specific counterexamples to the affine symmetry principles). It turns out that these two natural and seemingly unrelated opinions are not logically independent: the affine symmetry principles already imply a value of $\ln 2$ for the Pasadena prospect, given L^1 Continuity:

Theorem 38 (DTU + Sym + L^1 Continuity)**.** *The Pasadena prospect is equally good as the outcome of utility* $\ln 2$*.*

Proof. Consider a modification of the Pasadena prospect, the *Highland Park prospect* of Table 4.5, H, which sours the negative parts of the prospect by decreasing the denominator by 1. The Highland Park prospect worsens the Pasadena prospect by a prospect with a finite expectation; specifically the prospect

$$\frac{2}{3}\delta_0 + \frac{1}{3}\sum_{n=1}^{\infty}\frac{3}{2^{2n}}\delta_{\frac{-2^{2n}}{2n(2n-1)}}$$

which has expectation $-\ln 2$. By Relative Expectation Theory, the Highland Park prospect is $\ln 2$ units of utility worse than the Pasadena prospect.

Now, notice that the Highland Park prospect H has the following property

$$H = \frac{2}{3}Q + \frac{1}{3}[-2U]Q,$$

where

$$Q := \sum_{n=1}^{\infty}\frac{3}{2^{2n}}\delta_{\frac{2^{2n-1}}{2n-1}}.$$

Notice that Q and

$$\frac{3}{4}\delta_0 + \frac{1}{4}[4U]Q$$

also differ by a prospect of finite expectation, namely

$$\frac{3}{4}\delta_2 + \frac{1}{4}\sum_{n=1}^{\infty} \frac{3}{2^{2n-1}}\delta_{\frac{-2^{2n+1}}{(2n+1)(2n-1)}}$$

which has an expectation of 0, so

$$Q \sim \frac{3}{4}\delta_0 + \frac{1}{4}[4U]Q$$

and thus

$$Q \sim \frac{1}{2}\delta_0 + \frac{1}{2}[2U]Q$$

by Totality and Positive Affine Invariance. Then, by Reflection Invariance, H is equally good as 0. Therefore, since H worsens the Pasadena prospect by $\ln 2$ units of utility, the Pasadena prospect is equally good as the outcome of utility $\ln 2$. □

A natural conjecture is that $\mathsf{DTU}+\mathrm{Sym}+L^1$ Continuity entails one of Easwaran's Weak Expectation Theory or Principal Value Theory in addition to mere Folded Expectation Theory. Theorem 38 lends some plausibility to this conjecture, which is left for further work.

4.6 Models of Affine Symmetry

The main purpose of this section is to establish the consistency of $\mathsf{DTU} + \mathrm{Sym}$.

Theorem 39. *The affine symmetry principles are consistent with* DTU*, and with* $\mathsf{DTU} + L^1$ *Continuity.*

This will be done by constructing a class of $\mathsf{DTU} + \mathrm{Sym}$. Further investigation of this class of models will lead to better understanding of the limits of the theory.

A harmless simplification employed in the models will be the identification of the outcomes with the real numbers $\mathbf{R}$, with the real number x representing the outcome of utility x. Prospects may then be identified with the real-valued probability distributions. However, it will be useful to extend the convex space of probability distributions on $\mathbf{R}$ to the real vector space of *signed measures of finite total variation*.

Definition 4 (Signed measure/finite total variation)**.** A *signed measure* on $\mathbf{R}$ is an additive real-valued function on $\mathcal{M}$. A signed measure μ has *finite total variation* if where $\mathcal{I}$ ranges over disjoint classes of intervals,

$$\sup_{\mathcal{I}} \sum_{I \in \mathcal{I}} |\mu(I)| < \infty$$

Equivalently, a signed measure of finite total variation is a difference $\mu^+ - \mu^-$, where μ^+ and μ^- are finite positive measures. The probability distributions are signed measures that are positive and that have total variation 1; the other signed measures have no decision-theoretically significant interpretation, and may be thought of as imaginary sums and differences of prospects where probabilities are allowed to be greater than 1 or less than 0. In this setting, principles of decision theory like Independence and Stochastic Dominance still make sense when applied to imaginary prospects, and assuming this additional structure makes the models easier to work with. In particular, we may use the standard identification of linear orderings with *cones* in the vector space:

Definition 5 (Cone/Convex Cone/Strict Cone)**.** A (weak) *cone* is a set $C \subseteq V$ such that $x\mu \in C$ when $\mu \in C$ and $x > 0$.

A cone is *convex* when $x\mu + y\nu \in C$ whenever $\mu, \nu \in C$ and $x, y > 0$.

A *strict* cone is the set $C \setminus (-C)$, where C is a (weak) cone. When C is a cone, we write C^+ for the corresponding strict cone $C \setminus (-C)$, C^0 for the difference $C \setminus C^+ (= C \cap (-C))$, and C^- for $(-C^+)$.

A cone corresponds to a preordering $\geq_C$ of V, where $\mu \geq_C \nu$ iff $\nu \in \mu + C$, and $\mu >_C \nu$ iff $\nu \in \mu + C^+$, and $\mu =_C \nu$ iff $\nu \in \mu + C^0$.

When the cone is convex, the preordering $\geq_C$ is *linear*, in the sense that it satisfies for any positive x, y,

$$\mu \geq_C \nu \leftrightarrow x\xi + y\mu \geq_C x\xi + y\nu,$$

which is analogous to the Independence principle generalized to settings where probabilities needn't lie between 0 and 1.

We will build models of the affine symmetry principles by interpreting '$\succsim$' as the preordering of prospects determined by a cone $C_{\succsim}$ on the space $\mathcal{F}$ of signed measure of finite total variation. The principles governing '$\succsim$' are satisfied at least when the cone $C_{\succsim}$ satisfies the following corresponding properties:

Totality $\mathcal{F} = C \cup C^-$.

Preordering $\mathbf{0} \in C$ (reflexivity), C is convex (transitivity).

Independence C is convex.

Simple Expected Utility Theory If $\int x \,\mathrm{d}\,\mu(x) \geq 0$ converges, then $\mu \in C$ if and only if $\int x \,\mathrm{d}\,\mu(x) \geq 0$.

Stochastic Dominance If $\mu([x, \infty)) \geq \nu([x, \infty))$ for all $x \in \mathbf{R}$, then $\mu - \nu \in C$. If in addition, $\mu([x, \infty)) > \nu([x, \infty))$, then $\mu - \nu \in C^+$.

The affine symmetry principles are decomposed into the following three:

Scale Invariance When $a > 0$, $[aU]C \subseteq C$ and $[aU]C^+ \subseteq C^+$.

Reflection Anti-Invariance $[-U]C = (-C)$.

Prospect Shift Invariance When μ is a prospect (a probability distribution), $[U + b]\mu - (\mu + \delta_b) \in C^0$

In the setting of preorderings on $\mathcal{F}$ rather than the space of prospects, there is an important difference between scaling and shifting utility. Consider the imaginary prospect $\mathbf{0}$ which gives zero probability to every set. It is to be equally good as δ_0, which gives 100% probability to utility 0. But improving the payoffs of $\mathbf{0}$ by a unit returns $\mathbf{0}$ again, which is worse than δ_0 so-shifted, which is δ_1. So, whereas we will be considering Scale Invariance as applying to all signed measures in $\mathcal{F}$, the analogous principle to Shift Invariance is restricted to prospects.

It is straightforward to come up with cones that satisfy these constraints besides Totality. One example is the cone that determines the continuous eventual dominance of truncated expectation ordering $\sqsupseteq$ of Lauwers 2016 and Chapter 2, defined thus:

$$\mu \sqsupseteq \nu := (\forall \varepsilon > 0)(\exists t > 0)(\forall s > t)\left(\int_{-s}^{s} x \,\mathrm{d}\,\mu - \int_{-s}^{s} x \,\mathrm{d}\,\nu > -\varepsilon\right)$$

$C_{\sqsupseteq} = \{\mu - \nu : \mu \sqsupseteq \nu\}$ is the corresponding cone. $C_{\sqsupseteq}$ can also be expressed as the set of signed measures whose principle value is eventually always greater than any negative number:

$$\mu \in C_{\sqsupseteq} = (\forall \varepsilon > 0)(\exists t > 0)(\forall s > t)\left(\int_{-s}^{s} x \,\mathrm{d}\,\mu(x) > -\varepsilon\right)$$

Theorem 40. *The cone $C_{\sqsupseteq} = \{\mu - \nu : \mu \sqsupseteq \nu\}$ satisfies the constraints besides Totality.*

Proof. Chapter 2, Section 2.3.3. □

Note that $C_{\sqsupseteq}$ is not the only such cone, but it is a natural one with plausible decision-theoretic significance, and is easy to work with.

A model of the theory, then, can be generated by *extending* such a cone satisfying the constraints besides Totality, such as $C_{\sqsupseteq}$, to a cone that does have Totality while respecting the other constraints.

Definition 6 (Extension of a cone)**.** A cone D is said to (weakly) *extend* C if $C \subseteq D$ and $C^+ \subseteq D^+$; if $D \neq C$ then D is said to *strictly* extend C.

Some of the constraints are automatically had by any extension of a cone that satisfies them. These are the principles which specify exactly which elements a cone must have to satisfy them: Simple Expected Utility Theory, Stochastic Dominance, Prospect Shift Invariance, the reflexivity part of Preordering ($\mathbf{0} \in C$), and Totality (because there are no strict extensions to a cone which is already total). The other principles, Independence, Scale Invariance, and Reflection Anti-Invariance, constrain how the cone must relate to itself, and so will not be had by all extensions of a cone that has them.

Our strategy for constructing a cone satisfying the constraints including Totality, then, will be as follows: beginning with a cone like $C_{\sqsupseteq}$ that satisfies the constraints besides Totality, we will show that it is always possible to strictly extend this cone while still respecting the constraints of Independence, Scale Invariance, and Reflection Anti-Invariance. By Zorn's lemma, it will follow that there is an inextensible cone satisfying these principles, which will satisfy Totality.

Theorem 41 (Extensibility)**.** *Let C be a non-total convex cone satisfying Scale Invariance and Reflection Anti-Invariance. Then there is a strict extension of C satisfying those principles as well.*

Proof. Take $\mu \notin C \cup C^-$. Let C_μ be the smallest scale invariant and reflection anti-invariant convex cone containing μ:

$$C_\mu = \left\{ \sum_{i=1}^{n} \operatorname{sgn}(a_i) x_i [a_i U] \mu : n \geq 1, x_i > 0, a_i \neq 0 \right\}$$

We first show that C_μ intersects at most one of C^+, C^0, or C^-. Suppose without loss of generality the transformation

$$T_x^a := \sum_{i=1}^{n} \operatorname{sgn}(a_i) x_i [a_i U]$$

maps μ to an element of C^+ (otherwise substitute C^+ for C^0 or C^-), and consider an arbitrary transformation

$$T_y^b := \sum_{i=1}^{m} \operatorname{sgn}(b_i) y_i [b_i U].$$

$T_y^b(T_x^a \mu) \in C^+$ by the hypothesis that C is convex, scale invariant, and reflection anti-invariant. Notice that the two operations commute: $T_x^a \circ T_y^b = T_y^b \circ T_x^a$, so $T_x^a(T_y^b \mu) \in C^+$ as well. Therefore $T_y^b \mu$ cannot be in either C^0 or C^- since those two sets are scale invariant, reflection anti-invariant, convex, and disjoint from C^+.

Let D be $C + (C_\mu \cup \{\mathbf{0}\})$ if C_μ intersects C^+, $C + C_\mu + (-C_\mu)$ if C_μ intersects C^0, or $C + (-C_\mu \cup \{\mathbf{0}\})$ if C_μ intersects C^-, otherwise let D be $C + (C_\mu \cup \{\mathbf{0}\})$ (in effect supposing μ to be good when we could have consistently supposed it to be bad or neutral). Since D is the sum of scale invariant and reflection anti-invariant convex cones, D is also a scale invariant and reflection anti-invariant convex cone.

It is clear that $C \subseteq D$ and $C \neq D$. Now, suppose that $D = C + (C_\mu \cup \{\mathbf{0}\})$, then C_μ does not intersect $-C$, so $-C - (C_\mu \cup \{\mathbf{0}\})$ does not intersect C^+, so $C^+ \subsetneq D^+$. Similarly in the other two cases. So $C^+ \subsetneq D^+$, so D is a strict extension of C, as required. □

Remark 3. *The feature of rescaling and negating utility that ensures their extensibility is that they commute with the vector space operations (i.e., they are linear) and also commute with each other. This technique can therefore be used to establish the consistency of* DTU *with the monotonicity of any such class of operations with respect to of* $\succsim$.

This establishes the first part of Theorem 39. For the consistency of L^1 Continuity, it suffices to note that any extension of $C_\sqsupseteq$ has this property, since when μ and ν have a finite distance in the L^1 metric, $\mathbf{E}_{\mu-\nu}U$ is finite and is equally good as $\mu - \nu$ on $\sqsupseteq$.

References

Bartha, Paul (2016). "Making Do Without Expectations". In: *Mind* 125.499, pp. 799–827.

Buchak, Lara (2013). *Risk and Rationality*. Oxford University Press.

Colyvan, Mark (2008). "Relative Expectation Theory". In: *Journal of Philosophy* 105.1, pp. 37–44.

Easwaran, Kenny (2008). "Strong and Weak Expectations". In: *Mind* 117.467, pp. 633–641.

— (2014b). "Principal Values and Weak Expectations". In: *Mind* 123.490, pp. 517–531.

Gwiazda, Jeremy (2014). "Orderly Expectations". In: *Mind* 123.490, pp. 503–516.

Hájek, Alan (2014). "Unexpected Expectations". In: *Mind* 123.490, pp. 533–567.

Lauwers, Luc (2016). "Why Decision Theory Remains Constructively Incomplete". In: *Mind* 125.500, pp. 1033–1043.

Levi, Isaac (1990). *Hard Choices: Decision Making Under Unresolved Conflict*. Cambridge University Press.

Nover, Harris and Alan Hájek (2004). "Vexing Expectations". In: *Mind* 113.450, pp. 237–249.

Savage, Leonard (1954). *The Foundations of Statistics*. Wiley Publications in Statistics.

Seidenfeld, Teddy, Mark Schervish, and Joseph Kadane (2009). "Preference for Equivalent Random Variables: A Price for Unbounded Utilities". In: *Journal of Mathematical Economics* 45, pp. 329–340.

Smith, Nicholas J. J. (2014). "Is Evaluative Compositionality a Requirement of Rationality?" In: *Mind* 123.490, pp. 457–502.

Struik, D.J. (1953). *Lectures on Analytic and Projective Geometry*. Addison-Wesley mathematics series. Addison-Wesley Publishing Company.

Chapter 5

A St Petersburg Paradox for Risky Welfare Aggregation

An extremely natural and plausible principle of welfare aggregation is that the overall value of a prospect depends only on what it offers for each possible person. This is the principle of *Anteriority* (McCarthy, Mikkola, and Thomas 2020: 81):

Anteriority If each possible person is equally likely to exist in either of two prospects, and for each welfare level, each person is, conditional on their existence, equally likely to have a life at least that good on either prospect, then those prospects are equally good overall.

Anteriority must be rejected, on the grounds that it conflicts with an even more fundamental principle: the principle of *(Stochastic) Dominance*. Nevertheless, the attractiveness of Anteriority should not be underestimated.

The positive argument for Anteriority is that it meshes nicely with a plausible welfarist picture of population axiology, according to which overall value supervenes on the distribution of individual welfare. To see why, imagine that you have two prospects that satisfy the antecedent of Anteriority. From the perspective of every possible person's welfare, then, the two prospects look exactly the same: there is the same probability of that person not existing, and conditional on her existence, there is the same probability of her ending up with any particular welfare level on either prospect. If two prospects are exactly the same from the perspective of every possible individual's welfare in this way, then it is hard to see how the distribution of individual welfare throughout the prospects could support any discrimination between them.

Anteriority is a sort of weak *ex ante* Pareto principle for prospects with variable population. Unlike *ex ante* Pareto principles, Anteriority remains neutral between various popular ways of cashing out the welfarist vision. The principle is prima facie acceptable for Prioritarians (Parfit 1997) as well as Totalists, including those who deny that existence can be better or worse for an individual than nonexistence.

Ex Post Egalitarianism is straightforwardly incompatible with Anteriority (Fleurbaey and Voorhoeve 2013). Adherents of this view might hope to recover a weakened version of Anteriority that is restricted to pairs of prospects whose possible outcomes contain no inequality. Even this weakened version of Anteriority will be impugned by the following argument, given a slight strengthening of an auxiliary assumption.

Despite its apparent virtues, Anteriority must be rejected. The issue is not that it yields a counterintuitive verdict on a particular case, but that it conflicts with an even more fundamental principle of value:

Stochastic Dominance If for each possible outcome, the first of two prospects is at least as likely as the second to yield an outcome at least as good as that outcome, then the first prospect is at least as good as the second (in this case we say the first *dominates* the second). If for some possible outcome the first prospect is strictly more likely than the second to yield something at least that good, then the first is strictly better (and in this case we say it *strictly* dominates).

Dominance is a fundamental assumption of decision theory that has garnered widespread acceptance. The principle is also extremely intuitively compelling: the value of a prospect, one would have thought, is solely a result of the values of the outcomes it might yield, and the probability with which it yields them, and these considerations are unanimously in favour of a prospect that dominates.

Given some widely accepted assumptions about the structure of individual welfare, Anteriority and Dominance conflict. (The conflict is closely related to puzzles arising from the possibility of unbounded value (Broome 1995; Russell and Isaacs 2021) and of infinite populations (Vallentyne and Kagan 1997; Bostrom 2011). Neither possibility is assumed here.) Our first assumption concerns how welfare levels might be distributed in the world:

Mix-and-Match

(a) There are some infinitely many possible people such that for any way of assigning welfare levels to any finite number of them, there is an outcome in which those people are the only people to exist and they all have their assigned welfare levels.

(b) For any sequence of positive real numbers that sum to one, and for any way of assigning outcomes to those numbers, there is a prospect that yields each of those outcomes with probability equal to the corresponding number.

Both parts of Mix-and-Match are typically assumed without note (Nover and Hájek defend part (b) (2004: 246–47)). Three more principles codify very plausible assumptions about the structure of value:

TRANSITIVITY The relation *being equally good overall as* is transitive.

PAIRWISE ANONYMITY Switching out someone for another person with the same welfare level doesn't change the overall value of an outcome.

ADDITION COMPARABILITY There is some outcome Ω with a finite population, and some welfare level x, such that Ω is either better or worse overall than an outcome in which two people at welfare level x are added to the people in Ω, and everybody else's welfare stays the same. (Refuting the restriction of Anteriority to prospects whose outcomes contain no equality will require the further assumption that everybody in Ω has welfare level x.)

Transitivity is usually taken to be obvious. Pairwise Anonymity is mostly uncontroversial, at least as applied to finite-population outcomes as it will be here. Vallentyne and Kagan (1997) show that problems arise for a natural strengthening of Pairwise Anonymity in infinite-population cases, but this does not make the failure of the weaker Pairwise Anonymity in finite-population cases very plausible. Addition Comparability is also seemingly acceptable: surely you would discover that things are worse than you thought if there turn out to be two more people than you thought, and their lives are horrifically terrible. Notice that Addition Comparability does not require the controversial assumption that some lives are better or worse than nonexistence for the people who live them.

The inconsistency between Anteriority and Dominance can now be demonstrated. Here is one way to do so, where the outcome Ω and the welfare level x are witnesses to Addition Comparability. First, take two infinite sequences of possible people $i_1, i_2, \ldots$ and $j_1, j_2, \ldots$, such that these sequences have no members in common with each other or with the population of Ω. Then, for each number n, let v_n be that outcome in which exactly i_1 through i_n are added to the population of Ω, all at welfare level x. Similarly, let w_n be that outcome where j_1 through j_n are added at welfare level x instead. Let $v_n^\frown w_n$ be an outcome where, in addition to the people in Ω, those from v_n and w_n also exist at welfare level x.

Now consider the following sequence of prospects (A) through (F). I write the prospects as tables, whose top row specifies the probability with which the outcome below it will come about, given that prospect:

(A)

$1/2$	$1/4$	$\ldots$	2^{-n}	$\ldots$
v_2	v_4	$\ldots$	v_{2^n}	$\ldots$

(B)	$1/4$	$1/4$	$1/8$	$1/8$	$\dots$	$\dots$	$2^{-(n+1)}$	$2^{-(n+1)}$	$\dots$	$\dots$
	v_2	v_2	v_4	v_4	$\dots$	$\dots$	v_{2^n}	v_{2^n}	$\dots$	$\dots$

(C)	$1/4$	$1/4$	$1/8$	$1/8$	$\dots$	$\dots$	$2^{-(n+1)}$	$2^{-(n+1)}$	$\dots$	$\dots$
	v_2	w_2	v_4	w_4	$\dots$	$\dots$	v_{2^n}	w_{2^n}	$\dots$	$\dots$

(D)	$1/4$	$1/4$	$1/8$	$1/8$	$\dots$	$\dots$	$2^{-(n+1)}$	$2^{-(n+1)}$	$\dots$	$\dots$
	Ω	$v_2{}^\frown w_2$	Ω	$v_4{}^\frown w_4$	$\dots$	$\dots$	Ω	$v_{2^n}{}^\frown w_{2^n}$	$\dots$	$\dots$

(E)	$1/4$	$1/4$	$1/8$	$1/8$	$\dots$	$\dots$	$2^{-(n+1)}$	$2^{-(n+1)}$	$\dots$	$\dots$
	Ω	v_4	Ω	v_8	$\dots$	$\dots$	Ω	$v_{2^{n+1}}$	$\dots$	$\dots$

(F)	$1/2$	$1/4$	$\dots$	2^{-n}	$\dots$
	Ω	v_4	$\dots$	v_{2^n}	$\dots$

Dominance entails that (A) and (B) are equally good. By Pairwise Anonymity, v_n and w_n are equally good for each n, so (B) and (C) must be equally good by Dominance. Anteriority implies that (C) and (D) are equally good. Applying Pairwise Anonymity and Dominance again yields that (D) and (E) are equally good. (E) and (F) must be equally good by Dominance, because $1/4$, $1/8$ and so on add up to $1/2$ exactly. The welfare level x was chosen, using Addition Comparability, so that v_2 is better or worse than Ω. By Dominance, then, (A) and (F) are not equally good, contradicting Transitivity.

There are many ways to respond to this puzzle, but investigating each option would take us too far afield. Notice, though, that the reasoning establishing the equal value of any adjacent pair of prospects from (A) through (F) seems unimpeachable except, perhaps, in the case of (C) and (D). For example, if you prefer (B) to (C), then either you regard v_n better than w_n for some n, which would be quite strange since for all I have said the people who would exist in either outcome may be exactly similar, or you think that v_n and w_n are equally good for any n, in which case it is completely obscure what feature of (B) is supposed to make it better than (C). The same goes for the comparison of (D) and (E), and that (A) and (B) and (E) and (F) are equally good is surely undeniable. Once it is granted that v_2 is not equally good as Ω, it is almost undeniable that (A) and (F) are also not equally good. The equal value of (C) and (D) seems by far the weakest link in the paradoxical chain of reasoning. So Anteriority must go.

Let us now consider a case wherein a choice between prospects (C) and (D) arises.

> Planetary Prospects:
> God will toss a fair coin until it lands tails, and she will record the number of tosses. Then, she will choose one of two buttons fairly at

> random, and press it. One of the buttons is red and the other is blue. As things stand, neither button does anything when pressed: your choice affects what they do when pressed. You are to modify one of the buttons so that when it is pressed 2^n Jovians are created on Jupiter, where n is the number of God's tosses. Just before you make your choice, the representative of Saturn will act similarly: she will modify the blue button so that should it be pressed, 2^n Saturnians will be created. Neither modification will affect the other, and any people created will have welfare level x.

Choosing to modify the red button yields prospect (C), and the blue button prospect (D), assuming that we would have outcome Ω were God to abstain from pressing. It was established above that (C) is better or worse than (D), so this choice is a weighty one. It is weighty even though no possible person gets a greater chance of existence, or a greater chance of a better life, given either choice.

Planetary Prospects may be used to cast doubt on patterns of reasoning employed elsewhere in population axiology. Parfit (1984: 419) observes that a certain version of Averagism implies that whether it is good to have children now depends on the average welfare and number of the ancient Egyptians. This consequence seems absurd. He writes: 'research in Egyptology cannot be relevant to our decision whether to have children'. By the same token, research into how things will go on Saturn ought to be irrelevant to what we should do on Jupiter, assuming that there will be no interaction between the planets. But Planetary Prospects shows that this thinking is wrong. Counterintuitively, information that has nothing to do with what you can affect *can* make a difference to the comparative value of the prospects available to you. Specifically, the information that the representative of Saturn chose to modify the blue button is relevant to which button you should modify. Parfit's Egyptology objection to Averagism is therefore rendered indecisive.

(Superficial adjustments to the case show exactly that information about what went on in ancient Egypt could be relevant to the values of prospects now. Specifically, make it so that God's coin tossing and button pressing all occurred in the past, and so that your choice is between 2^n children at welfare level x being created now if God pressed the red button, and those same children being created if God pressed the blue button. Furthermore, make it so that Ramesses II faced a similar situation in ancient Egypt thousands of years ago, and that any Egyptian children created in this way were cordoned off so as to not affect the rest of history.)

Given that Anteriority is to be rejected, it is reasonable to ask whether some weaker principle in the vicinity can replace it. One natural candidate is the restriction of Anteriority to finite-population prospects. To be precise, let a *finite-population prospect* be a prospect on which there are only finitely many possible people who have a chance of existing. *Finite Anteriority* is the restriction of Anteriority to finite-population prospects:

FINITE ANTERIORITY If each possible person is equally likely to exist in either of two *finite-population* prospects, and for each welfare level, each person is conditional on their existence equally likely to have a life at least that good on either prospect, then those prospects are equally good overall.

Since most of the classical problems in population axiology arise in finite-population cases, Finite Anteriority will, in those cases, do any theoretical work that one could hope to use Anteriority for. Moreover, Finite Anteriority is consistent with all of the previously employed principles besides Anteriority (the proof is too involved to include here).

Should Finite Anteriority be accepted by those initially inclined towards Anteriority? The answer to this question turns in part on whether motivation can be found for Finite Anteriority that does not extend to Anteriority proper. After all, if an argument for Finite Anteriority also works in favour of Anteriority, there must be something wrong with the argument.

The project of motivating Finite Anteriority without overgeneration is a difficult one, but not without promise. Many initially plausible principles begin to break apart when applied in full generality, but often the core idea need not be abandoned. The above argument against Anteriority in its full generality should be taken as a warning for those who endorse Finite Anteriority, that utmost care must be taken in this vicinity.

There is a related difficulty for Finite Anteriority that also warrants consideration. It is a problem even for the fixed-population restriction of Finite Anteriority. To generate the problem, suppose that Totalism is true for finite populations, which is to say that the overall value of an outcome is given by the sum of real-valued welfare levels of its inhabitants. Now consider the prospects (I) and (II), each of which involves the certain existence of only Ann and Bob (this example is inspired by one of Seidenfeld, Schervish and Kadane (2009: 333)):

(I)		$1/2$	$1/4$	$\ldots$	2^{-n}	$\ldots$
	Ann	2	4	$\ldots$	2^n	$\ldots$
	Bob	2	4	$\ldots$	2^n	$\ldots$
	Total	4	8	$\ldots$	2^{n+1}	$\ldots$

(II)		$1/4$	$1/4$	$1/8$	$1/8$	$\ldots$	$\ldots$	$2^{-(n+1)}$	$2^{-(n+1)}$	$\ldots$	$\ldots$
	Ann	4	2	8	2	$\ldots$	$\ldots$	2^{n+1}	2	$\ldots$	$\ldots$
	Bob	2	4	2	8	$\ldots$	$\ldots$	2	2^{n+1}	$\ldots$	$\ldots$
	Total	6	6	10	10	$\ldots$	$\ldots$	$2+2^{n+1}$	$2+2^{n+1}$	$\ldots$	$\ldots$

By Finite Anteriority (and even the restriction of Finite Anteriority to fixed population cases), (I) and (II) are equally good. Since $1/4$, $1/8$ and so on add up to exactly $1/2$, both people are in both prospects faced with a half chance of welfare level 2, a quarter chance of welfare level 4 and so on. On the other hand, given Totalism for outcomes with finite populations, (II) strictly dominates (I) overall and is thus strictly better.

There are numerous ways out of this puzzle, for those attracted to Finite Anteriority. One is to deny Totalism for fixed-population comparisons. Another is to deny that there is a sequence of welfare levels with the required structure, for instance by positing that there are finite bounds on how good or bad a life can get. Nevertheless, it is clear that even extensive restrictions of Anteriority do not automatically get us out of trouble. Those of us who still hope to employ Anteriority-like principles in the foundations of population axiology have a lot of work to do.

5.1 Postscript: Consistency of Finite Anteriority

The original paper suggests but does not prove the consistency of Finite Anteriority or related weakenings of Anteriority. Finite Anteriority is consistent when each person's individual utility function is bounded. To see why, suppose that every possible person's welfare ordering is represented by a $[0, 1]$-valued utility function, and satisfies the axioms of DTU besides Unbounded Utility (Chapter 2). Then Finite Anteriority is guaranteed to hold if the utility scale for the overall value ordering extends the ordering given by the expected sum of individual utility. This in turn holds if the overall utility function is given by the sum of individual utility functions, for outcomes with finite population, which is a conservative extension of DTU.

This strategy for proving consistency will also work if individual utility scales are unbounded, when Finite Anteriority is further restricted to prospects where each individual utility scale has an expectation.

In settings with the possibility of infinite populations, DTU for the overall ordering is implausible given Finite Anteriority. Since Finite Anteriority (with Pairwise Anonymity) says that increasing the number of happy people increases overall utility without bound, one would expect that an infinite number of happy people would be above any finite utility, which is denied by DTU. A preferable theory would be DTU minus the assumption that outcomes have real-valued utilities, and with additional assumptions about how individual welfare interacts with the overall value of prospects. Exactly which form such a theory should take is left for further work.

References

Bostrom, Nick (2011). “Infinite Ethics”. In: *Analysis and Metaphysics* 10, pp. 9–59.

Broome, John (1995). “The Two-Envelope Paradox”. In: *Analysis* 55.1, pp. 6–11.

Fleurbaey, Marc and Alex Voorhoeve (2013). “Decide As You Would With Full Information! An Argument Against Ex Ante Pareto”. In: *Inequalities in Health: Concepts, Measures, and Ethics*. Ed. by Ole Norheim et al. Oxford University Press.

McCarthy, David, Kalle Mikkola, and Teruji Thomas (2020). “Utilitarianism with and Without Expected Utility”. In: *Journal of Mathematical Economics* 87, pp. 77–113.

Nover, Harris and Alan Hájek (2004). “Vexing Expectations”. In: *Mind* 113.450, pp. 237–249.

Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press.

— (1997). “Equality and Priority”. In: *Ratio* 10.3, pp. 202–221.

Russell, Jeffrey Sanford and Yoaav Isaacs (2021). “Infinite Prospects”. In: *Philosophy and Phenomenological Research* 103.1, pp. 178–198.

Seidenfeld, Teddy, Mark Schervish, and Joseph Kadane (2009). “Preference for Equivalent Random Variables: A Price for Unbounded Utilities”. In: *Journal of Mathematical Economics* 45, pp. 329–340.

Vallentyne, Peter and Shelly Kagan (1997). “Infinite Value and Finitely Additive Value Theory”. In: *Journal of Philosophy* 94.1, pp. 5–26.

Bibliography

Arntzenius, Frank, Adam Elga, and John Hawthorne (2004). "Bayesianism, Infinite Decisions, and Binding". In: *Mind* 113.450, pp. 251–283.

Arntzenius, Frank and David McCarthy (1997). "The Two Envelope Paradox and Infinite Expectations". In: *Analysis* 57.1, pp. 42–50.

Askell, Amanda (2018). "Pareto Principles in Infinite Ethics". PhD thesis. New York University.

Bales, Adam, Daniel Cohen, and Toby Handfield (2014). "Decision Theory for Agents with Incomplete Preferences". In: *Australasian Journal of Philosophy* 92.3, pp. 453–70.

Bartha, Paul (2016). "Making Do Without Expectations". In: *Mind* 125.499, pp. 799–827.

Beckstead, Nick and Teruji Thomas (2021). "A Paradox for Tiny Probabilities and Enormous Values". In: *GPI Working Papers* 7.

Bernoulli, Daniel (1954). "Exposition of a New Theory on the Measurement of Risk". In: *Econometrica* 22.1, pp. 23–36.

Blackwell, David A and Meyer A Girshick (1954). *Theory of games and statistical decisions*. Wiley.

— (1979). *Theory of games and statistical decisions*. Courier Corporation. *Theory of games and statistical decisions*. Wiley, 1954.

Bolker, Ethan D. (1967). "A Simultaneous Axiomatization of Utility and Subjective Probability". In: *Philosophy of Science* 34.4, pp. 333–340.

Bostrom, Nick (2011). "Infinite Ethics". In: *Analysis and Metaphysics* 10, pp. 9–59.

Broome, John (1995). "The Two-Envelope Paradox". In: *Analysis* 55.1, pp. 6–11.

Buchak, Lara (2013). *Risk and Rationality*. Oxford University Press.

Chalmers, David (2002). "The St. Petersburg Two-Envelope Paradox". In: *Analysis* 62.2, pp. 155–157.

Colyvan, Mark (2008). "Relative Expectation Theory". In: *Journal of Philosophy* 105.1, pp. 37–44.

Colyvan, Mark and Alan Hájek (2016). "Making Ado Without Expectations". In: *Mind* 125.499, pp. 829–857.

de Finetti, Bruno (1972). *Probability, Induction, and Statistics*. New York: John Wiley.

Dietrich, Franz and Christian List (2005). "The Two-Envelope Paradox: An Axiomatic Approach". In: *Mind* 114.454, pp. 239–248.

Easwaran, Kenny (2008). "Strong and Weak Expectations". In: *Mind* 117.467, pp. 633–641.

Easwaran, Kenny (2014a). “Decision Theory Without Representation Theorems”. In: *Philosophers’ Imprint* 14.

— (2014b). “Principal Values and Weak Expectations”. In: *Mind* 123.490, pp. 517–531.

Fine, Terrence (2008). “Evaluating the Pasadena, Altadena, and St Petersburg Gambles”. In: *Mind* 117.467, pp. 613–632.

Fleurbaey, Marc and Alex Voorhoeve (2013). “Decide As You Would With Full Information! An Argument Against Ex Ante Pareto”. In: *Inequalities in Health: Concepts, Measures, and Ethics*. Ed. by Ole Norheim et al. Oxford University Press.

Gibbard, Allan and William L. Harper (1978). “Counterfactuals and Two Kinds of Expected Utility”. In: *Foundations and Applications of Decision Theory*. Ed. by A. Hooker, J. J. Leach, and E. F. McClennen. D. Reidel, pp. 125–162.

Goodsell, Zachary (2022). “Tossing Morgenbesser’s Coin”. In: *Analysis* 82.2, pp. 214–221.

Gwiazda, Jeremy (2014). “Orderly Expectations”. In: *Mind* 123.490, pp. 503–516.

Hájek, Alan (2014). “Unexpected Expectations”. In: *Mind* 123.490, pp. 533–567.

Hájek, Alan and Michael Smithson (2012). “Rationality and Indeterminate Probabilities”. In: *Synthese* 187.1, pp. 33–48.

Hare, Caspar (2010). “Take the Sugar”. In: *Analysis* 70.2, pp. 237–247.

Harsanyi, John C. (1955). “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility”. In: *Journal of Political Economy* 63.4, pp. 309–321.

Hong, Frank (forthcoming). “Know Your Way Out of St. Petersburg: An Exploration of "Knowledge-First" Decision Theory”. In: *Erkenntnis*, pp. 1–20.

Huemer, Michael (2016). *Approaching Infinity*. New York: Palgrave Macmillan.

Jech, Thomas (2006). *Set Theory: The Third Millennium Edition, Revised and Expanded*. Springer Monographs in Mathematics. Springer Berlin Heidelberg.

Jeffrey, Richard (1965). *The Logic of Decision*. University of Chicago Press.

Joyce, James M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Kaplan, Mark (1983). “Decision Theory as Philosophy”. In: *Philosophy of Science* 50.4, pp. 549–577.

Kosonen, Petra (2022). “Tiny Probabilities of Vast Value”. PhD thesis.

Lauwers, Luc (2016). “Why Decision Theory Remains Constructively Incomplete”. In: *Mind* 125.500, pp. 1033–1043.

Lauwers, Luc and Peter Vallentyne (2016). “Decision Theory Without Finite Standard Expected Value”. In: *Economics and Philosophy* 32.3, pp. 383–407.

— (2017). “A Tree Can Make a Difference”. In: *Journal of Philosophy* 114.1, pp. 33–42.

Lenman, James (2000). “Consequentialism and Cluelessness”. In: *Philosophy and Public Affairs* 29.4, pp. 342–370.

Levi, Isaac (1990). *Hard Choices: Decision Making Under Unresolved Conflict*. Cambridge University Press.

Lewis, David (1981). “Causal Decision Theory”. In: *Australasian Journal of Philosophy* 59.1, pp. 5–30.

McCarthy, David, Kalle Mikkola, and Teruji Thomas (2020). “Utilitarianism with and Without Expected Utility”. In: *Journal of Mathematical Economics* 87, pp. 77–113.

McGee, Vann (1999). “An Airtight Dutch Book”. In: *Analysis* 59.4, pp. 257–265.

Meacham, Christopher J. G. (2019). “Difference Minimizing Theory”. In: *Ergo: An Open Access Journal of Philosophy* 6.

Nalebuff, Barry (1989). “Puzzles: The Other Person’s Envelope is Always Greener”. In: *The Journal of Economic Perspectives* 3.1, pp. 171–181.

Nover, Harris and Alan Hájek (2004). “Vexing Expectations”. In: *Mind* 113.450, pp. 237–249.

Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press.

— (1997). “Equality and Priority”. In: *Ratio* 10.3, pp. 202–221.

Quirk, James P and Rubin Saposnik (1962). “Admissibility and measurable utility functions”. In: *The Review of Economic Studies* 29.2, pp. 140–146.

Russell, Jeffrey Sanford (2020). “Non-Archimedean Preferences Over Countable Lotteries”. In: *Journal of Mathematical Economics* 88.May 2020, pp. 180–186.

— (2021). “On Two Arguments for Fanaticism”. In: *GPI Working Papers* 17.

Russell, Jeffrey Sanford and Yoaav Isaacs (2021). “Infinite Prospects”. In: *Philosophy and Phenomenological Research* 103.1, pp. 178–198.

Savage, Leonard (1954). *The Foundations of Statistics*. Wiley Publications in Statistics.

Schoenfield, Miriam (2014). “Decision Making in the Face of Parity”. In: *Philosophical Perspectives* 28.1, pp. 263–277.

Seidenfeld, Teddy, Mark Schervish, and Joseph Kadane (2009). “Preference for Equivalent Random Variables: A Price for Unbounded Utilities”. In: *Journal of Mathematical Economics* 45, pp. 329–340.

Smith, Martin (forthcoming). “Decision Theory and de Minimis Risk”. In: *Erkenntnis*, pp. 1–24.

Smith, Nicholas J. J. (2014). “Is Evaluative Compositionality a Requirement of Rationality?” In: *Mind* 123.490, pp. 457–502.

Struik, D.J. (1953). *Lectures on Analytic and Projective Geometry*. Addison-Wesley mathematics series. Addison-Wesley Publishing Company.

Vallentyne, Peter and Shelly Kagan (1997). “Infinite Value and Finitely Additive Value Theory”. In: *Journal of Philosophy* 94.1, pp. 5–26.

von Neumann, John and Oskar Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Wilkinson, Hayden (2022). “In Defense of Fanaticism”. In: *Ethics* 132.2, pp. 445–477.