The asymmetry, uncertainty, and the long term

Teruji Thomas

Global Priorities Institute | October 2019

GPI Working Paper No. 11-2019

The Asymmetry, Uncertainty, and the Long Term

Teruji Thomas*

Working Paper, October 2019.
Your comments and questions welcome at teru.thomas@oxon.org.

Abstract

The Asymmetry is the view in population ethics that, while we ought to avoid creating additional bad lives, there is no requirement to create additional good ones. The question is how to embed this view in a complete normative theory, and in particular one that treats uncertainty in a plausible way. After reviewing the many difficulties that arise in this area, I present general 'supervenience principles' that reduce arbitrary choices to uncertainty-free ones. In that sense they provide a method for aggregating across states of nature. But they also reduce arbitrary choices to one-person cases, and in that sense provide a method for aggregating across people. The principles are general in that they are compatible with total utilitarianism and ex post prioritarianism in fixed-population cases, and with a wide range of ways of extending these views to variable-population cases. I then illustrate these principles by writing down a complete theory of the Asymmetry, or rather several such theories to reflect some of the main substantive choice-points. In doing so I suggest a new way to deal with the intransitivity of the relation 'ought to choose *A* over *B*'. Finally, I consider what these views have to say about the importance of extinction risk and the long-run future.

1 Introduction

Consider two possible long-term futures for humanity: the *Good Future*, containing 10^{20} flourishing future human lives, and the *Extinct Future*, containing no future human lives at all.[1] According to some views of population ethics, we would have

*Global Priorities Institute and Faculty of Philosophy, University of Oxford. Besides the other researchers at GPI, I am grateful to Timothy Campbell, Matthew Clarke, Daniel Cohen, Tomi Francis, Anders Herlitz, Kacper Kowalczyk, David McCarthy, Aidan Penn, Stefan Riedener, and Tatjana Višak for much useful discussion.

[1] To keep things simple, I'll consider only effects on the welfare of humans, or more generally 'people'. The numbers are simply for illustration, but are indicative of those found in Bostrom (2003).

incredibly strong reasons to bring about the Good Future rather than the Extinct one in a straight choice. Take, for example, *Totalism*, the view that we ought to act in a way that maximizes expected total welfare. The sheer number of people in the Good Future means that the total amount of welfare at stake in this choice would be vastly greater than any sacrifice the current generation could plausibly make, or any benefit they could endow upon themselves. Indeed, merely replacing a one-in-a-billion chance of the Extinct Future by a one-in-a-billion chance of the Good Future would justify the destruction of all the wellbeing of all the eight billion people currently alive—ten times over.

One could, perhaps, avoid such striking conclusions by appealing to non-welfarist considerations such as the rights of the present generation. But even within the domain of welfarist population ethics, many people are drawn to

> **The Asymmetry.** In a straight choice between creating no one and creating some additional people, with no effect on those who independently exist…
>
> 1. If the additional people would certainly have bad lives, we ought not to create them.
> 2. If the additional people would certainly have good lives, it is *permissible but not required* to create them.[2]

The Asymmetry entails that it would be permissible to choose the Extinct Future over the Good Future in a straight choice. This might not be the full story: just as one might supplement Totalism with a story about rights, one might interpret the Asymmetry as a *pro tanto* principle, and bring in other ingredients that would speak in favour of the Good Future. For example, if the continuance of humanity is morally important in non-welfarist ways, then it might turn out that, all things considered, we ought to choose the Good Future, but not at anything like the cost implied by Totalism. Be that as it may, my presumption is that there is *some* class of considerations—something like considerations of impartial beneficence—such that Totalism and the Asymmetry are straightforwardly disagreeing theories about what one ought to do as far as those considerations go; for the rest of this paper I am talking about what one ought to do, in just that sense.[3]

[2] See for example Roberts (2011) for a survey.

[3] I regret that I do not have more to say to elucidate this point. In practice, I will generally suppose that Expected Totalism holds in fixed-population cases; the question then is how fixed-population Expected Totalism can be plausibly extended to include the Asymmetry. However, most of the discussion will *not* be premissed on fixed-population Expected Totalism, and this way of developing the project should be of interest even to readers who think fixed-population Expected Totalism must ultimately be modified to incorporate egalitarian concerns, personal prerogatives, special obligations, deontic constraints, risk aversion, and whatever else.

The problem is that, unlike Totalism, the Asymmetry is nowhere near a complete theory, and in particular it is silent about what to do when we are uncertain about the outcomes of our acts—as indeed we always are.[4] What if, again, the most we can do is reduce the probability of extinction while imposing some more certain cost on those currently alive? Totalism, by incorporating expected value theory, provides a clean story about how to think about such choices in principle, no matter how complicated things might be in practice. As far as I am aware, there is no worked-out view that combines the Asymmetry with a plausible story about uncertainty.[5]

The goal of this paper is to fill this gap, and more generally to present a complete, extensionally plausible theory of the Asymmetry. The thrust of the paper is therefore more constructive than critical: the point is to beat a defensible path through the thicket of intuitions and theoretical puzzles that surround the Asymmetry, clearing the way for further exploration. This path-beating inevitably involves picking sides in some controversies, and I will make clear some of the main turning-points along the way. Indeed, I will ultimately present *four* possible destinations corresponding to different ways of resolving what strike me as the most important types of trade-off.

Here is the plan. In section 2, I present the best known extant approach to the Asymmetry, the so-called Harm Minimization View. I use this to introduce the main difficulties that arise in theorizing about the Asymmetry, and to lay down some markers. In particular I explain why it is difficult to reconcile the Asymmetry with expected value theory.

The centerpiece of the paper is section 3, in which I introduce some generic principles for choice under uncertainty. These principles are 'generic' in that they have nothing to do with the Asymmetry per se. In fixed-population cases (that is, in cases where the same people exist no matter what) they are compatible both with totalism and with 'ex post' prioritarianism, and allow for a wide range of views about how those fixed-population theories should extend to variable populations. These lead to a 'Supervenience Theorem' that reduces arbitrary choice scenarios to a class of simple choice scenarios, and which I prove in the Appendix. This class of simple choices can be taken either to be uncertainty-free choices or uncertain choices involving only one person. The Supervenience Theorem can thus be seen either as a way of aggregating across states of nature or across people.

[4] Throughout, I will understand uncertainty in an orthodox Bayesian way, using a probability distribution to represent the epistemic state of the agent (but not taking a stand on whether the relevant probabilities are purely subjective or objective, e.g. evidential). I will say nothing on the topic of 'Knightian uncertainty', 'ambiguity', or 'imprecise probabilities', although this is arguably an important area, especially when thinking about the long-run future. An out-of-the-box view would represent the agent's epistemic state by a set of probability measures, and use a 'liberal' decision principle, on which an option is permissible if and only if it would be permissible with respect to some probability function in the set (see Weatherson (2000); Moss (2015) for discussion).

[5] McDermott (1982), Meacham (2012), and Cohen (2019) do make brief suggestions on this front, which I will criticize below.

It remains, then, to produce a plausible theory of the Asymmetry for these simple choice scenarios. In section 4, I consider extant proposals for determining the set of permissible options in any given choice scenario by comparing available options two at a time. I show why these proposals are unsatisfactory, and make a better one, based on Schulze's 'beatpath' method in voting theory. In section 5 I use this proposal to sketch several detailed theories of the Asymmetry, incorporating different responses to the issues raised in section 2. I conclude in section 6 by illustrating what some of these views say about extinction risk and more generally about the importance of the long-run future. The Appendix contains a formal statement and proof of the Supervenience Theorem.

2 The Harm-Minimization View

To illustrate the initial problems that arise when we take the Asymmetry seriously, let me discuss a simple view, the so-called Harm Minimization View (HMV), which does seem to capture some important intuitions behind the Asymmetry.[6] While I must apologise to the reader for the intimidating array of cases, it is important to appreciate the range of issues raised by the Asymmetry in order to make progress towards a more satisfactory view. There is in any case a short summary at the end of the section, to which the reader may wish to refer.

The Harm Minimization View is formulated as follows, on the supposition that there is no relevant uncertainty. First, given an option set $\mathscr{C}$, we calculate the harm to each person in each option. By definition, the harm to person s in option x is the largest amount by which s could have been better off had some other option been chosen. Next, we can define the total harm of option x by adding up the harms to the people who exist in x. Finally, the rule is that we ought to choose the option with the least total harm.

Note that, in interpreting this view, we should at least formally count non-existence as a zero level of welfare. This ensures that having a bad life counts as a harm if one would not have existed under some alternative option, a fact crucial to the way in which the view validates the Asymmetry. Consider Case 1.

Case 1: Basic Asymmetry

Act 1		Act 2		Act 3	
Adam	0	Adam	0	Adam	0
Eve	—	Eve	−1	Eve	1

[6] See McDermott (1982). Another type of view is Kath's Shortfall Utilitarianism (Kath, 2016), which is more similar to some of the views I sketch in section 5.

Here Adam has the same welfare, zero, under each act. Under Acts 2 and 3 (but not Act 1) Eve also exists, with welfare -1 or 1 respectively, corresponding to a slightly bad or a slightly good life. Suppose first that only Acts 1 and 2 are available options. Act 1 contains no harm to anyone: the only existing person, Adam, would not have been better off under any alternative. But Act 2 contains one unit of harm to Eve, since she could have had zero (in the form of non-existence) rather than -1. So one ought to choose Act 1 over Act 2. (Whenever I say one ought to choose x over y, I mean it in this sense: one ought to choose x when only those two options are available.) Similarly focusing on Acts 1 and 3 only, neither option contains any harm, and therefore, out of those two acts, either one is permissible. These are the verdicts characteristic of the Asymmetry.

Although this seems like an insightful account of the Asymmetry *per se*, as a general theory it raises many issues.

2.1 Condorcet Violations

First of all, HMV violates popular requirements from decision theory, including contraction consistency. The essential point is that the relation 'ought to choose x over y' is not transitive.[7] This intransitivity is brought out by Case 2, a version of the well-known mere addition paradox.

Case 2: Mere Addition Paradox

Act 1		Act 2		Act 3	
Adam	100	Adam	105	Adam	95
Eve	—	Eve	105	Eve	125
Cain	—	Cain	105	Cain	125

HMV says that we ought to choose 2 over 1, 3 over 2, and yet 1 over 3. Moreover, this failure of transitivity is strongly suggested by the Asymmetry itself, and is not special to HMV. We ought to choose 2 over 1, because it benefits Adam while bringing Eve and Cain into existence with good lives.[8] We ought to choose 3 over 2, given that the same people exist in either option, and they are on average better off under Act 3.[9]

[7] Recall that a binary relation R is transitive iff for all x, y, z in its domain, $R(x, y) \wedge R(y, z) \to R(x, z)$.

[8] Although I think most people sympathetic to the Asymmetry will endorse this verdict, intransitivity may deliver some pressure to deny it. In particular, critical range utilitarianism, with an infinite critical range, is transitive and compatible with expected value theory, and delivers the Asymmetry; but denies that it is *ever* required to create extra people, however well off they would be and however much it would benefit independently existing people.

[9] It is true that, under Act 2, everyone is equally well off, a consideration that some people take to speak in its favour, compared to Act 3. However, if the numbers are chosen carefully enough, this should not seem to outweigh the difference in average welfare.

But it is at least permissible to choose 1 over 3, since doing so benefits Adam and there seems to be no obligation to bring Eve and Cain into existence.

Given that transitivity failures seem to be common in non-consequentialist theories (and, if we believe Temkin, perhaps even in consequentialist theories), I do not think that this amounts to a strong objection to HMV or to the Asymmetry itself. We should, however, be alive to the difficulties it raises, and in particular we should attend to how the theory produces verdicts in cases in which there are more than two options.[10]

When 'ought to choose x over y' is transitive, there will automatically be, in any finite set of options, at least one *undominated* option, that is, at least one option x such that, for any other option y, it is permissible to choose x over y. A very natural (if not unquestionable) view is that the permissible options include the undominated ones. Now, the problem is that when 'ought to choose x over y' is not transitive, there may be no undominated options, so—unless we want to recognize widespread moral dilemmas—we cannot rely on *this* principle for selecting permissible options. Nonetheless, when there are undominated options, they still intuitively have a special status. The following example from Jacob Ross illustrates this point.

Case 3: Improvable Life Avoidance[11]

Act 1		Act 2		Act 3	
Adam	10	Adam	20	Adam	−200
Eve	—	Eve	20	Eve	100

Act 2 is the outstanding option. Indeed, even by the lights of HMV, Act 2 is not only undominated but is the *Condorcet winner*: we ought to choose Act 2 over each other option in a pairwise choice. But HMV does not yield the verdict that, when all three options are available, we ought to choose Act 2. Instead, Act 1 is required. That's because Act 1 only contains 10 units of harm: Adam could have had 20 instead of 10. In contrast, Act 2 contains 80 units of harm, because Eve could have had 100 instead of 20. And Act 3 contains 220 units of harm. So Act 1 is the option that minimizes total harm.

To avoid such implausible patterns of permissibility, one might require that, if there are undominated options, then they are the permissible ones. However, there are cases in which this does not seem correct either (see my discussion of Diagram B in section

[10] I should mention two objections to intransitivity. One objection is that 'better than' is transitive, and so the verdicts about Case 2 cannot be understood in terms of betterness. I simply agree with this, which partly explains my focus on the 'ought to choose x over y' relation, instead of betterness. A second objection has to do with diachronic choice (see e.g. Andreou, Rabinowicz, and my discussion of Diagram D in section 4).

[11] Ross (2015, p. 440); in unpublished work, Višak emphasises this as a problem for Meacham's version of HMV.

4; I thank Matt Clark for pressing this point). What does seem generally compelling is the principle that the undominated options should be *among* the permissible ones. HMV violates this criterion,[12] and the similarly compelling

> **Condorcet Criterion.** One ought to choose the Condorcet winner, if there is one.

2.2 Uncertainty

The Harm Minimization View, as stated above, does not say how to handle uncertainty. Both McDermott and Meacham (2012) in their developments of HMV do make brief suggestions on this front, but both suggestions are implausible.

A first possibility, close to McDermott's proposal, is to compare acts based on their total *expected* harm (or, equivalently, their expected total harm). But consider Case 4.

Case 4: Anti-Natalism

Act 1	H	T	E	Act 2	H	T	E
Eve	—	—	—	Eve	100	100	-1

Here we are thinking of creating Eve, as in Act 2.[13] Her welfare depends on the toss of a coin. If the coin lands heads H or tails T, she will get welfare 100. But if it lands on its edge E—which we can suppose has a fantastically small probability—she will have a slightly bad life, at level -1. Because Eve would not suffer any harm on heads or tails, but would suffer a harm in the edge case, Act 2 contains some expected harm, whereas Act 1 contains none. The suggested extension of HMV then requires Act 1. Given that any given person *might* have a bad life, this amounts to a radical and, I think, unwanted form of anti-natalism. Indeed, a plausible extension of the Asymmetry would be

> **The Expectational Asymmetry.** In a straight choice between creating no one and creating some additional people, with no effect on those who independently exist…

[12] In the revised version of McDermott's view (McDermott, 2019) revised version of his view, both Acts 1 and 2 are permissible (though Act 2 is still the Condorcet winner). I think this is wrong—one really ought to choose the Condorcet winner—but at least more plausible than the verdict that Act 2 is impermissible. In any case the other problems with HMV are unaffected.

[13] In previous payoff tables, I included the independently existing Adam as well as Eve, and he might have been identified as the agent facing the choice. Here I am assuming that only Eve's interests are at stake, and for brevity exclude anyone else from the table. If bothered by the absence of an agent, the reader may suppose that an additional person, such as Adam, exists independently of the choice and has completely unaffected welfare. On all the concrete views I will consider, we are justified in ignoring such a person in evaluating the available options.

1. If the additional people would all have bad lives in expectation, we ought not to create them.
2. If the additional people would all have good lives in expectation, it is *permissible but not required* to create them.

While perhaps we should consider some variations on this principle (e.g. allowing moral oughts to be risk averse with respect to individual welfare), the Expectational Asymmetry seems to be much more in the right ballpark than the expected harm view.

McDermott's own very brief suggestion was to minimize the total expected harm under each act, but where, unusually, the harm to each person in each state is calculated relative to the expected wellbeing of that person under the alternatives. So, for example, in Case 4, in calculating the harm to Eve in state E under Act 2, the relevant fact is not that Eve would have been better off than -1 in state E under Act 1, but that she would have been better off than -1 in expectation under Act 1. But of course this yields the same verdict in this particular case. So Case 4 is a counterexample to McDermott's view as well.[14]

A different possibility is to identify the ('ex ante') harm to person s under act x as the largest amount of expected wellbeing by which s could have been better off under some alternative. This appears to be Meacham's proposal. The question then arises what to do about the possibility that Eve does not exist. If we treat non-existence as zero welfare, then we are in danger of lapsing into Totalism. The tension here is brought out by Case 5.

Case 5: Almost Certain Non-Existence

Act 1	H	T	E	Act 2	H	T	E	Act 3	H	T	E
Adam	10	10	10	Adam	10	10	10	Adam	15	15	15
Eve	10	10	10	Eve	—	—	—	Eve	—	—	15

Consider first Acts 1 and 2. Eve is sure to exist under Act 1. In Act 2, she is sure not to exist. The Asymmetry commits us to saying that either one of these acts is permissible, in a pairwise choice. On the other hand, Eve has zero expected welfare under Act 2, so if we just naively identify harm with differences in expected welfare, Eve suffers a significant harm under Act 2, and the harm minimization requires Act 1. To escape this conclusion, it won't do to carve out an exception based on the fact that Eve is *certain* not to exist under Act 2. For consider Acts 1 and 3. In Act 3, Eve has a fantastically tiny chance of existence, making her expected welfare close to zero. On the view on offer, Eve suffers a significant harm—almost 10 units worth—under Act 3.

[14]The view proposed by Cohen (2019) sounds initially like the 'expected total harm view' although it become clear in his discussion of the non-identity problem that this is not what's intended. Case 4 seems to be a counterexample to his view, either way.

This is more than enough to offset the significant benefit to Adam. But it is difficult to believe that it is permissible to choose Act 2 over Act 1 and yet impermissible to choose Act 3 over Act 1.

In comparing Acts 1 and 3 in Case 5, it is tempting to appeal to Eve's expected welfare *conditional on existence*: 10 under Act 1 and 15 under Act 3. The fact that 10 is less than 15 suggests a sense in which Eve would be harmed by Act 1, so that one ought to choose Act 3 over Act 1, even ignoring Adam. But yet another case shows that this cannot be the relevant sense of harm.

Case 6: Better Chance of a Bad Life

Act 1	H	T	E	Act 2	H	T	E
Eve	-10	-10	-10	Eve	—	—	-11

In Case 6, Act 1 gives Eve slightly higher expected welfare conditional on existence: -10 rather than -11. On the view of harm under consideration, one ought to choose Act 1 over Act 2. But Act 1 makes it certain that Eve will exist, and have a bad life, whereas Act 2 makes it all but certain that she will not exist at all. The first clause of the Asymmetry strongly suggests that one ought to choose Act 2 over Act 1, and indeed this seems like the right verdict.

Further views may come readily to mind. Instead of trying to settle the matter right away, let me point out one neglected kind of case that raises further puzzles.

Case 7: Interstate Non-Identity 1

Act 1	H	T	Act 2	H	T
Eve	1	—	Eve	—	10

Here Act 1 creates Eve if and only if H obtains; Act 2 creates Eve, rather better off, if and only if T obtains. I think it is quite unclear what to say about this case. On the one hand, it is very tempting to say that one ought to choose Act 2 over Act 1, and that one ought to do so for Eve's sake. And certainly it would seem *gratuitous* to choose Act 1. However, by the Asymmetry, if the (unaffected) agent were certain that H obtained, he would be permitted to choose Act 1; and so too if he knew that T obtained. That is, if he knew the outcome of the coin toss *either way*, he would be permitted to choose Act 1. Keeping that point very firmly in mind, it is hard to explain how the agent's as-it-happens ignorance about the coin could be relevant to the permissibility of these acts.[15]

[15] The case is similar to the sort of case considered by Hare (2010). A different way to run the argument is to rely on a distinction between objective and subjective oughts. One might think that, whether the coin is heads or tails, Act 1 is objectively permissible, so that the agent is (or can be) certain that Act 1 is objectively permissible; it seems strange then to claim that Act 1 is subjectively impermissible.

In thinking about this case, and others I will consider below, it is worth bearing in mind that the judgment that Act 2 is permissible, in the sense relevant to this paper, is compatible with the judgment that Act 2 is instrumentally better than Act 1. Indeed, insofar as one desires a theory of value in addition to a theory of oughts or requirements, a leading contender must be the view that the intrinsic value of outcomes tracks total welfare, and the instrumental value of acts tracks expected total welfare. It may however be that the underlying considerations that ground such evaluative facts do not always generate normative requirements: Act 2 is instrumentally better than Act 1, but not required. This picture gives at least some sense to the judgment that choosing Act 1 would be 'gratuitous'.

2.3 The Non-Identity Problem

A third issue is the famous Non-Identity Problem. Consider Case 8.

Case 8: Basic Non-Identity

Act 1		Act 2		Act 3	
Eve	1	Eve	—	Eve	10
Eve′	—	Eve′	10	Eve′	—

According to HMV, we ought to choose Act 3 over Act 1: Act 1 harms Eve. But it often happens that our actions change the identities of future people; instead of simply increasing Eve's wellbeing, as in Act 3, consider the option of creating a different person, Eve′, with higher wellbeing, as in Act 2. HMV says that it is permissible to choose Act 1 over Act 2. Nor is this an accidental peculiarity of HMV. At least heuristically, the second clause of the Asymmetry suggests that there is no reason to create additional people with good lives, and if so it is awkward to explain why there would be *more* reason to create one additional person rather than another.

It is in fact unclear to me what the right answer is here. Many people are attracted to the 'No Difference View' (Parfit, 1984, p. 367), to the effect that the choice between Acts 1 and 2 must be treated in the same way as the choice between Acts 1 and 3: it makes no difference that the identities of the people change. In an imperfect match to traditional terminology, I will call such views 'wide', as opposed to 'narrow' views like HMV. On the other hand, Boonin (2014) gives counterexamples to the No Difference View that others may find persuasive. For example, suppose that Eve in Act 1 is horse rather than a human, with the relatively low welfare typical of even good horse lives. It does not seem objectionable to bring this horse into existence (Act 1) rather than to bring human Eve′ into existence (Act 2); it certainly does not seem *as* objectionable as choosing Act 1 over Act 3.[16] A distinct possibility, in harmony with my comments about Case 7, is that although Acts 2 and 3 are equally good in instrumental terms, but they still have different normative status relative to Act 1.

In any case, what to say about Non-Identity cases is one of the main choice points in developing a theory that includes the Asymmetry. There are several different ways that one might try to modify HMV to obtain the No Difference View (or more generally the 'wide' verdict that one ought to choose Act 2 over Act 1 in Case 8). One might think (cf. Ross) that whatever it is that that explains the No Difference View involves evaluative considerations independent of those that are intended to be reflected in HMV. One must therefore supplement HMV with such heterogenous considerations. Meacham (2012), on the other hand, emphasises the temptation to identify Eve′ as a counterpart of Eve, and one might then understand harm as a difference in the wellbeing of counterparts. Even if Meacham's detailed characterisation of the counterpart relation is not especially compelling, the basic approach offers some useful flexibility. If we take the counterpart relation just to be transworld identity, then we obtain the original version of HMV.[17] But the counterpart relation might instead work in subtle ways that explain why the No Difference view is more compelling in some cases than in others; it might, for example, relate Eve and Eve′ when both are human, but not when Eve is equine. And of course the counterpart relation might suffer from some indeterminacy, allowing for different views about how to deal with that fact.

Independently of *how* the No Difference View is implemented, someone sympathetic to it might accept the following extension to situations involving uncertainty:

> **Same Expected Number Totalism.** If Acts 1 and 2 contain the same expected number of people, then in a choice between them, we ought to choose on the basis of expected total welfare.

The principle is illustrated by the verdict that one ought to choose Act 2 over Act 1 in Case 7, and similarly in the following closely related case.

Case 9: Interstate Non-Identity 2

Act 1	H	T	Act 2	H	T
Ann	1	—	Ann	—	10
Bea	—	1	Bea	10	—

[16] McDermott similarly denies the No Difference View. But he also suggests that putative non-identity cases in which the No Difference View is most compelling (like Parfit's classic conception case) may actually be identity cases.

[17] Indeed, a sanguine interpretation of Meacham's view holds that his counterpart relation *is* what counts as transworld identity in the appropriate normative context, so that his view is more of a clarification than a modification of HMV.

Here, Act 1 creates Ann if the coin lands heads and Bea if it lands tails. Act 2 creates Ann on tails and Bea on heads, each rather better off than she would have been, had she existed, under Act 1. Looking at each of Ann and Bea separately, Case 9 has essentially the same structure as Case 7, and one should presumably give parallel verdicts in Cases 7 and 9: either permitting both acts in each case, or, as Same Expected Number Totalism recommends, requiring Act 2 in each. What makes Case 9 of additional interest is that, considering each of heads and tails separately, the choice has essentially the same structure as the choice between Acts 1 and 2 in Case 8. In fact, the connection between Cases 8 and 9 is very close, in light of the following curious observation, inspired by Mahtani (2017) (see section 3.2 for further discussion of this argument). I described 'H' and 'T' in Case 9 as corresponding to the results of a coin flip, but abstractly they could just be states of any sort over which the agent is uncertain. And that uncertainty might include uncertainty about the identities of future people. Suppose in particular that, on H, Ann is Eve and Bea is Eve′, whereas, on T, Ann is Eve′ and Bea is Eve. Then a moment's inspection shows that the payoff tables for Acts 1 and 2 in Case 8, rewritten in terms of Ann and Bea, are exactly the same as those in Case 9.

The upshot of this discussion is that there is significant pressure to give parallel verdicts in Cases 7 and Case 8.[18] The theories I develop in this paper will all endorse this conclusion, because of the general principles I adduce in section 3.

2.4 Pro-Extinctionism

Here's a fourth issue. Consider Case 10.

Case 10: Mixed Addition 1

Act 1		Act 2	
Eve	—	Eve	10
Cain	—	Cain	10
Abel	—	Abel	-1

Here the option in Act 2 is to create Eve, Cain, and Abel; Eve and Cain would have good lives, but Abel would have a slightly bad one. HMV requires Act 1 over Act 2: the only harm falls to Abel in Act 2. As this shows, HMV typically favours extinction: insofar as the continuation of life on earth will lead to some people, like Abel, having bad lives, we ought not to allow them into existence.[19]

[18] A similar claim is made by Roberts, in her forthcoming 'The Better Chance Puzzle and the Value of Existence'.

[19] Note that we don't typically know *which* people would have bad lives. Applied to any realistic case, then, the argument here relies implicitly on assumptions about how we should treat certain kinds of uncertainty. The argument does go through given the principles I introduce in section 3. That we need to be able to argue in this way (or some other way) underlines the point of this paper.

If one does not like this view, one might instead require

> **The Group-Level Asymmetry.** Between creating no one, and creating some additional people, with no effect on those who independently exist…
>
> 1. If the additional people would certainly, *on average*, have bad lives, we ought not to create them.
> 2. If the additional people would certainly, *on average*, have good lives, it is permissible but not required to create them.

Just as there was a parallel between Cases 7 and 8, there is a parallel between Cases 4 and 10. For consider Case 11.

Case 11: Mixed Addition 2

Act 1	H	T	E	Act 2	H	T	E
Ann	—	—	—	Ann	10	10	−1
Bea	—	—	—	Bea	10	−1	10
Cat	—	—	—	Cat	−1	10	10

Looking at each of Ann, Bea, and Cat separately, the choice in Case 11 has the same structure as the choice in Case 4, and should presumably be decided in the same way. On the other hand, considering each state separately, Case 11 has the same structure as Case 10, and should also be decided in the same way. (As before, we can strengthen this point by noting that, given appropriate uncertainty about which of Ann, Bea, and Cat are Eve, Cain, and Abel, the payoff tables for Case 10 and Case 11 are equivalent.) So there is at least some pressure for parallel verdicts in Cases 4 and 10. In particular, we should not accept the radically pro-extinction verdict of HMV in Case 10 any more than we should accept the radically anti-natalist verdict in Case 4.[20] A more general argument—based on the Supervenience Theorem I will discuss in section 3—yields an equivalence between the Expectational Asymmetry and the Group-Level Asymmetry. Since I think the Expectational Asymmetry is clearly in the right ballpark, I take this as an argument that the Group-Level Asymmetry is in the right ballpark as well.[21]

Be that as it may, there is yet another sense in which HMV is pro-extinction. Consider Case 12.

[20] Meacham recognizes that the pro-extinction verdict seems problematic, and offers a brief discussion of external considerations that might overturn this verdict. As mentioned in the introduction, I am sympathetic to the existence of such external considerations. But it's important to realise just *how strongly* pro-extinction HMV seems to be: we must consider the harm implicit in creating potentially trillions of bad future lives, disregarding the potentially many more good ones. So it is at least unclear that there are external considerations strong enough to overturn the pro-extinction verdict of HMV.

[21] For example, I mentioned that one might modify the Expectational Asymmetry to take account of risk aversion; that corresponds to modifying the Group-Level Asymmetry to take account of inequality

Case 12: Costly Addition (Hard vs Soft)

Act 1		Act 2	
Adam	10	Adam	9
Eve	—	Eve	9
Cain	—	Cain	9

Here, in Act 2, Eve and Cain both have good lives. But their existence comes at a slight cost to Adam. According to HMV, the only relevant fact is the harm to Adam, and therefore Act 1 is required. So even if allowing the continuation of life on earth is permissible, it would be impermissible to do it at even the slightest cost to the present generation.

For later reference, I'm going to call this the *hard* verdict: more generally, a hard version of the Asymmetry is one on which creating additional good lives is impermissible (in an uncertainty-free pairwise choice) if it comes at any cost to the average welfare of those who exist independently. One might, on the other hand, think that both acts in cases like Case 12 are permissible: call this the *soft* verdict. There is a spectrum of possible views, but a paradigmatic soft version of the Asymmetry would hold that it is permissible to create additional good lives at some cost to those who exist independently, as long as total welfare does not decrease overall.

As a matter of fact, many theorists are attracted to the hard verdict. For example, consider a 'Repugnant Conclusion' choice between a world with many people with very good lives, and an arbitrarily more populous world, containing those same people and more, in which everyone has a barely good life. A typical response is that one ought to choose the first of these worlds, but the soft verdict supports the view that the second option is permissible: the sheer number of additional good lives can make up for the cost to the independently existing people. As another example, emphasised by Roberts , suppose you can either feed a starving child, or create an additional happy child; you ought to feed the starving child. While there is certainly something to this, intuitions surrounding the Repugnant Conclusion are notoriously murky, and I am unclear on whether Roberts's verdict relies on considerations that I would want to bracket from my discussion—for example, the kind of power relation in which the agent is naturally imagined to stand to the starving child. For this and for two more theoretical reasons, I will be non-committal about whether the hard verdict or the soft is the right one, and treat this as a major choice-point.

The two more theoretical reasons are as follows. First, there is a reasonably natural explanation of the Group-Level Asymmetry, at least in slogan form: creating an additional good life can *offset*, but not *outweigh*, the harm of creating an additional

aversion. It seems to me that, at least at a first pass, we may bracket both risk-aversion and inequality aversion, and in general I stick to views that agree with Expected Totalism in fixed-population cases. This is, however, an obvious direction in which one could explore further.

bad life. Or to put it another way, though there is no *requirement* to create additional good lives, doing so can *justify* the creation of additional bad ones.[22] But it is at least puzzling why the creation of additional good lives would be able to justify the creation of additional bad lives without also being able to justify harms to independently existing people.

Second, consider the following case:

Case 13: Better Chance of a Good Life

Act 1	H	T	E	Act 2	H	T	E
Eve	10	—	—	Eve	9	9	9

Roberts herself thinks that either act in Case 13 is permissible, and I agree that this seems plausible. But one can argue for the same sort of parallel between Cases 12 and 13 as there is between Cases 7 and 8 and between Cases 4 and 10. So there is some tension between the hard verdict that one ought to choose Act 1 in Case 12 and Roberts's preferred verdict that either act is permissible in Case 13.

2.5 Summary

Let me briefly sum up the main take-aways from this discussion. First, there are two broadly 'structural' difficulties in theorizing about the Asymmetry: dealing with the intransitivity of the 'ought to choose over' relation (section 2.1), and handling uncertainty in a sensible way (section 2.2). There are also at least two more 'substantive' issues: the Non-Identity Problem, leading to a 'wide' vs 'narrow' distinction (section 2.3); and Pro-Extinctionism, leading to a 'hard' vs 'soft' distinction (section 2.4). Along the way, I endorsed the Condorcet Criterion, and, more tentatively, the Expectational Asymmetry (section 2.2) and the Group-Level Asymmetry (section 2.4). Moreover, I argued for parallel verdicts in a number of cases (Cases 4 and 10, Cases 7 and 8, and Cases 12 and 13), prefiguring the supervenience principles I will develop in section 3.

3 Supervenience Principles

In the rest of this paper, I sketch out my own strategy for thinking about the Asymmetry. The strategy involves three stages, corresponding to the different types of issues raised in section 2.

[22] See Gert (2003), and McMahan (2013), for this sort of idea. While Gert contrasts 'requiring' and 'justifying' strength, McMahan contrasts the 'reason giving' and 'cancelling' (elsewhere 'offsetting') weight of various considerations. Although McMahan's paper is about population ethics, he applies his distinction most clearly to considerations within a single life. Presumably he would allow the same picture in interpersonal cases.

The first stage is to develop some relatively generic principles for dealing with uncertainty; these 'supervenience principles' are the topic of this section. They are sufficient to reduce arbitrary choices to the simpler class of ones in which there is no uncertainty, or, alternatively, to ones in which only one person's wellbeing is at stake.

The second stage, taken up in section 4, deals with the intransitivity of the 'ought to choose over' relation, as discussed in 2.1. I suggest a way of reducing multi-option choices to pairwise choices that is compatible with the Condorcet Criterion.

The third stage, discussed in section 5, is to decide the substantive issue of how to implement the Asymmetry in the remaining cases: pairwise uncertainty-free (or alternatively one-person) choices. Primarily, this means taking sides on the narrow/wide and hard/soft distinctions I introduced in section 2.

I said that the principles for dealing with uncertainty that I am about to introduce are 'generic', and it is worth saying something about their generality. In fixed-population cases—cases in which the same people would exist in every state under every act—the principles are most closely aligned with expected totalism and 'ex post' prioritarianism; they reject, for example, certain kinds of welfare-egalitarian considerations. The intention here is not rule out such forms of egalitarianism as false theories, but rather, as I sketched in the introduction, to focus on whatever sorts of reasons are well-captured by expected totalism (or if you like, ex post prioritarianism) in fixed-population cases. When it comes to variable population cases, I am, obviously, focussing on the Asymmetry, and I see nothing in the Asymmetry per se that lies in tension with these principles: they are, for example, compatible with the Harm Minimization View. But I should say that, conversely, there is nothing in the principles that requires the Asymmetry: they are compatible with expected totalism in its full generality, and also with weak versions of the Asymmetry on which our reasons to create good lives are merely weaker than our reasons not to create correspondingly bad ones. Relatedly, in the context of prioritarianism, they are also compatible with the plausible view that creating a good life has lower priority than improving someone's life from a neutral level to a good one. Finally, such principles are of interest not only to those who are inclined to accept them: they provide signposts for constructing theories of choice under uncertainty, and, insofar as they are prima facie plausible, make clearer what sorts of bullets one may need to bite.

I note that the supervenience principles, and the resulting theorem, are stated more formally in the Appendix, and the technically minded may prefer to read the presentation there. I should also mention that this material is inspired by the 'aggregation theorems' in McCarthy et al. (2016).

3.1 Personwise Supervenience

In what follows, I consider two sets of options: $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. The supervenience principles are going to give conditions under which a choice between the xs must be treated in the same way as a choice between the ys, in the sense that the kth option in the first choice is permissible if and only if the kth option in the second choice is permissible as well. Formally, I assume that there is a *choice function* C that maps each option set with n options to a subset of $\{1, \ldots, n\}$; the interpretation is that the kth option x_k is permissible in a choice between the xs if and only if $C(x_1, \ldots, x_n)$ contains k. When I say that a choice between the xs must be treated in the same way as a choice between the ys, I mean that $C(x_1, \ldots, x_n) = C(y_1, \ldots, y_n)$.

The first principle reflects the idea that if the two choice scenarios are the same from the point of view of each person, then the two choices must be treated in the same way:

> **Personwise Supervenience.** Suppose that for each person i, and every selection $w_1, \ldots, w_n$ of welfare levels (including possibly non-existence), the probability that i would get w_1 on the first option, but w_2 on the second, ..., and w_n on the nth, is the same whether the option set is $(x_1, \ldots, x_n)$ or $(y_1, \ldots, y_n)$. Then $C(x_1, \ldots, x_n) = C(y_1, \ldots, y_n)$.

The idea is illustrated by Case 14. In the first choice, between Acts 1 and 2, there is a 0.5 probability that Adam would get a_1 under the first act and a_2 under the second (namely, this happens on Heads), and a 0.5 probability that he would get b_1 under the first act and b_2 under the second (namely, this happens on Tails). Acts 3 and 4 are just the same except that I've switched the roles of the two equiprobable states from Adam's point of view. It's still true, for example, that there's a 0.5 probability that Adam would get a_1 under the first act and a_2 under the second, but now this happens on Tails rather than on Heads. (As far as Eve's welfare goes, the two option sets are exactly the same.) Personwise Supervenience says that the switch makes no difference to whether the first or second option is permissible. Note, to be clear, that Personwise Supervenience does *not* require us to treat a choice between Acts 1 and 2 in the same way as a choice between Acts 1 and 4: in the latter scenario, the probability that Adam would get a_1 under the first act but a_2 under the second is 0 rather than 0.5.

Case 14: Personwise Supervenience

Act 1	H	T
Adam	a_1	b_1
Eve	c_1	d_1

Act 2	H	T
Adam	a_2	b_2
Eve	c_2	d_2

Act 3	H	T		Act 4	H	T
Adam	b_1	a_1		Adam	b_2	a_2
Eve	c_1	d_1		Eve	c_2	d_2

Personwise Supervenience is related to the ex ante Pareto principle, and the motivation is broadly the same.[23] Some reasons one might have for denying ex ante Pareto—for example, egalitarian reasons—also suggest denying Personwise Supervenience.[24] But, again, I think it is reasonable to bracket such things as egalitarian considerations, and I don't see any reason directly relevant to the Asymmetry for denying Personwise Supervenience. Some other reasons for denying ex ante Pareto—for example, prioritarian ones—are compatible with Personwise Supervenience.

3.2 Statewise Supervenience

Here is the second supervenience principle.[25]

> **Statewise Supervenience.** Suppose that for each state s, and each selection $w_1, \dots, w_n$ of welfare levels (including possibly non-existence), the number of people who would, in s, get w_1 on the first option, but w_2 on the second, ..., and w_n on the nth, is the same whether the option set is $(x_1, \dots, x_n)$ or $(y_1, \dots, y_n)$. Then $C(x_1, \dots, x_n) = C(y_1, \dots, y_n)$.

To illustrate the principle, I'll use the same Acts 1 and 2 from the previous Case 14. For the choice between Acts 1 and 2, on heads, the first act gives one person a_1 rather than a_2 and gives one person c_1 rather than c_2. That's still true when we look at Acts 5 and 6 in Case 15.

[23] The ex ante Pareto *indifference* principle, specifically, says that if options x and y are equally good for each person considered separately, then they are equally good overall. One can try to apply ex ante Pareto to Case 14 in the following way: Acts 1 and 3, and Acts 2 and 4, are equally good for each person, and so equally good overall, by ex ante Pareto indifference; therefore Act 1 must be better than Act 2 to the same degree that Act 3 is better than Act 4. If one is happy to interpret $R(y/x)$ as the degree to which y is better than x, then this is also the conclusion of Personwise Supervenience. However, it is a bit obscure whether Acts 1 and 3 (say) *are* equally good for Adam if one of a or c stands for non-existence, and so a bit obscure whether ex ante Pareto indifference applies. The argument also relies on a version of transitivity. Personwise Supervenience does not suffer from these problems.

[24] Some egalitarians deny ex ante Pareto because they care not only about the welfare of individuals considered separately, but also about correlations between the welfare of different people. See for example McCarthy (2015). In Case 14, suppose that $a = 5$, while $b = c = d = 9$. Then a potential source of disanalogy between Acts 1 and 2 on the one hand and Acts 3 and 4 on the other is that Acts 1, 2, and 4 are all certain to result in perfect equality between Adam and Eve whenever they both exist, whereas Act 3 might result in inequality. So an egalitarian might deny Personwise Supervenience.

[25] As written, Statewise Supervenience only applies when literally the same set of states, with the same probabilities, are relevant to the two choice scenarios. But it could obviously be extended to cases in which there is a probability-preserving bijection between the relevant states. This extended principle is implied by the unextended one, given Personwise Supervenience.

Case 15: Statewise Supervenience

Act 5	H	T		Act 6	H	T
Adam	c_1	b_1		Adam	c_2	b_2
Eve	a_1	d_1		Eve	a_2	d_2

I've just switched *who*, on Heads, gets a_1 rather than a_2 and who, on Heads, gets c_1 rather than c_2. Meanwhile, what happens on Tails is unchanged. Statewise Supervenience says that this switch does not alter whether the first or second available option is permissible. Note that it does not say that the choice between Acts 1 and 2 must be treated in the same way as the choice between Acts 1 and 6: in the latter choice zero people rather than one person would, on Heads, get a_1 under the first option and a_2 under the second.

The basic motivation for Statewise Supervenience is that it encodes a form of impartiality: it doesn't matter *who* in each state is affected by the choice, only *that* some number of people are affected in the way that they are. This enables a plausible, and rather minimal, form of statewise reasoning. As in the case of Personwise Supervenience, there are possible counterexamples to Statewise Supervenience. Some of these have little to do with the Asymmetry and it seems reasonable to set them aside.[26] But other doubts are more immediately relevant. Recall the Interstate Non-Identity Case 7, in which we can either (Act 1) create Eve at level 1 on heads, or (Act 2) create her at level 10 on tails. Contrast it with the following case, in which we can either (Act 1) create Ann at level 1 on heads, or (Act 2) create *someone else* at level 10 on tails.

Case 16: Interstate Non-Identity 3

Act 1	H	T		Act 2	H	T
Ann	1	—		Ann	—	—
Bea	—	—		Bea	—	10

It isn't immediately obvious that we must treat Cases 7 and 16 in the same way, but that is what Statewise Supervenience demands.

While my main goal here is to put forth Statewise Supervenience as a modest working hypothesis, let me throw up one surprising obstacle to denying it, prefigured in section 2. Suppose that the agent's epistemic state involves uncertainty about whether (as in state H) Eve and Eve′ are Ann and Bea, in that order, or (as in state T) Bea and Ann. Then the payoff tables in Case 7, rewritten for Ann and Bea, are exactly the payoff tables in Case 16. More generally, if the xs and ys stand in the relation supposed by Statewise Supervenience, then, by changing the way in which we designate people ('Eve' and 'Eve′' or 'Ann' and 'Bea'?), we can transform the payoff tables of

[26] I include here the famous fairness-based example of Diamond (1967).

the xs into into payoff tables isomorphic to those of the ys. The natural conclusion is that a choice between the xs must be must be treated in the same way as a choice between the ys. (For further examples, recall that I already used this kind of move in section 2 to argue that Cases 8 and 9, and similarly Cases 10 and 11, must be decided in parallel, as Statewise Supervenience more directly requires.) Thus denying Statewise Supervenience carries with it the burden of explaining what is illicit about this (admittedly suspicious!) way of transforming cases, and providing a theory of choice under uncertainty that disallows it in just the right way. Mahtani,[27] in a slightly different context, convincingly argues that there is no easy way out. In particular, there is no easy way to identify a privileged way of designating people with respect to which we ought to analyse the cases.

3.3 Scale Invariance

In addition to Personwise Supervenience and Statewise Supervenience, I propose a third principle of *Scale Invariance*. It is also a kind of supervenience principle, and easiest to understand by looking at an example like Case 17.

Case 17: Scale Invariance

Act 1	H	T
Adam	10	0

Act 2	H	T
Adam	5	—

Act 3	H	T
Adam	10	0
Eve	10	0

Act 4	H	T
Adam	5	—
Eve	5	—

In Acts 1 and 2 only Adam has any possibility of existence. Acts 3 and 4 are very similar except that I introduce Eve as a clone of Adam, meaning that she has exactly the same welfare as Adam in each state. The Scale Invariance condition is that a choice between Acts 1 and 2 must be treated in the same way as a choice between Acts 3 and 4. More generally, say that one choice scenario is a *scaling* of another if it is obtained by simply 'scaling up' the number of people involved (see the Appendix for a formal statement).

> **Scale Invariance.** Suppose that the choice between the xs is a scaling of the choice between the ys. Then Then $C(x_1, \ldots, x_n) = C(y_1, \ldots, y_n)$.

[27] Mahtani (2017). See also Chalmers (2011) for the issue of identity relations across epistemic scenarios.

As with the other supervenience principles, there are some reasons one might resist Scale Invariance, as a general principle. One might, for example, be attracted to 'variable value' views, which behave like total utilitarianism for small populations and like average utilitarianism for large populations; this will lead to violations of Scale Invariance. Even setting aside standard critiques of such views, my main claim is that Scale Invariance is plausible as a restricted principle about reasons of beneficence.

3.4 The Supervenience Theorem

Combining the three supervenience principles brings us to the following result, which is the technical crux of this paper.

> **The Supervenience Theorem.** If Personwise Supervenience, Statewise Supervenience, and Scale Invariance hold, then the choice function is completely determined by its restriction to one-person cases (those in which only one person might exist), or, alternatively, by its restriction to uncertainty-free cases.[28]

Let me give some examples.[29] Suppose we agree with Totalism that in uncertainty-free cases one ought to choose an option with greatest total welfare. The unique choice function compatible with this rule, and satisfying the three supervenience principles, is Expected Totalism. Expected Totalism is also the unique theory satisfying the three supervenience principles such that, in one-person cases, one ought to choose an option with greatest expected welfare, treating non-existence as a zero level.[30] Varying this by treating non-existence as if a non-zero level of welfare leads to 'critical level' utilitarianism; using priority-weighted welfare rather than welfare itself leads to forms of ex post prioritarianism.

Suppose on the other hand we adopt the Harm Minimization View for uncertainty-free choices. The unique choice function compatible with this view, and satisfying the three supervenience principles, selects options that minimize expected total harm. So it decides one-person cases on the basis of expected harm. This view does capture the Asymmetry, but I reject it for reasons explained in section 2: that is, I reject the

[28]There are inevitably some technicalities, which I discuss in the Appendix, before proving the theorem. The main caveat is that the theorem only considers acts that result in finitely many possible welfare distributions, each with a rational probability. But it is hard to imagine that two individually plausible theories would agree about these rational-probability cases and disagree in general.

[29]The proof of the Supervenience Theorem, given in the Appendix, indicates how to actually reconstruct the choice function from its restriction to one-person or uncertainty-free cases. But one can verify the following examples by simply checking that the stated theory does satisfy the three supervenience principles and does restrict to one-person or uncertainty-free cases in the stated way.

[30]This is a variation on the way of understanding Expected Totalism developed in McCarthy et al. (2016), and indeed the Supervenience Theorem is a generalisation of the aggregation theorems found there.

HMV in general, for a number of reasons including the implausible pro-extinction verdict in Case 10, and I specifically reject this way of extending it to uncertain cases because, in the context of Case 4, it leads to an implausible form of anti-natalism.

Indeed, I explained in section 2.4 that these two implausible verdicts are related, by an informal argument that can be formalized by appeal to the three supervenience principles. Examining that argument helps to indicate the mechanics of the Supervenience Theorem. Statewise Supervenience entails that we must treat Case 10 in the same way as Case 11; Personwise Supervenience and Scale Invariance entail that we must treat Case 11 in the same way as Case 4. So these three cases must be decided in parallel. A general form of this argument shows that the Group Level Asymmetry and the Expectational Asymmetry are equivalent (with the caveat mentioned in fn. 28).

The argument for a parallel between Case 7 and the choice between Acts 1 and 2 in Case 8 admits a similar reconstruction. Statewise Supervenience entails that Case 9 must be treated in the same was as a choice between Acts 1 and 2 in Case 8, and Personwise Supervenience and Scale Invariance entail that it must be treated in the same was as Case 7. So these three cases must also be decided in parallel. A similar argument shows that Case 12 must be treated in the same was as Case 13.

Each of these arguments illustrates how, given the three supervenience principles, verdicts in uncertainty-free cases constrain verdicts in other cases, including one-person cases, and vice versa.

I should, however, explain the general rule that emerges from the Supervenience Theorem for reducing general choice scenarios to uncertainty-free ones. Suppose in an uncertainty-free choice scenario between $x_1, \dots, x_n$, there are at most m salient people, i.e. at most m people who exist under some option or another. For each of these m people there is an n-tuple of welfare levels, including possibly non-existence, corresponding to that person's welfare in each of the n options. For the choice function to satisfy Scale Invariance, the permissible options must be determined by the *number* of people who face each n-tuple of welfare levels, divided by m. What one sees from the Supervenience Theorem is that, in general, the permissible options must be determined in the same way by the *expected* number of people who face each n-tuple of welfare levels, divided by m.

4 Intransitivity

Having suggested how choice under uncertainty should be related to choice without uncertainty, I turn to writing down some illustrative theories of the Asymmetry compatible with this suggestion. I will develop these theories around a common core: a rule for determining the permissible options in any option set by comparing two options at a time, without regard to which other options are available. For example, the Condorcet Criterion partially characterises permissible options in terms of pairwise comparisons, and the idea is to strengthen this characterisation. The claim is not that such a rule *must* be available, but it is certainly convenient for theory-construction since it is much clearer what kinds of trade-offs are reasonable in two-option cases. In any case, several authors have already proposed rules along these lines. My contribution in this section is to explain some problems with these proposals, and to get a better one on the table.

Concretely, I will adopt a view from voting theory. Intransitivity arises in that context from the voting paradox: even if each individual has transitive preferences, it is possible for a majority to prefer x to y, a majority to prefer y to z, and a majority to prefer z to x. One may thus draw an analogy between the size of the majority in favour of x over y and how strong the reasons are for choosing x over y. Many of the methods developed for elections can be adapted to the present context, and deserve careful consideration. I am going to focus on Schulze's method (Schulze, 2011), commonly known as 'beatpath'; for further investigation, I can recommend Ranked Pairs (Tideman, 1987), and the variety of uncovered choice views surveyed by Duggan (2013), all of which have attractive formal properties and some conceptual plausibility.

In any case, say that one option x is *beatpath better* than another option y if and only if, for some α:

(i) There is a sequence of available options, leading from y to x, such that one ought to choose each option over the one before, and such that the reasons for doing so have strength at least α;

(ii) There is no such sequence leading from x to y.

Heuristically, one can deliberate from y to x, but not from x to y, by attending to reasons of strength at least α. The proposal is that an option is permissible if and only if there is none beatpath-better than it.

Schulze shows that beatpath-betterness is a transitive relation on each option set, with the effect that permissible options always exist.[31] Moreover, undominated options are always permissible, and any Condorcet winner will be the sole permissible option.

A proper defence of the beatpath view would be far beyond my purposes here; however, I will compare it to some of the competitors that have been proposed in similar contexts. To do so, let me first illustrate the verdicts of beatpath in some useful test cases, illustrated in Figure 1. In each diagram, the nodes $(x, y, z, \ldots)$ represent the

[31] Two important issues arise here, with more general relevance: incommensurable reasons, and infinite choice sets. First, the argument for transitivity relies on the strengths of reasons being completely ordered; it is not clear how one would accommodate incommensurability. Second, the claim that permissible options always exist depends on there being a finite set of options—but of course this is a familiar problem for all maximizing views. One familiar solution is to adopt a scalar view on which some options are simply better than others, with no fundamental role for categorical judgments of permissibility. It is easy to extend the Supervenience Theorem to such a scalar setting. The possibility of infinite option sets raises a more fundamental problems for some other views: e.g. Tideman's Ranked Pairs method is defined using induction from strong reasons to weaker ones, and this may not always make sense.

available options, and an arrow from, say, x to y indicates that one ought to choose y over x; the number labelling the arrow indicates how strong the reasons are to do so. Bolded options are the permissible ones, according to beatpath.

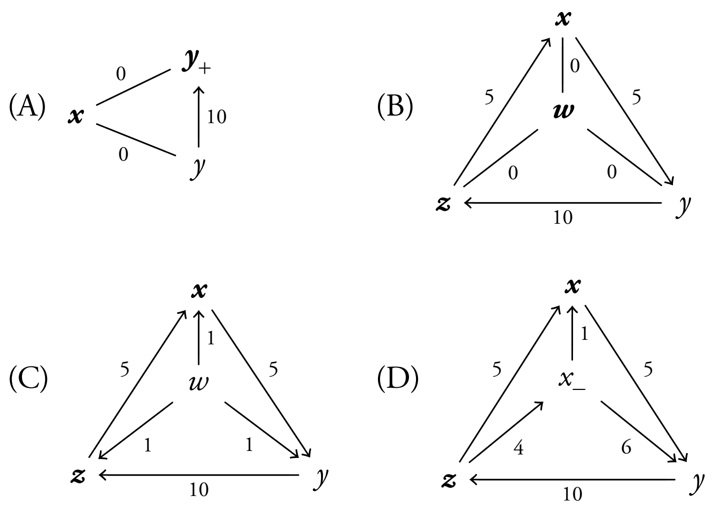

Figure 1: Diagrams A, B, C, D.

The first diagram, A, might correspond to a choice between a status quo (x) and creating Eve either with a good life (y) or with 10 additional units of welfare (y_+). When it comes to pairwise choices, either of x or y, and either of x or y_+, would be permissible, but y_+ ought to be chosen over y, and the strength of the reasons to do so is proportional to the benefit to Eve. Using beatpath, when all three options are available, one can deliberate from y to y_+, but not vice versa, by attending to reasons of strength at least 10; this renders y (but no other option) impermissible. This squares with the typical verdict about such a case.

The second diagram, B, represents a case in which there are three options x, y, and z which form a cycle. The fourth option, w, is such that the pairwise considerations are neutral between it and each other option. For example, x, y, and z might be something like the three options in Case 2, and w might be a fourth option in which no one exists at all.[32] Here the bottom line is that w, x, and z are permissible. This illustrates the point that, even though w is the sole undominated option (in the sense that there are no pairwise considerations in favour of something else), other options may still be permissible. Indeed, this seems like the right view about the four-option version of Case 2 just mentioned: it seems implausible that the only permissible option would be to create no one at all.

[32] I have chosen the strengths of the relevant reasons to illustrate what would happen on the 'narrow, hard' view of the Asymmetry, which yields a cycle in Case 2.

Diagram C presents a variant in which which there are some pairwise considerations that favour each other option over w, so that w would be the 'Condorcet loser'. Condorcet losers are always impermissible according to beatpath. As in B, we might identify x, y, z with the three options in Case 2, and w now with in option in which only one person, Abel, exists with a slightly bad life. It would seem wrong to choose w, even though the pairwise considerations against it are relatively weak.

Diagram D represents the sort of situation that arises in a 'value pump'. There is again a cycle between x, y, and z, as before. There is in addition a fourth option x_-, which is much like x but with a small penalty attached. Correspondingly, the reasons to choose x_- over z are slightly weaker than the ones to choose x over z, and the reasons to choose y over x_- are slightly stronger than those to choose y over x. (If, as in Case 2, Adam exists in all of x, y, and z, then x_- might differ from x only in that Adam is slightly worse off.) The value pump consists in the fact that, given a sequence of pairwise choices, one apparently ought to choose y over x, then z over y, then x_- over z, thus ending up with an option that is, in some intuitive sense, gratuitously worse than the original. To make the point sharper, the agent could be forced through a long spiral of options y_-, z_-, x_{--}, $y_{--}, \ldots$, where each minus sign represents a small penalty inflicted on Adam, finally ending up with a truly terrible outcome, e.g. one just like x except that Adam is a million units worse off. I am not addressing the issue of sequential choice here (see Rabinowicz (1995) for a survey), but in the scenario represented by the diagram, in which all the options that arise in the value pump are available at once, it seems clear that x_- must be impermissible, and beatpath delivers this verdict.

Vong (2018), being influenced by Temkin, frames the problem as one of dealing with cases in which the 'better than' relation is intransitive. What I call the strength of the reasons to choose y over x, and which I will denote $R(y/x)$, corresponds in his framework with the degree to which y is better than x. I don't find this a helpful way of looking at things; for example, one might instead identify betterness with the beatpath-betterness, in which case the beatpath view is that one ought to choose one of the best options. Still, let me accept Vong's framing in order to ease comparison. Vong's suggestion is that, relative to the option set $\mathscr{C}$, we should choose an option x that minimizes the sum $\sum_{y \in \mathscr{C}} R(y/x)$. Thus the availability of each option that is better than x weighs against x to the extent that it is better, and the availability of each option that is worse weighs in favour of x to the extent that it is worse. This is a natural thing to try: it reflects the ideology that the availability of each other option y yields some reason for or against choosing x, and one ought to weigh these reasons against each other. However, we should reject Vong's suggestion on three counts (thus indicating that this is not an appropriate ideology). First, Vong's rule does not automatically pick Condorcet winners, when they exist. It does not even do so for option sets with respect to which 'better than' is transitive. Second, in the case illustrated by diagram A, the availability of y weighs in favour of y_+, on Vong's view, but not in favour of x, incorrectly rendering x impermissible. Third, Vong's rule is implausibly sensitive to how options are individuated, violating a 'clone-independence' condition. Suppose that in an option set $\mathscr{C} = \{x, y, z\}$, y is better than x, z better than y, and x better than z, and by the same amount in each case. Then Vong's rule says, plausibly enough, that any option is permissible. But now consider the option set $\mathscr{C}' = \{x, y, y', z\}$, where y' is effectively a clone of y, like it in all relevant respects: y' is just as good as y, and worse than z and better than x to the same extent that y is. Perhaps y' is just like y except the agent twitches her nose. Vong's rule says that z is now required. In contrast, the beatpath view will continue to say that any option is permissible, and this is surely the right result.[33] A related problem for Vong's view is that it does not make any sense when there are infinitely many available options.

Carlson (1996, p. 159) suggests that we first define a relation of 'weak preference$_n$' for each $n > 0$: translated into my framework, x is weakly preferred$_n$ to y if and only if there is a sequence $y = y_0, y_1, \ldots, y_m = x$, with $m \leq n$, such that it would be permissible to choose each option over the one before. One then finds the smallest n such that at least one option is weakly preferred$_n$ to each other option. If there is only one such option, then it is required (with some possibility of tie-breaking in general). However, in diagram B above, this gives the implausible result that only w is permissible.[34]

A view with similar problems is *Minimax*, also known as the Simpson-Kramer method in voting theory; I have not seen it defended in print in this context, although I long found it attractive. It claims that one ought to choose an option against which there are the least pairwise reasons to do something else. In other words, choose an x that minimizes $\max_{y \in \mathscr{C}} R(y/x)$. This rule satisfies the Condorcet criterion and a limited form of clone-independence. But like Carlson's view, it requires option w in diagram C, and, worse than that, it sometimes selects Condorcet losers, requiring option w in diagram C.

Schwartz (1972), Ross, and Herlitz (2019) all suggest views on which any option in the so-called Schwartz set is permissible. We can say that x is Schwartz-better than y if one can deliberate from y to x, but not from x to y, again in the sense that there is a sequence of options leading from y to x such that one ought to choose each option over the one before. The Schwartz set consists of the Schwartz-best options. Thus the

[33] More generally, the beatpath (in the context of voting theory) satisfies Tideman's clone-independence axiom.

[34] Carlson seems to suggest that his view will render all options in this case permissible, but does not explain why. Carlson's paper is mainly concerned with giving a different solution to the case of Quinn's 'self-torturer'. For those who know the case, I'll mention that, on a natural way of modelling it, the beatpath view would say that any one of the first few settings on the torture device would be permissible. This seems like an acceptable verdict (noting again that I am talking about the non-sequential version of the case), although a common view seems to be that the self-torturer ought to choose some non-trivial amount of pain.

beatpath view is a refinement of Schwartz's view that pays attention to the strengths of pairwise reasons. Abstractly, Schwartz's view has a lot going for it: for example, it is compatible with backwards induction in sequential choice scenarios. However, precisely because of this compatibility, it leads to the kind of non-sequential value pump exemplified by Diagram D: it permits $x\_$, and in the elaboration of the case with a whole spiral of options $x\_, y\_, z\_, x\_\_, \ldots$ it will still permit any option. This is an unacceptable result.

As an alternative to the Schwartz method, Herlitz suggests a principle of 'strongly uncovered choice'. This method does not face the same objection, and my reasons for rejecting it are more tentative. An option x is *strongly covered* if and only if there is an option y that is 'unambiguously better' in the sense that one ought to choose over x anything that one ought to choose over y, and one ought to choose y over anything over which one ought to choose x.[35] The rule is that one may choose any option that is not strongly covered. Note that, like the Schwartz method, strongly uncovered choice determines permissibility directly in terms of the verdicts in pairwise choices, without reference to the strength of reasons behind such choices. So one objection is that, in some cases, the strength of such reasons does seem relevant. In diagram B above, one really ought to avoid y. For a different case, not depending on the strengths of reasons, suppose that among the three options x, y, z, one ought to choose x over y and y over z, but either of x or z would be permissible in a straight choice between them. Strongly uncovered choice invariably permits y. This seems wrong to me, although I admit that intuitions here are fragile and may depend on the substance of the view. Some other uncovered choice rules discussed in Duggan do better in this example, but none take the strengths of reasons into account.

5 Some Concrete Views

Assuming that we can reduce arbitrary choices to uncertainty-free choices (section 3) and arbitrary (including uncertainty-free) choices to pairwise choices (section 4), it only remains to write down an extensionally plausible theory of these uncertainty-free pairwise cases. (Or one could focus on one-person cases, which may be equally enlightening.)

In section 2, I identified two key choice-points: the wide/narrow distinction embodied in Case 8 (and the parallel one-person Case 7) and the soft/hard distinction embodied in Case 12 (and Case 13). Thus there are four types of views corresponding to different ways of resolving these issues. My goal is to give a simple example of each kind, explaining how each extends to general choice scenarios in light of the

[35] Herlitz's formulation contains a typo, as he explained to me in correspondence; at least when degrees of betterness are commensurable, the set of permissible options is supposed to be the McKelvey uncovered set discussed in Duggan.

Supervenience Theorem. In section 6, I examine what they say about extinction risk.

In describing these views, it suffices to give a formula for $R(x/y)$, the strength of the reasons for choosing x over y in a straight choice, for each pair of options. By convention, $R(x/y)$ is positive if one ought to choose x, negative if one ought to choose y, and 0 if either would be permissible. To do this, it is convenient to formalize the ideology of 'offsetting' I introduced in my discussion of Case 12. If r_1, r_2 are numbers (measuring the strength of different considerations relevant to choosing some option x over some alternative y) define

$$r_1 \text{ offset by } r_2 = \begin{cases} r_1 & \text{if both positive or both negative} \\ \max\{0, r_1 + r_2\} & \text{if } r_1 \geq 0 \geq r_2 \\ \min\{0, r_1 + r_2\} & \text{if } r_2 \geq 0 \geq r_1. \end{cases}$$

Thus if r_1 and r_2, being positive, both point in favour of x, or, being negative, both in favour of y, no offsetting occurs; if they point in different directions, then we allow r_2 to weigh against r_1, but not to outweigh it. So, for example, if r_1, being positive, points in favour of x and r_2, being negative, points in favour of y, then r_1 offset by r_2, being at least zero, does not point in favour of y.

5.1 A narrow, soft view

I will start with the type of view that strikes me as the most theoretically natural of the ones I will consider. It is 'narrow' in that it denies the No Difference View and 'soft' in that it allows either option in Case 12. The first of these features will raise some hackles, but, as I suggested in section 2, the fact is that the No Difference View is awkward to explain in the context of the Asymmetry, and, if Boonin is right, need not be defended. Meanwhile, as I argued in section 2.4, the 'soft' verdict is more natural than the 'hard' one if (as I claim) Act 2 is permissible in Case 10.

Suppose we are choosing between two uncertainty-free ourcomes x and y. I suggest attending to the following features of the case. First, there are *necessary* people, those who exist in both x and y. I'll let $T_{\text{nec}}(x)$ and $T_{\text{nec}}(y)$ denote the total welfare of the necessary people in x and y respectively. Second, there are contingent people, the people who exist in only one option or the other. I'll assume that the contingent people in x who have good lives have total welfare $T_{\text{good}}(x)$, and those in x who have bad lives have total welfare $T_{\text{bad}}(x)$; similarly for $T_{\text{good}}(y)$ and $T_{\text{bad}}(y)$.

We can now define the strength of the reasons for choosing x over y to be

$$R(x/y) = [(T_{\text{nec}}(x) - T_{\text{nec}}(y)) + (T_{\text{bad}}(x) - T_{\text{bad}}(y))] \text{ offset by } (T_{\text{good}}(x) - T_{\text{good}}(y)).$$

The first term, in square brackets, reflects in the simplest way possible the balance of considerations about the welfare of necessary people and of contingent people with bad lives. This term is also equal to the difference in the total harm of the two outcomes, and, if we left it at that, we would get the Harm Minimization View for pairwise choices, a view I briefly considered and rejected at the end of section 3. But we also allow these harm-based considerations to be offset by the balance of considerations about the welfare of contingent people with good lives.

I should now state how this view extends to cases of uncertainty in light of the Supervenience Theorem. In each state, each act leads to the existence of some necessary people (those who would also exist, in that state, under the alternative act) and some other contingent people (those who would not exist, in that state, under the alternative). We can thus in general define $T_{\text{nec}}(x)$ to be the expected total welfare of necessary people under act x, and $T_{\text{good}}(x)$ and $T_{\text{bad}}(x)$ to be the expected total welfare of contingent people with good lives or bad lives respectively. The same formula for R is then the appropriate formula for all pairwise choices.

Combining this formula with the beatpath view about intransitivity, we obtain a view that validates the Asymmetry in both the Expectational and the Group-Level forms, and satisfies the Condorcet Criterion and the three supervenience principles. It also agrees with Expected Totalism in fixed-population cases. In fact, something stronger is true: quite generally, any option that maximizes expected total welfare will be undominated, and therefore permitted. This narrow, soft view can then be seen as a permissive cousin of Expected Totalism that incorporates the Asymmetry and denies the No-Difference View.[36] This fits well with the idea I suggested in section 2.2 that expected total welfare might be the right theory of instrumental value, but not all considerations that ground value facts generate reasons with requiring strength.

5.2 A narrow, hard view

My second example will be a theory that again denies the No Difference View but which is 'hard' in the sense of requiring Act 1 in Case 12. Of course, HMV is one such theory, but I take it to give the wrong answer in Case 10.

I this time define

$$\begin{aligned} R(x/y) = &(T_{\text{nec}}(x) - T_{\text{nec}}(y)) \\ &+ \big[(T_{\text{bad}}(x) - T_{\text{bad}}(y)) \text{ offset by } (T_{\text{good}}(x) - T_{\text{good}}(y))\big]. \end{aligned}$$

This is just like the formula of section 5.4, but modified so that now the balance of considerations concerning contingent people with good lives is only allowed to offset the balance of considerations concerning contingent people with bad ones; it does not interact in the same way with harms or benefits to necessary people.

[36] Of course, this point relies on the weight with which $T_{\text{good}}(x) - T_{\text{good}}(y)$ is allowed to offset the other considerations. One could, for example, offset by *one half* of $T_{\text{good}}(x) - T_{\text{good}}(y)$ to obtain a view that is both less permissive and less similar to Expected Totalism. But this variant denies the Expectational and Group-Level Asymmetries and may run into other difficulties as well.

A different version of the view would use $T_{\text{good}}(x)$ to offset $T_{\text{bad}}(x)$ and $T_{\text{good}}(y)$ to offset $T_{\text{bad}}(y)$ separately. To see the significance of this, consider the following case:

Case 18: Mixed Addition 3

Act 1		Act 2	
Eve	5	Eve	—
Cain	-1	Cain	—
Abel	—	Abel	5

As I've defined R, Act 1 is impermissible, because $T_{\text{good}}(\text{Act }1) - T_{\text{good}}(\text{Act }2)$, the difference between Eve's welfare in Act 1 and Abel's in Act 2, is zero, and cannot offset the badness of Cain's life. On the alternative way of offsetting, the goodness of Eve's life would (on its own) offset the badness of Cain's, and either act would be permissible. I think that, even if one denies the No Difference View, the more plausible verdict is that Act 1 is impermissible (and the theory of section 5.1 agrees). But this kind of subtlety has not, as far as I know, been discussed in the literature.

The extension to cases of uncertainty proceeds in the same way as in section 5.1: in general we define T_{nec}, T_{good}, and T_{bad} to be the shares of expected total welfare attributable to the relevant types of people. So in general the governing considerations are (i) differences in the expected total welfare of necessary people, and (ii) differences in the expected total welfare of contingent people with bad lives, offset by differences in the expected total welfare of contingent people with good ones.

5.3 A wide, hard view

I now state two theories that endorse the No Difference View of the non-identity problem, claiming that one ought to choose Act 2 over Act 1 in Case 8. As I explained in my discussion of that case, one way of implementing the No Difference View is to choose a counterpart relation between the individuals who exist in the relevant outcomes, and to use this counterpart relation instead of transworld identity in analysis: starting from the narrow views above, we would reinterpret 'necessary' people as those with counterparts and 'contingent' people as those without. It's natural to insist that the counterpart relation be *saturating* (to use Meacham's term), pairing as many people in x to people in y as possible. Thus if x and y have the same number of people, a saturating counterpart relation will define a bijection between them.

Although I have some qualms about Meacham's particular view about the counterpart relation (principally that it is inconsistent over time), I do no mean to press them here, and someone who likes Meacham's view could certainly combine it with my general framework. I do, however, wish to propose an alternative that seems at least as natural (and doesn't suffer from temporal inconsistency): consider *all* saturating counterpart relations that extend transworld identity, and proceed as if uniformly uncertain which one is correct.

We can get an explicit formula for R by combining this move with either of the previous two 'narrow' views, and I will simply state those formulae without deriving them, as I think they are fairly natural in themselves. We need a little more notation, however. I will suppose (without loss of generality) that outcome x has at least as many people as outcome y. So suppose x has $N_{\text{con}}(x)$ contingent people, with average welfare $W_{\text{con}}(x)$, and so too y has $N_{\text{con}}(y)$ with average $W_{\text{con}}(y)$; the number of *excess* people in x is $N_{\text{con}}(x) - N_{\text{con}}(y)$. Of course these 'excess people' are purely statistical: it is not that some particular contingent people count as excess and others don't. Nonetheless, it is natural to attribute to these excess statistical people total welfare $T_{\text{exc}}(x) = (N_{\text{con}}(x) - N_{\text{con}}(y))W_{\text{con}}(x)$; to the other 'non-excess' contingent people we can attribute total welfare $T_{\neg\text{exc}}(x) = N_{\text{con}}(y)W_{\text{con}}(x)$ and $T_{\neg\text{exc}}(y) = N_{\text{con}}(y)W_{\text{con}}(y)$ in the corresponding outcomes.[37]

We can now define the cardinal choice function R in the following way (still on the supposition that x has at least as many people as y):

$$R(x/y) = (T_{\text{nec}}(x) - T_{\text{nec}}(y)) + (T_{\neg\text{exc}}(x) - T_{\neg\text{exc}}(y)) \\ + T_{\text{exc}}(x) \text{ if } T_{\text{exc}}(x) \text{ is negative.}$$

Thus non-excess contingent people are treated on a par with necessary people, and excess contingent people only make a difference insofar as their total welfare is negative. Although I have not overtly used the idea that additional good lives can offset additional bad ones, leading to the Group-Level Asymmetry, this happens automatically here because we always consider the average welfare of contingent people. On the other hand, this view is still 'hard' in that it requires Act 1 in Case 12.

To extend this theory to handle uncertainty in line with the Supervenience Theorem, we should, as in my discussion of the preceding views, interpret T_{nec} in general as the expected total welfare of the necessary people in x. We should also interpret N_{con} in general as the expected number of contingent people. The number of excess statistical lives is still $N_{\text{con}}(x) - N_{\text{con}}(y)$, supposing that this is positive. W_{con} is not, as one might guess, the expected average welfare of contingent people, but what I will call their *ex ante average welfare*: the expected total welfare of contingent people divided by their expected number. We can then define T_{exc} and $T_{\neg\text{exc}}$ in the same way as before, representing the shares of total expected welfare attributable to excess and non-excess but contingent statistical lives.

The formula for R thus combines the following key considerations:

1. Differences in the expected total welfare of necessary people.

[37] In stating these definitions using *average* welfare, I run into the objection that if there are no contingent people, then their average welfare is undefined. In that case one has instead $T_{\text{exc}}(x) = T_{\text{exc}}(y) = T_{\neg\text{exc}}(x) = T_{\neg\text{exc}}(y) = 0$.

2. Differences in expected total welfare attributable to non-excess contingent people.
3. The badness of excess statistical lives.

5.4 A wide, soft view

Finally, we can modify the previous view to obtain a 'soft' theory if we simply let the welfare of excess contingent lives offset other considerations. Again supposing that x has at least as many people as y (or, under uncertainty, at least as many expected people),

$$\begin{aligned} R(x/y) = (T_{\text{nec}}(x) - T_{\text{nec}}(y)) + (T_{\neg\text{exc}}(x) - T_{\neg\text{exc}}(y)) \\ + T_{\text{exc}}(x) \text{ if } T_{\text{exc}}(x) \text{ is negative} \\ \text{or offset by } T_{\text{exc}}(x) \text{ if } T_{\text{exc}}(x) \text{ is positive.} \end{aligned}$$

The interpretation of this formula in the presence of uncertainty works just as in section 5.3. In other words, we simply add to the list of three considerations a fourth:

4. A potential offset by the *goodness* of excess statistical lives.

With a little thought one can see that this view, like the other soft view of section 5.1, always permits actions that maximize expected total welfare. It can therefore be seen as a modification of Expected Totalism to incorporate the permission claimed in the second clause of the Asymmetry, while still maintaining the No Difference View.

5.5 Choice Points

At least as important as having a few precisely stated views on the table, I hope I've made it clear how one could try to generate alternatives.

First, one might try to find alternatives to my key axioms, Personwise Supervenience, Statewise Supervenience, and Scale Invariance, for handling uncertainty. But I think all of these axioms are plausible for a theory of impartial beneficence, and denying Statewise Supervenience in particular raises the problem of designation-dependence discussed in section 3.2.

Second, I've proposed using the beatpath Condorcet voting method, but one could consider other voting methods (I mentioned Schwartz, Ranked Pairs, and various kinds of uncovered choice as obvious candidates). One could also pursue a completely different strategy for dealing with multi-option choices, though I would be loath to give up the Condorcet Criterion.

Finally, we can argue about pairwise verdicts in uncertainty-free or one-person pairwise comparisons, and how to implement them. Of particular significance here are choices about what to say about non-identity cases and costly creation cases, or about their analogues in one-person cases.

6 Extinction Risk Revisited

I began with some remarks concerning the extinction of humanity that make vivid the stakes involved in population ethics and the importance of developing systematic theories about what we ought to do under uncertainty. By way of conclusion, let me return to a stylised version of the extinction scenario to illustrate where we have ended up.

To construct an interesting but tractable case, I will think of a *great* future as containg some huge number of human lives—say 10^{20}—all of which are well worth living. An *extinct* future will be one with no future human lives at all, and I will also consider a *drab* future, with just as many people as a great one, but on average barely worth living. Consider then three options:

(A) Maintain the status quo — leading to a great future with high probability, or to a drab future or an extinct Future, with equal low probabilities.

(B) Remove extinction risk — replacing the extinct future with a great one, at some cost to present people.

(C) Change trajectory — replacing the drab future with a great one, at the same cost to present people.

To keep the analysis very simple, let me make some further assumptions:

* Assume that options B and C have the same expected total welfare, and the probabilities are such that B and C would be permissible and A would be impermissible by the lights of Expected Totalism.
* Assume that the possible drab future under A and B would contain completely different people from the great future with which C would replace it. (But otherwise the identities of future people are unaffected.)
* Assume that the possible drab futures contain a mix of good and bad lives, and that the expect total welfare of the bad lives exceeds any costs to present people.

For the last point, note that a drab future could contain vastly more bad lives than there are present people, even if they form only a tiny proportion of the total.

What do the views I developed in section 5 say about this choice? The narrow, soft view of section 5.1 regards B and C as permissible, undominated options, since they are permissible by the lights of Expected Totalism. It is also permissible to choose A over B, since A has the advantage with respect to the welfare of necessary people, and never creates contingent people. But because of the assumption about the bad lives in the drab future under A, one ought to choose C over A. So, when all three options are available, beatpath rules A impermissible.

The narrow, hard view described in section 5.2 gives slightly different answers. Characteristically for a hard view, it recognizes strong reasons to choose A over B, since

B involves costs to necessary people. When we focus on a pairwise choice between B or C, either would be permissible. When we focus on a pairwise choice between A or C, the view again requires C over A, because of the bad lives in the drab future. C, then, is the beatpath-best option.

On the wide, hard view of section 5.3, there is strong reason in favour of A over B, again because B involves costs to necessary people. When we look at A and C, there is potentially even stronger reason in favour of C. This is the verdict of Same Expected Number Totalism. Finally, out of B and C, C increases the ex ante average welfare of contingent people, giving strong reason in favour of C. Thus C is the Condorcet winner, and therefore the beatpath-best option.

On the wide, soft view of section 5.4, however, both the reason to choose A over B and the reason to choose C over B are fully offset by the excess good lives in B. But there is still strong reason to choose C over A. When all three options are available, then, beatpath declares that B and C are the permissible options.

What to make of this? The main thing that is new here is that we have the resources to do the analysis properly, using theories that validate the Asymmetry but also take uncertainty fully into account. On the substantive issues, the general lean of these theories against B is unsurprising, given the Asymmetry. Although some of the views permit B, it would be difficult to tweak the case in such a way that any of them would *require* it. More surprising is that all four views reject option A, at least under some reasonable assumptions about how the drab future plays out. (But in section 5.2, I mentioned and tentatively rejected an alternative way in which a narrow, hard view could be implemented; on this alternative, option A is required.) I find this fact particularly surprising in the context of the narrow, soft view, since it is intuitively the most permissive of the four. It still recognizes the great importance of the long-term future.

Appendix. The Supervenience Theorem

To develop the Supervenience Theorem into a formal result, I use a version of Savage's decision-theoretic framework, in which the objects of choice, 'acts', are modelled as functions from 'states' to 'outcomes'. I will make the welfarist assumption that outcomes are adequately described by welfare distributions.

Formally, let a *state space* be a finite set S with a probability measure $\Pr_S$, which I will require to take rational numbers as values.[38] Let $\mathbb{W}$ be a set of welfare levels, including

[38] The rationality assumption is only strictly required for the part of the Supervenience Theorem that deals with uncertainty-free cases. One could replace the assumption with an appropriate continuity axiom. But the basic point is that, if two plausible views agree about rational-probability cases, then they presumably agree in general. A similar point justifies my focus on finite state spaces, although one could adapt most of the discussion to arbitrary probability spaces.

the possibility $\Omega \in \mathbb{W}$ of non-existence. A *population* $\mathbb{I}$ is a finite set (interpreted as a set of possible people); given $\mathbb{I}$, a welfare distribution is any function from $\mathbb{I}$ to $\mathbb{W}$.[39]

Thus, given a state space S and a population $\mathbb{I}$, an *act* x based on S and $\mathbb{I}$ is a function on S whose values are functions from $\mathbb{I}$ to $\mathbb{W}$; $x(s)(i)$ is to be interpreted as the welfare level, including possibly non-existence, attained by person i in state s if x is chosen. An *option set* based on S and $\mathbb{I}$ is a tuple $x = (x_1, \ldots, x_n)$ of such acts. From now on fix an integer $n > 1$, and consider only option sets with n options.

Putting this all together, a *scenario* $(S, \mathbb{I}, x)$ consists of a state space S with its probability measure Pr_S implicit, a population $\mathbb{I}$, and an option set x based on S and $\mathbb{I}$. A *choice function* C is a function mapping each scenario (S, Pr_S, x) to a subset of $\{1, \ldots, n\}$: the interpretation is that x_k is a permissible choice in this scenario if and only if $k \in C(S, \mathrm{Pr}_S, x)$.

Now I turn to formalising the supervenience principles. Say that scenarios $(S, \mathbb{I}, x)$ and $(S', \mathbb{I}, x')$ with the same population are *personwise equivalent* if, for every $i \in \mathbb{I}$ and every n-tuple $(w_1, \ldots, w_n) \in \mathbb{W}^n$, the probability of the world being such that i would get w_1 under the first act, but w_2 under the second, and so on, is the same in each scenario. Formally, the condition is:

$$\begin{aligned}&\mathrm{Pr}_S\{s \in S : \forall k \in \{1, \ldots, n\},\, x_k(s)(i) = w_k\} \\ &\qquad\qquad\qquad = \mathrm{Pr}_{S'}\{s \in S' : \forall k \in \{1, \ldots, n\},\, x'_k(s)(i) = w_k\}.\end{aligned}$$

Similarly, say that scenarios $(S, \mathbb{I}, x)$ and $(S, \mathbb{I}', x')$, here with the same state space, are *statewise equivalent* if for each $s \in S$ and each n-tuple $(w_1, \ldots, w_n) \in \mathbb{W}^n$, the number of people who, in state s, would get w_1 under the first act but w_2 under the second, and so on, is the same in each scenario. Formally, the condition is

$$\begin{aligned}&\#\{i \in \mathbb{I} : \forall k \in \{1, \ldots, n\},\, x_k(s)(i) = w_k\} \\ &\qquad\qquad\qquad = \#\{i \in \mathbb{I}' : \forall k \in \{1, \ldots, n\},\, x'_k(s)(i) = w_k\}.\end{aligned}$$

The first two supervenience principles then read:

Personwise Supervenience. The choice function takes the same value on scenarios that are personwise equivalent.

Statewise Supervenience. The choice function takes the same value on scenarios that are statewise equivalent.

Next, say that $(S, \mathbb{I}, x)$ is a *scaling* of $(S, \mathbb{I}', x')$ if, for some natural number N, there is an N-to-1 function f from $\mathbb{I}$ onto $\mathbb{I}'$ such that $x'(s)(f(i)) = x(s)(i)$ for all $s \in S$ and $i \in \mathbb{I}$. I will say that two scenarios are *related by scaling* if one is a scaling of the other.

[39] The restriction to finite populations is necessary to avoid the well-known problems of infinitary ethics and decision theory.

Scale Invariance. The choice function takes the same value on scenarios that are related by scaling.

Finally, to state the Supervenience Theorem, say that a scenario $(S, \mathbb{I}, x)$ is *one-person* if $\#\mathbb{I} = 1$ and *uncertainty-free* if $\#S = 1$. It is also convenient to say that the scenario is *unanimous* if it is a scaling of a one-person scenario: this is in fact the only case in which we need Scale Invariance. Unanimity is equivalent to the condition that $x_k(s)(i) = x_k(s)(j)$ for all $k \in \{1, \ldots, n\}$, $s \in S$, and $i, j \in \mathbb{I}$.

Theorem 1 (Supervenience Theorem)**.** *Suppose a choice function C satisfies Personwise Supervenience, Statewise Supervenience, and Scale Invariance.*

1. *C is determined by its values on the class of one-person scenarios.*
2. *C is determined by its values on the class of uncertainty-free scenarios.*

Proof. Consider an arbitrary scenario $(S, \mathbb{I}, x)$. The strategy is to construct from it a one-person scenario, as well as an uncertainty-free scenario, on which the choice function must take the same value. Let me begin by stating what these scenarios are.

Let $T = \{(s, i) : s \in S, i \in \mathbb{I}\}$, equipped with the probability measure $\Pr_T\{(s, i)\} = \Pr_S\{s\}/\#\mathbb{I}$. Choose some one-element set $\{*\}$ and define a one-person scenario $(T, \{*\}, y)$ by putting $y_k(s, i)(*) := x_k(s)(i)$. The first claim is, more specifically,

$$C(S, \mathbb{I}, x) = C(T, \{*\}, y).$$

Second, let M be the least common denominator of the rational numbers $\Pr_S\{s\}$, for $s \in S$. Define $\mathbb{J} := \{(s, i, m) : s \in S, i \in \mathbb{I}, m \in \{1, \ldots, M\}\}$. We can define an uncertainty-free scenario $(\{*\}, \mathbb{J}, z)$ by putting $z_k(*)(s, i, m) := x_k(s)(i)$. The second claim is, more specifically,

$$C(S, \mathbb{I}, x) = C(\{*\}, \mathbb{J}, z).$$

Now, to warm up, let me point out two typical ways of applying the supervenience principles. Suppose, first, that we have a function $\tau : S \to \Sigma$, where Σ is the group of permutations of $\mathbb{I}$. Define a new scenario $(S, \mathbb{I}, x')$ by putting $x'_k(s)(i) := x_k(s)(\tau(s)i)$. Then it is easy to see that $(S, \mathbb{I}, x')$ is statewise equivalent to $(S, \mathbb{I}, x)$, and I will say that the former is a *statewise permutation* of the latter with respect to τ.

Second, suppose that we have another state space S' with probability measure $\Pr_{S'}$. Let R be the set of all probability preserving functions $f : S' \to S$, meaning that, for any $s \in S$, $\Pr_{S'} f^{-1}(s) = \Pr_S\{s\}$. Suppose we have a function $r : \mathbb{I} \to R$. Define a new scenario $(S', \mathbb{I}, x')$ by putting $x'_k(s)(i) = x_k(r(i)(s))(i)$, for all $k \in \{1, \ldots, n\}$, $s \in S'$, and $i \in \mathbb{I}$. Then it is easy to see that $(S', \mathbb{I}, x')$ is personwise equivalent to $(S, \mathbb{I}, x)$, and I will say that the former is a *personwise refinement* of the latter with respect to r.

To apply these ideas, let S' be the set of all pairs (s, σ) with $s \in S$ and $\sigma \in \Sigma$. Define a probability measure on S' by the rule that $\Pr_{S'}\{(s, \sigma)\} := \frac{1}{\#\Sigma}\Pr_S\{s\}$, and define a scenario $(S', \mathbb{I}, x')$ by putting $x'_k(s,\sigma)(i) := x_k(s)(i)$. Then $(S', \mathbb{I}, x')$ is a personwise refinement of $(S, \mathbb{I}, x)$ with respect to the function $r(i)(s,\sigma) := s$. So Personwise Supervenience entails $C(S', \mathbb{I}, x') = C(S, \mathbb{I}, x)$.

Next define a scenario $(S', \mathbb{I}, x'')$ by $x''_k(s,\sigma)(i) := x'_k(s,\sigma)(\sigma i) = x_k(s)(\sigma i)$. This time $(S', \mathbb{I}, x'')$ is a statewise permutation of $(S', \mathbb{I}, x')$, with respect to the function $\tau(s,\sigma)(i) = \sigma i$. So Statewise Supervenience entails $C(S', \mathbb{I}, x'') = C(S', \mathbb{I}, x')$.

Next fix some individual $j \in \mathbb{I}$ and define a scenario $(S', \mathbb{I}, x''')$ by putting

$$x'''_k(s,\sigma)(i) := x''_k(s,\sigma)(j) = x_k(s)(\sigma j).$$

For each $i \in \mathbb{I}$ choose some $\sigma_i \in \Sigma$ with $\sigma_i i = j$. Then we have $x''''_k(s,\sigma)(i) = x''_k(s,\sigma\sigma_i)(i)$, so that $(S', \mathbb{I}, x''')$ is a personwise refinement of $(S', \mathbb{I}, x'')$ with respect to the map r defined by $r(i)(s,\sigma) := (s,\sigma\sigma_i)$. So, by Personwise Supervenience, $C(S', \mathbb{I}, x''') = C(S', \mathbb{I}, x'')$.

Now, $(S', \mathbb{I}, x''')$ is a unanimous scenario; specifically, it is a scaling of the one-person scenario $(S', \{*\}, y')$ defined by $y'_k(s,\sigma)(*) := x''''_k(s,\sigma)(j) = x_k(s)(\sigma j)$. Therefore, by Scale Invariance, $C(S', \{*\}, y') = C(S', \mathbb{I}, x''')$. As a final step, note that $(S', \{*\}, y')$ is a personwise refinement of $(T, \{*\}, y)$ with respect to $r(*)(s,\sigma) := (s,\sigma j)$. Applying Personwise Supervenience, and combining the calculations so far, we find $C(S, \mathbb{I}, x) = C(T, \{*\}, y)$, as claimed.

Now for the second construction. The first construction, applied to $(\{*\}, \mathbb{J}, z)$ instead of $(S, \mathbb{I}, x)$, yields $C(\{*\}, \mathbb{J}, z) = C(T', \{*\}, y')$, where $T' := \{(*, s, i, m) : s \in S, i \in \mathbb{I}, m \in \{1, \ldots, N\}\}$ and $y'_k(*, s, i, m)(*) := z_k(*)(s, i, m) = x_k(s)(i)$. The probability measure on T' is given by $\Pr_{T'}\{(*, s, i, m)\} := 1/\#\mathbb{J}$. But $(T', \{*\}, y')$ is a personwise refinement of $(T, \{*\}, y)$ with respect to $r(*)(*, s, i, m) := (s, i)$. So, by Personwise Supervenience, $C(\{*\}, \mathbb{J}, z) = C(T, \{*\}, y) = C(S, \mathbb{I}, x)$, as claimed. □

References

Boonin, D. (2014). *The Non-Identity Problem and the Ethics of Future People*. OUP Oxford, New York, NY.

Bostrom, N. (2003). Astronomical waste: The opportunity cost of delayed technological development. *Utilitas*, 15(3):308–314.

Carlson, E. (1996). Cyclical preferences and rational choice. *Theoria*, 62(12):144–160.

Chalmers, D. J. (2011). The Nature of Epistemic Space. In Egan, A. and Weatherson, B., editors, *Epistemic Modality*, pages 60–107. Oxford University Press.

Cohen, D. (2019). An actualist explanation of the procreation asymmetry. *Utilitas*, page 1–20.

Diamond, P. A. (1967). Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment. *Journal of Political Economy*, 75(5):765–766.

Duggan, J. (2013). Uncovered sets. *Social Choice and Welfare*, 41(3):489–535.

Gert, J. (2003). Requiring and Justifying: Two Dimensions of Normative Strength. *Erkenntnis*, 59(1):5–36.

Hare, C. (2010). Take the sugar. *Analysis*, 70(2):237–247.

Herlitz, A. (2019). Non-transitive Better than Relations and Rational Choice. *Philosophia*.

Kath, R. (2016). *Shortfall Utilitarianism: A Theory for Variable Population Decisions*. PhD thesis, University of Sydney.

Mahtani, A. (2017). The Ex Ante Pareto Principle. *Journal of Philosophy*, 114(6):303–323.

McCarthy, D. (2015). Distributive Equality. *Mind*, 124(496):1045–1109.

McCarthy, D., Thomas, T., and Mikkola, K. (2016). Utilitarianism with and without Expected Utility.

McDermott, M. (1982). Utility and population. *Philosophical Studies*, 42(2):163–177.

McDermott, M. (2019). Harms and objections. *Analysis*, 79(3):436–448.

McMahan, J. (2013). Causing People to Exist and Saving People's Lives. *The Journal of Ethics*, 17(1/2):5–35.

Meacham, C. J. G. (2012). Person-affecting views and saturating counterpart relations. *Philosophical Studies*, 158(2):257–287.

Moss, S. (2015). Time-Slice Epistemology and Action Under Indeterminacy. In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, pages 172–94. Oxford University Press.

Parfit, D. (1984). *Reasons and Persons*. Clarendon Press, Oxford.

Rabinowicz, W. (1995). To have one's cake and eat it, too: Sequential choice and expected-utility violations. *Journal of Philosophy*, 92(11):586.

Roberts, M. A. (2011). An Asymmetry in the Ethics of Procreation. *Philosophy Compass*, 6(11):765–776.

Ross, J. (2015). Rethinking the Person-Affecting Principle. *Journal of Moral Philosophy*, 12(4):428–461.

Schulze, M. (2011). A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method. *Social Choice and Welfare*, 36(2):267–303.

Schwartz, T. (1972). Rationality and the Myth of the Maximum. *Noûs*, 6(2):97–117.

Tideman, T. N. (1987). Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4(3):185–206.

Vong, G. (2018). Making Non-Transitive Betterness Behave. *Journal of Moral Philosophy*, 15(5):495–515.

Weatherson, B. (2000). Decision Making with Imprecise Probabilities. Manuscript.