Pareto Principles in Infinite Ethics

by

Amanda Askell

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Philosophy

New York University

May 2018

Professor Cian Dorr

Abstract

It is possible that the world contains infinitely many agents that have positive and negative levels of well-being. Theories have been developed to ethically rank such worlds based on the well-being levels of the agents in those worlds or other qualitative properties of the worlds in question, such as the distribution of agents across spacetime. In this thesis I argue that such ethical rankings ought to be consistent with the Pareto principle, which says that if two worlds contain the same agents and some agents are better off in the first world than they are in the second and no agents are worse off than they are in the second, then the first world is better than the second. I show that if we accept four axioms – the Pareto principle, transitivity, an axiom stating that populations of worlds can be permuted, and the claim that if the 'at least as good as' relation holds between two worlds then it holds between qualitative duplicates of this world pair – then we must conclude that there is ubiquitous incomparability between infinite worlds. I show that this is true even if the populations of infinite worlds are disjoint or overlapping, and that we cannot use any qualitative properties of world pairs to rank these worlds. Finally, I argue that this incomparability result generates puzzles for both consequentialist and non-consequentialist theories of objective and subjective permissibility.

Contents

List of Figures

Introduction

Infinite ethics is the investigation of the ethical implications of living in a universe that contains infinitely many agents whose lives are of moral worth. Aggregative ethical theories like utilitarianism identify the value of a world with the total utility or wellbeing at that world. In infinite worlds, however, the total utility of a world is often not well-defined. Therefore alternative aggregative principles have been proposed to allow aggregative theorists to rank worlds that contain infinitely many agents. Attempts to formulate such principles bring to light tension between certain fundamental ethical principles in worlds that contain infinitely many agents.

In this thesis, I argue that if we accept four fundamental axioms – Pareto, the claim that the 'at least as good as relation' $\succcurlyeq$ is a qualitative relation, the Permutation Principle, and transitivity, then we must conclude that most infinite worlds are ethically incomparable, meaning that world w_1 is not at least as good as world w_2 and world w_2 is not at least as good as world w_1. We must therefore reject the claim that $\succcurlyeq$ is a complete relation.

In chapter 1 I survey the existing literature in infinite ethics. I begin by arguing that we have reason to be uncertain about whether the world we occupy contains infinitely many agents. I introduce the basic problems in infinite ethics and the solutions to these problems that have been proposed in the literature.

In chapter 2 I consider examples of worlds that contain infinitely many agents. I argue that we should accept agent-based Pareto principles over 'expansionist' principles in infinite worlds. I also defend the view that the at least as good as relation $\succcurlyeq$ is a qualitative, transitive relation. I defend the Permutation Principle, which says that we can permute the populations of world pairs without altering the qualitative properties that hold at and between the pair of worlds in question.

In chapter 3 I demonstrate that the four axioms I defend in chapter 3 entail that many infinite worlds are ethically incomparable. I begin by showing that world pairs with disjoint populations of agents such that infinitely many agents are better off in w_1 than in w_2 and infinitely many agents are better off in w_2 than in w_1 are ethic ally incomparable by a 'four world' argument. I then extend this result to identical population world pairs and overlapping world pairs. I extend these results further by introducing a 'cyclic' argument, which allows us to demonstrate the incomparability of more world pairs.

In chapter 4 I extend the results of the previous chapter by introducing extensions of the Pareto principle. I argue that we can extend Pareto to a further class of world pairs, which extends the results of the previous chapter. I show that attempts to extend Pareto further entail implausible rankings of infinite worlds. I argue that we should accept 'addition' principles in infinite ethics, which further extend the incomparability of the previous chapter.

In chapter 5, I formulate the incomparability results as an impossibility result: we cannot jointly accept Pareto, transitivity, the qualitativeness of $\succcurlyeq$, the Permutation Principle, and completeness. I consider each of the first four axioms in turn and argue that giving up each is highly undesirable. I then consider the implications of accepting ubiquitous incomparability between infinite worlds. I argue that this generates highly troubling puzzles for both consequentialist and non-consequentialist theories of objective and subjective permissibility.

Chapter 1

The Foundations of Infinite Ethics

In this chapter I provide a comprehensive summary of the motives for and problems in infinite ethics. In section 1.1 I point out that there is a distinct possibility that the universe we live in contains infinitely many agents with lives that are of moral worth. I then show in section 1.2 that in infinite worlds aggregative ethical principles are more sensitive to what we take to be the 'basic locations of value'. I survey existing results in section 1.3, showing that it can be difficult to jointly satisfy certain core ethical commitments if generations are treated as basic locations of value. In section 1.4 review the literature on Pareto principles in infinite ethics, which treat agents rather than generations as basic locations of value. I conclude with a review of infinite aggregation principles in section 1.5.

1.1 The Possibility of an Infinite World

Infinite ethics is the investigation of the ethical implications of living in a universe that contains infinitely many agents whose lives are of moral worth: agents that have preferences like we do or are able to feel pleasure and pain like we do. The ethical implications of this possibility are, however, somewhat immaterial if we can be certain that we do not live in such a universe.[1] In this section I will show that we have compelling reasons to believe that the universe does contain infinitely many agents whose lives are of moral worth.

If the universe is infinite and the probability that agents will come into existence in an arbitrarily large finite region of spacetime is positive and non-infinitesimal, then we should expect there to be infinitely many agents in the universe. We do not know if our universe is finite or infinite in size, but our current evidence suggests that the universe may in fact be infinite. In standard cosmology, for example, a curvature of the universe that is zero or negative suggests that the universe is infinite, and recent data from the Wilkinson Microwave Anisotropy Probe (WMAP) and the Planck Collaboration are strong evidence that the curvature of the observable universe is zero.[2] So the standard model of cosmology and our current data are at least consistent with the hypothesis that we live in a universe that is spatially infinite and contains infinitely many agents.

The standard model of cosmology is not the only reason we have to believe that the universe may be infinite, however. Perhaps the most compelling evidence we currently have that the universe is infinite comes from inflationary cosmology.

[1]The implications are important even if we achieve near-certainty but not absolute certainty. For example, as I will show in Chapter 5, many of the problems raised in infinite ethics affect our ethical decision making even if our credence that the universe contains infinitely many agents is extremely low.

[2]By 'standard cosmology' I mean models that use the Friedmann-Lemaâtre-Robertson-Walker (FLRW) metric. On Planck and WMAP data see Adam et al [2] and Bennett et al [33]. The universe can have zero or negative curvature and be finite if it has a non-trivial topology. The 2013 Planck results do not provide support for a non-trivial topology [159] but this remains an open question. See Ellis and Brundrit [77] on the possibility of infinite populations in FLRW cosmology.

Inflationary cosmology was initially formulated as a solution to several important problems facing the standard Big Bang model. One important problem this model faces is the 'horizon problem'. Regions of the early universe are assumed to be extremely homogeneous on the Big Bang model despite the fact that, given the finite speed of light and the finite age of the universe, these regions must have been out of causal contact with one another. We do indeed observe such homogeneity. The microwave background that we observe if we look in one direction is extremely similar to the microwave background that we observe if we look in the opposite direction (it is highly isotropic) and yet the different regions that this background radiation is being from are outside of each other's particle horizons: light from the radiation background in one direction cannot have reached the area we observe in the other direction or vice versa. It is difficult to explain how such homogeneity could occur if the two regions have never causally interacted with one another.[3]

The inflationary model, originally proposed by Guth [93] posits that there was a phase of exponential expansion a very short period after the Big Bang singularity, after which expansion continued at a much slower rate. This expansion occurred because the early universe existed in a false vacuum state: an unstable state with an energy level higher than the 'true minimum'.[4] This false vacuum decays into bubbles of true vacuum that grow at the speed of light.[5] A period of early inflation could explain the homogeneity that we observe, since the areas of the universe that we observe now would have been able to causally interact prior to this period of exponential expansion.

Guth's version of the theory could not, however, explain why the universe transitioned into a phase of less rapid expansion that would result in a post-inflationary universe with the

[3] A further problem with the initial conditions of the Big Bang model is the 'flatness problem'. See Guth [93] and Maudlin [161, p. 40-44] for a discussion of these problems.

[4] A false vacuum can be metastable, meaning that it can remain at its local minimum energy level for a relatively long time, but can decay to a true vacuum by quantum tunneling. See Coleman and De Luccia [58].

[5] Steinhardt and Turok [215, p. 91-93] give an accessible description of this process.

properties we observe: one that is non-empty and non-homogeneous.[6] A solution to this 'graceful exit' problem was proposed by Linde [155] and Albrecht and Steinhardt [4]. On this new theory, there was a slower transition from the false vacuum into a true vacuum. The process of inflation can continue outside of the 'pockets' that have decayed and the inflationary universe can give rise to infinitely many such 'pocket universes' that are spatially infinite (see Aguirre [3]).[7] This has become known as 'eternal inflation theory'.

Infinitely many of these pocket universes will have conditions that are conducive to life, meaning that if the eternal inflation theory is true then there will be infinitely many agents:

> [Inflationary] theory is fantastically successful by normal scientific standards. But models of the universe as a whole which provide a mechanism for such inflation typically feature eternal inflation – a kind of universe in which pockets of ordinary, non-inflating space keep forming, but in such a way that that the inflating portion of space is never completely filled, but keeps expanding and giving rise to new noninflating pockets. In the most plausible such models, there are many different kinds of pockets, only a few of which are hospitable to life. Nevertheless there is plenty of life: in fact there will be infinitely many life-friendly pockets as well as infinitely many life-unfriendly ones, and the life-friendly pockets will typically contain infinitely many agents each.
>
> Dorr and Arntzenius [71, p. 419]

Therefore, if eternal inflation theory is true then we can predict that infinitely many agents exist in the universe, and that there are likely infinitely many agents in our 'life friendly' pocket of the universe. Many of these agents – even those within our own pocket – will be beyond our ability to causally interact with[8] but if the eternal inflation theory is true then

[6] See Linde [154] for a description of the transition from 'old inflation' to 'new inflation'.

[7] Steinhardt [214] showed that inflation need not end, while Vilenkin [238] showed that this generalizes. Linde [156] offers an early account of eternal inflation theory. The advantages of eternal inflation theory are explored in Guth [94].

[8] On causality and the particle horizon, see Ellis and Stoeger [78]. Bostrom [42] discusses views that restrict ethical consideration to regions that we can causally interact with. I will discuss this more in Chapter 5 when I consider subjective principles in infinite ethics. Until then I will be concerned primarily with objective ethical principles that rank entire worlds (i.e. entire universes).

each bubble universe is infinite and so our future light cone – the part of the universe that our current actions can affect – may contain infinitely many agents.[9]

We now have two models that result in an infinite universe. The first is the standard model of cosmology in which the curvature of spacetime is negative or zero. The second is the standard model of cosmology with the addition of eternal cosmological inflation. The possibility of an infinite universe is therefore not a mere flight of fancy, but is consistent with some of the most successful theories in recent cosmology.[10] It is therefore important to explore the ethical implications of living in a universe that contains infinitely many agents.

1.2 Basic Locations of Value

Some ethical theories like utilitarianism are wholly aggregative. According to wholly aggregative ethical theories, the normative status of an action is entirely a function of how much of some basic value that it produces and the goodness of a world is just a function of the value that exists at that world. Some ethical theories are partially but not wholly aggregative. For example, a partially aggregative ethical theory might treat rights as side constraints on action – it is never permissible to act in a way that violate a person's rights – but say that if we are choosing between acts that don't violate any rights then it is better

[9] See sections 4 and 7 of Dorr and Arntzenius [71]. The possibility that the causal ramifications of our actions are infinite produces a much more troubling version of Lenman's [151] 'cluelessness' objection to consequentialism. In Chapter 5 I will discuss how this affects the subjective permissibility of actions.

[10] There are other hypotheses that predict that the universe contains infinitely many agents. We can list a few examples. If the Everettian interpretation of quantum mechanics is true then there may be infintely many agents on different Everettian branches (though see Wallace [241] on the ontology of Everettian branches). In addition, some have hypothesized that if the Everettian theory is true then individual agents may live forever (see p. 5, Tegmark [221]). The 'cyclic' cosmological model allows infinitely many 'big bangs' and 'big crunches' (see Steinhardt and Turok [216]. Modal realism posits infinitely many agents in real worlds that are, from the perspective our world, not actual (on the potential ethical consequences of this, see Heller [108]). If the simulation hypothesis [41] is true, there may be infinitely many simulated universes besides our own or that our own (simulated) universe may be infinite.

to perform the act that produces more of some basic value.[11]

Most wholly or partially aggregative ethical theories share two key components. First, they have a theory about what is of *final value*.[12] According to hedonists, for example, pleasure is of final value and suffering is of final disvalue, while all other things, such as moving aesthetic experiences, are only of value insofar as they generate pleasure or suffering. Second, they have a theory about how to *aggregate* value: how to get from *pro tanto* final values to all-things-considered value.[13] Average utilitarianism, for example, identifies the overall value of a world with the average wellbeing level of its occupants, while total utilitarianism identifies the overall value of a world with the total wellbeing of its occupants.[14]

Wholly or partially aggregative theories face problems in infinite worlds if (i) their theory of final value includes something that there can be an infinite amount of if the world is infinite, such as pleasure and suffering, and (ii) their aggregative theory does not give diminishing weight to final values such that their overall value tends towards an upper and lower bound.[15]

The key problem that such aggregative theories face in infinite worlds is that their aggregations rules are underspecified: even though they produce a precise account of overall value in finite worlds, they do not produce a precise account of overall value in infinite worlds. Consider the total utilitarian rule for comparing possible worlds: *world* w_1 *is strictly better*

[11] Nozick [177] is a well-known advocate of such a view.

[12] Here I use Berker's [36] terminology. Bostrom [42] calls this a 'value rule'.

[13] Berker calls this a theory of all-things-considered value, while Bostrom calls it an aggregation rule.

[14] Aggregative theories also have a theory of *primary evaluands* and a *deontic theory*. A theory of primary evaluands is a theory about what things we should assess the rightness and wrongness of directly, which Shelly Kagan [118] calls *evaluative focal points* of a theory. A deontic theory is a theory about how we get from facts about total value to facts about rightness and wrongness. Bostrom [42] also argues that aggregative theories have a 'domain rule', which specifies what the relevant domain of an aggregative theory is. I believe that we can fold the domain rule into the theory of overall value rule: if something is outside of the domain of evaluation, then it is not of final value.

[15] Average utilitarianism may be a kind of special case here. If the utility of agents is bounded then technically the overall value of a world cannot be unbounded according to an average utilitarian aggregative theory. But if agents and their utilities are not given diminishing weight then, in infinite worlds, the average utility will often be undefined. Given this, I will treat average utilitarianism as a member of the class of aggregative views that face problems in infinite worlds.

than w_2 if and only if the total utility of w_1 is strictly greater the total utility of w_2. This rule is underspecified in infinite worlds because the different ways that we can aggregate utility in finite worlds produce conflicting results in infinite worlds.

For example, consider a finite world that contains five agents that are born on the same day and that each experience 10 'utils' (units of wellbeing) per year before dying painlessly 50 years later. We can find the total utility of this world by adding up the total number of utils being experienced each year – 10 utils times five agents, so 50 utils total – and then multiply this by the total number of years in which agents exist – 50 utils times 50 years, so 2500 utils total. Call this 'time-first' aggregation since we are aggregating across every year that the agents in a given spatial region exist.[16] Alternatively, we could add up the total number of utils that each agent experiences across his or her lifetime – 10 utils for 50 years, so 500 utils total – and then multiply this by the number of agents – so 500 utils times 5 agents, so 2500 utils total. Call this 'agent-first aggregation'. It doesn't matter which of these methods we use in finite worlds because they will always produce the same result and the total utility of the world will always be well-defined, regardless of which method we use. In this case, they both say that the total value of the world is 2500 utils.

As Cain [48] demonstrates, however, in infinite worlds it is possible to increase the utility across spacetime while decreasing the utility experienced by each agent. To show this, we can consider his case, 'The Sphere of Suffering', which it is worth quoting in full:

> Imagine the following situation. We have an infinite universe in which there are infinitely many persons (and there are only finitely many persons in any given finite volume). We imagine that, with respect to some given frame of reference, the spatial locations of these people remain fixed. These beings are immortal and no other living beings exist or will come into existence. I will assume that there are no other utilities to consider beyond those of these people. Both action

[16] Many 'time-first' views seem to assume a Newtonian spacetime but – as Arntzenius [6, p.41-2] points out – temporal or spatiotemporal orderings will be relative to frames of reference in Newtonian worlds.

> a1 and a2 will bring into existence a sphere one foot in diameter which will remain centred on the same point and will grow in diameter one foot per year. Action a1 brings it about that everyone within the sphere suffers disutility at a fixed finite level per unit of time, and anyone outside the sphere has a positive utility, again, at a fixed finite level per unit of time. Action a2 is similar, except that those within the sphere have a positive utility and those outside suffer. Suppose these are the only utilities to be considered. Which course of action is preferable?
>
> Cain [48, 401-2]

If we use the time-first aggregation method described above then we will conclude that the total utility of this world is positively infinite because there are always infinitely many agents in the space outside of the sphere experiencing positive utility and only finitely many agents in the space inside of the sphere experiencing negative utility at any given time. But if we use the 'agent-first' aggregation method then we will conclude that the total utility of the world in which the agent performs a1 is negatively infinite because every agent in this world has a life that contains a finite period of happiness when they are outside the sphere followed by an infinite period of happiness when they are inside the sphere. Since any positively infinite amount of utility is strictly better than any negatively infinite amount of utility, the total utilitarian criterion will say that the world in which the agent performs a1 is better than the world in which the agent performs a2 if we use the time-first aggregation method and it will say that the world the agent performs a2 is better than the world in which the agent performs a1 if we use the agent-first aggregation method.[17]

Broome [45] uses the term 'locations' to refer to the things that good occurs at, such as time, space, or agents. In infinite worlds, aggregative principles must be relative to a specified kind of location, such as agents or times, as we can see in the case above. Let us call this the *basic locations problem*. In the philosophical literature, the main candidates for basic locations of

[17] In Chapter 2 I will show that this kind of conflict still arises even if we assume that agents can only have lives of finite length and with finite utility.

value in infinite worlds are agents and times.[18] In order to avoid taking a stance on what the 'basic locations of value' are, some authors formulate versions of their principles that can be adapted to both agent-first and time-first aggregation.[19]

I am sensitive to the fact that personal identity for the purposes of ethics is not trivial. Many ethicists have argued that properties like psychological continuity,[20] biological continuity,[21] or having a unified narrative,[22] are what matter in ethics even if they are not necessary or sufficient for the identity relation. For example, Parfit [179] defends the view that an agent must bear what Parfit [179, p. 215] calls 'relation R' to her past and future self in order to remain the same agent across time for the purposes of ethics. Relation R holds between an agent at an earlier time and an agent at a later time if there is psychological continuity with 'the right kind of cause' between the two agents. The some agent-based principles are neutral about issues like personal identity but they do assume that the identity relation is ethically important. I return to this issue in Chapter 5, where I will also consider the view that it is subjective experiences and not agents that are basic locations of value.

Running in parallel to the philosophical literature on infinite ethics is the economics literature on intergenerational equity with infinite 'utility streams'. This literature can arguably be traced back to Frank Ramsey's 1928 paper 'A Mathematical Theory of Saving', which discusses the optimal rate of saving if we assume an infinite time horizon [188, p. 554-5]. It was revitalized in the 1960s by Koopmans [126] and Diamond [68]. These economists

[18] Vallentyne and Kagan [233, p. 5] argue that the locations of goodness could be 'points or regions in time, space or spacetime; or they might be people or states of nature'. I haven't included states of nature in my list here because if we change some state of nature like global temperatures (to use Broome's example [45, p. 1]) in an infinite world, the goodness of this change seems to depend on whether it improves agents' lives or increases utility across spacetime.

[19] Vallentyne [232], Kagan and Vallentyne [233], Lauwers and Vallentyne [148] and Arntzenius [6] all offer both time-first and agent-first versions of their principles.

[20] This view originates in Locke [158, II.27] and the relevance of of psychological continuity in ethics is explored more recently by Parfit [179, Ch. 10-11].

[21] This view, sometimes called 'animalism', has been defended by DeGrazia [67], for example.

[22] See Shoemaker [205, 2.3] for an overview of the narrative criterion.

consider how to weigh the interests of future *generations* of people. And so in this literature, generations are often treated as the basic location of value.[23]

What is a generation? In the intergenerational equity literature, we are typically asked to consider a set X of possible infinite utility streams. Each element x of X is an infinite vector $(x_1, x_2, x_3, ...)$ of utilities, where each x_i in the vector denotes the utility level of the generation of agents at time t_i. These generations are assumed to comprise finitely many agents and the utility of each generation is assumed to be finite. So if, in some vector x, there is a generation at the first time t_1 that we are considering with utility 4 and a generation at the second time t_2 with utility 5 then this vector starts $(4, 5, ...)$.

This generations framework is, I believe, highly underspecified. For example, I could find no discussion of whether each infinite vector is composed of the same agents or of entirely different agents, or if each generation in each infinite vector is composed of the same agents or of entirely different agents. As we saw above, however, the choice of interpretation is important here, since treating agents and times as basic locations of value can yield very different results in infinite worlds. For example, suppose that the future will contain the same agents regardless of what action we take now and we have to choose whether to undertake action A or action B, where the outcomes for each generation are as follows:

	g_1	g_2	g_3	g_4	g_5	g_6	$\cdots$	g_n	g_{n+1}	$\cdots$
A	3	3	3	3	3	3	...	3	3	...
B	2	2	2	2	2	2	...	2	2	...

Figure 1: Act outcomes for generations

If we treat generations as the basic locations of value, then it seems clear that action A is better than action B since it produces more utility at every future generation. But suppose

[23] An alternative idealization developed by Weil [242] appeals to 'infinitely-lived agents' rather than overlapping generations of agents.

that if action A is performed then each generation will consist of three agents. If action B is performed then the exact same agents will exist, but at a rate of one agent per generation (to make this more plausible, suppose that all agents are born from frozen fertilized eggs that can be incubated at any time). Since there is only one agent per generation, each agent can use more resources. The utility per agent conditional on A and B is as follows:

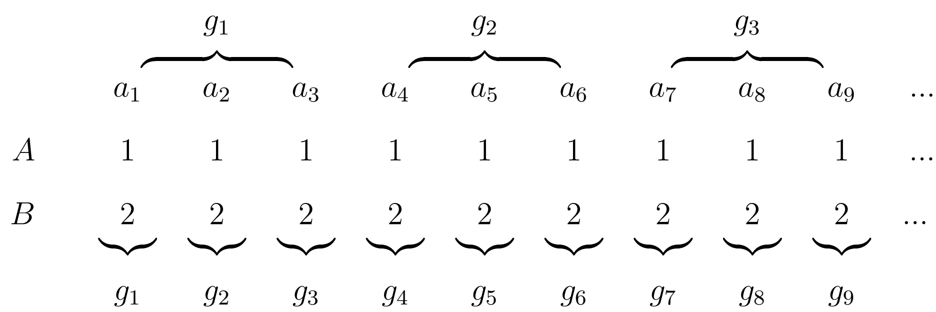


Figure 2: Act outcomes for agents within generations

Therefore the action that produces strictly more utility at each generation produces strictly less utility for every agent who will ever live. This is a strong reason to think that generations, insofar as they have been specified in this literature, are not basic locations of value.[24]

Blackorby et al [38, p. 580] claim that non-standard interpretations of the generations framework are possible. For example, that each x_i could be treated as the utility level of a single agent in a countably infinite population. But as we see above, which interpretation one adopts can reverse how one ranks the same pair of outcomes. Moreover, if we think that each x_i is the utility level of a single agent then it is not clear why we should be allowed to help ourselves to *ordered* utility streams since agents come in no natural order (unless we claim, quite implausibly, that birth order has some kind of moral significance).

[24]The generation theorist might try to avoid this problem by claiming that (i) if the future contains the same people then each generation must contain the same number of agents, or (ii) agents cannot be born at times other than the actual time at which they are born, or (iii) all futures contain entirely different people. None of these seem especially plausible. Further problems for this temporal interpretation arise if we believe that each future time may contain infinitely many agents across space, but I won't discuss these here.

For now I will assume the standard time-based interpretation of generations when discussing the intergenerational equity literature, since most of the problems and solutions in this literature assume that utilities come in a privileged ordering and are therefore inconsistent with an agent-based interpretation of generations. Theories that focus on generations are therefore best understood as time-based theories, rather than as distinct theories about what constitute basic locations of value. The case above shows that the generations-based view inherits the standard problems associated with time-based views.

I have identified two basic locations of value: agents and times, and identified generation-based views as a species of the latter, since generations are just the set of agents that exist during a given temporal period. It is worth pointing out, however, that those who treat agents as the basic locations of value may hold very different views about the nature of agents, or they may believe that subjective experiences and not entire lives are basic locations of value. I return to these issues in the final chapter of this thesis.

In this section I have surveyed agents and times as basic locations of value and demonstrated the ways in which they conflict. Throughout this thesis I will argue that when we are faced with choosing between treating agents or times or anything else as basic locations of value, the claim that agents are the basic locations of value is more plausible than the alternatives.

1.3 Sensitivity, Equity, and Completeness

Once we have established what our basic locations of value are, we need to specify how we are going to aggregate utility across these basic locations of value. In finite worlds this is fairly simple. If we favor purely aggregative principles then we could simply add up the utility across all of our basic locations of value and the result will be sensitive to any increases in utility. If we increase one agent's lifetime utility by 3 utils, then the total utility of the world will increase by 3 utils regardless of whether we use time-first or agent-first aggregation. Alternatively, we could aggregate utility in a way that pays attention to the distribution of utility across basic locations of value. For example, we might weigh utility above above some threshold more than we weigh the utility of agents below that threshold[25] or we might favor distributions that are more equal over those that are less equal.

In infinite worlds, it is not so simple to aggregate utility, let alone to do so in a way that is sensitive to the distribution of utility across agents.[26] In 'A Neglected Family of Aggregation Problems', Krister Segerberg [201] asks us to consider a firm whose sole goal is to maximize profit. It must consider what the perfect policy to employ is, where h is a policy and $h(n)$ is the amount of profit the firm will make in year n. At first, we might think that the value of a policy – $V(h)$ – is just the sum of the profit the policy will make for the firm each year, for all years the policy is implemented: $\sum_{n=0}^{\infty} h(n)$. As Segerberg points out, however, if the firm exists forever and implements the same policy every year, then this sum will not always be well-defined because the series will diverge. Segerberg then shows that the following naïve extension of the summation account of the value of a policy is also inadequate:

> We could perhaps agree to write $v(h) = \infty$ and $V(h) = -\infty$ if $\sum_{n=0}^{\infty} h(n)$ is larger respectively less than every finite bound as i grows towards infinity. The definition thus extended does not quite agree with our intuitions. For example, let $h_0(n) = 1$ and $h_1(n) = 1000000$. Then $V(h_0) = V(h_0) = \infty$, so according to the extended definition h_0 and h_1 would be equally good. But one need not be an ideal board member to see that h_1 is a better policy. [201, p. 224]

As this example shows, if we say that two sequences whose sums diverge to positive infinity are equally good then we will be insensitive to local increases in value at our basic locations of value. An infinitely-long life in which an agent experiences utility +1 every day and an

[25] This aggregation method is consistent with 'critical level utilitarianism' [37].

[26] I will touch on some of the problems with appealing to distributive principles in infintie ethics in Chapters 4 and 5.

infinitely-long life in which an agent experiences utility +20 every day will be deemed equally good because they both have the same cardinality of utility,[27] even though the second has more utility at each basic location – in this case each day of life – than the latter does.[28]

This problem of *insensitivity* has been repeated throughout the history of infinite aggregative ethics. For example, Mark Nelson [169] asks us to consider two actions, J and K. Act J produces 1 util every minute from the time that it is first performed, while act K produces 2 utils every minute from the time that it is performed.[29] It seems better to perform act J since this produces more utility at every basic location of value (in this case, times). But Nelson argues that since these two actions produce the same quantity of utility, the utilitarian would have no basis for choosing between them.[30]

The claim, then, is that an aggregation rule that ranks worlds or outcomes should be sensitive to improvements in utility at basic locations of value, even if these improvements do not affect the cardinality of the utility at the world or outcome as a whole.[31] Segerberg's naïve extension rule and Nelson's 'same quantity of utility' rule are not sensitive to such changes,

[27]Infinity can come in different cardinalities. In this case, if the agent lives for countably-many days then the cardinality of both lives is the same as the cardinality of the natural numbers, $\aleph_0$. If an agent lives for uncountably many days – for example, the days can be put in one-to-one correspondence with the real numbers – then her life will have a strictly greater cardinality of utility. (See Shelah [204] for an overview of cardinal arithmetic.) Even if our aggregative ethical theory is sensitive to different cardinalities of infinity, the problem of insensitivity remains. Many sequences with strictly greater value at each basic location will have the same cardinality of utility. I will discuss larger cardinalities of infinity in Chapter 4.

[28]In this case days of life are basic locations of value because we are assessing the value of a life and not the value of a world. If lives contain infinite utility then I believe any agent-first aggregation principle must be sensitive to improvements within the agent's life. I return to this in the next chapter.

[29]I have added the units of value and length of times. I follow Nelson in focusing on the utility produced by each action, but we could replace 'utility' with 'expected utility' in this case and the point would hold.

[30]The utility of most actions will be different be different at only finitely-many locations. This means that if the future is infinite, the utility of most of the actions available to us will be either positive or negative infinity (depending on whether the future is infinitely good or infinitely bad). If we accept Segerberg's naïve extension, most of the actions available to an agent in an infinite world will be equally good.

[31]It might be suggested that what it is to accept that x is a 'basic location of value' is to accept Sensitivity with respect to x. But there are some that accept that agents are the basic locations of value but reject agent-based Sensitivity principles: for example, those that adopt the Maximin principle (since, according to Maximin, the utilities (1,3) are no better than the utilities (1,1). I do not wish to rule out such views and so I won't equate basic locations of value with Sensitivity commitments.

since improvements that do not affect the cardinality of the utility at the world will not affect the ranking of that world under either rule.

Both Segerberg's rule and Nelsons's rule fail to meet the following desirable condition for aggregation rules, a condition that has appeared in slightly different forms and under several different names in the literature, and that I will call 'Sensitivity'[32]

Sensitivity (Locations)

($\succcurlyeq$) *If w_1 and w_2 share the same basic locations of value and every basic location of value has at least as much utility in w_1 as it does in w_2, then w_1 is at least as good as w_2.*

($\succ$) *If w_1 and w_2 share the same basic locations of value and every basic location of value has at least as much utility at w_1 as it does at w_2 and some basic location of value has strictly greater utility in w_1 than it does in w_2, then w_1 is strictly better than w_2.*

This general Sensitivity principle is relative to whatever our basic locations of value are. Cain's [48] sphere of suffering case shows that we cannot treat both agents and times as basic locations of value in infinite worlds. If the world is infinite then it is not possible to jointly satisfy a sensitivity principle over agents and a sensitivity principle over times. Any aggregation rule that entails Sensitivity for times will say that policy h_1 is better than policy h_0 in Segerberg's infinite firm case, and that act J is better than act K in Nelson's case.[33] Time-first aggregation rules that entail Sensitivity for temporal locations can therefore avoid the problems presented by Segerberg and Nelson. Relativizing Sensitivity to a basic location of value is not sufficient to produce well-defined results in all cases, but attempting to satisfy

[32]This basic idea has been called 'Monotonicity', 'Pareto', 'Sensitivity', or 'Basic Idea'. Vallentyne [230, p.215]; [232, p. 413] argues that this principle preserves the 'spirit' of traditional utilitarianism. Sensitivity also preserves the spirit of other aggregative ethical views, however, and I don't believe its appeal will be limited to traditional utilitarians.

[33]Endorsing Sensitivity is not sufficient to resolve a broader problem with subjective ethical decision making involving infinities that Bostrom [42] calls *infinitarian paralysis*. If the world contains infinite positive utility and infinite negative utility then, according to many aggregative ethical theories, nothing I can do will make a difference to the total utility of the world (assuming that my actions can only produce a finite utility). I return to the problem of subjective ethical decision making in Chapter 5.

a Sensitivity principle for multiple basic locations of value will often produce inconsistencies in infinite worlds even if those Sensitivity principles are consistent in finite worlds.

In the cases that Segerberg and Nelson construct *all* of the basic locations of value (times) and are better in one outcome than they are in the other. It would therefore be sufficient to endorse the 'Weak Sensitivity' principle, which says that if w_1 and w_2 share the same basic locations of value and every basic location of value has strictly greater utility at w_1 as it does at w_2 then w_1 is strictly better than w_2. But it is easy to construct cases that require the full strength of Sensitivity. For example, suppose that act L produces the same utility as act J at t_1 and it produces twice as much utility as J at every time after t_1.[34] Act L clearly seems better than act J, but this is not entailed by Weak Sensitivity. So if we want our aggregative ethical theories to be sufficiently sensitive to local improvements in value in infinite worlds, then it seems that they should at least entail full Sensitivity for whatever they take the basic bearers of value to be.

In the previous section I identified three candidates for 'basic locations of value'. These were times, generations, and agents.[35] Rather than talking about this general Sensitivity principle for locations, we can refer to the Sensitivity principles for each of these basic locations of value. In the previous section I argued that generations are underspecified and that a time-based interpretation of generations seems implicit in the literature on intergenerational equity. Since the intergenerational equity literature focuses primarily on generations, however, it will be helpful to formulate a Sensitivity principle for generations as well as for times and agents. I will assume that if two worlds contain the 'same generations', this means that they have the same temporal structure and that they share exactly the same set of agents at each temporal period. The three Sensitivity principles that result if we substitute basic

[34] Similarly, if an act produces more utility than J at t_1 and the same utility as J at every time after t_1, then this act seems better than J.

[35] Two further plausible candidates include subjective experiences and spatiotemporal regions.

locations with generations, times, and agents are as follows:[36]

Generational Sensitivity

($\succcurlyeq$) *If w_1 and w_2 contain the same generations and every generation has at least as much utility in w_1 as it does in w_2, then w_1 is at least as good as w_2.*

($\succ$) *If w_1 and w_2 contain the same generations and every generation has at least as much utility at w_1 as it does at w_2 and some basic location of value has strictly greater utility in w_1 than it does in w_2, then w_1 is strictly better than w_2.*

Temporal Sensitivity

($\succcurlyeq$) *If w_1 and w_2 have the same temporal structure and every time has at least as much utility in w_1 as it does in w_2, then w_1 is at least as good as w_2.*

($\succ$) *If w_1 and w_2 have the same temporal structure and every time has at least as much utility at w_1 as it does at w_2 and some basic location of value has strictly greater utility in w_1 than it does in w_2, then w_1 is strictly better than w_2.*

Pareto (Agent Sensitivity)

($\succcurlyeq$) *If w_1 and w_2 contain the same agents and every agent has at least as much utility in w_1 as they do in w_2, then w_1 is at least as good as w_2.*

($\succ$) *If w_1 and w_2 contain the same agents and each agent has at least as much utility at w_1 as they do at w_2 and some agent has strictly greater utility in w_1 than they do in w_2, then w_1 is strictly better than w_2.*

In the intergenerational equity literature it is common to refer to Generational Sensitivity as 'Pareto'. I have reserved the term Pareto for the agent-based Sensitivity principle because I will focus on agents as basic locations of value in subsequent chapters.

These sensitivity principles conflict with one another. We can see this when we consider the

[36] Temkin [225, p. 586] develops a similar set of dominance principles to demonstrate their inconsistency.

Sphere of Suffering case from the previous section. The world in which everyone is happy and a sphere of suffering expands throughout the universe is worse than the world in which everyone is unhappy and a sphere of happiness expands throughout the universe by Pareto because all agents have infinitely bad lives in the first world and all agents have infinitely good lives in the second world. The world in which everyone is unhappy and a sphere of happiness expands throughout the universe is worse than the world in which everyone is happy and a sphere of suffering expands throughout the universe by Temporal Sensitivity because at each time there is infinite suffering in the first world and infinite happiness in the second. Many have found Pareto to be more plausible than Temporal Sensitivity in cases where the two principles conflict. For example, Temkin calls Pareto the 'Personal Dominance Principle' and Temporal Sensitivity the 'Temporal Dominance Principle'. In cases where the two conflict, Temkin is inclined to endorse Pareto over Temporal Sensitivity:

> [I]f the same people would exist in each outcome, and they would each be better off in one of the outcomes than the other, then, in accordance with the Personal Dominance Principle, I would regard the outcome in which they were all better off as better than the other outcome, regarding utility, regardless of how the two outcomes compared in accordance with either the Spatial or Temporal Dominance Principles. [225, p. 590]

Although Temkin favors Pareto over Temporal Sensitivity, other ethicists like Vallentyne and Kagan [233] and Arntzenius [6] have formulated principles that are sensitive to improvements across space and time. The conflict between Pareto and sensitivity to improvements in utility across space and time will be explored in more detail in subsequent chapters.

Sensitivity captures the component of aggregative theories that says we make a world or outcome better by adding more utility to it. There is another important component shared by many aggregative theories, however, which we might call *equity*. A theory is thought to be equitable if it does not prefer one distribution over another if the two distributions are identical but differ only in terms of who has a given utility level. For example, if agents are basic locations of value then a theory is inequitable if it prefers a distribution in which agent A has utility 7 and agent B has utility 5 over a distribution in which agent A has utility 5 and agent B has utility 7. If times are the basic locations of value then adding 10 utils to time slice t_1 is just as good as adding 10 utils to any other time slice according to equitable theories. One of the earliest defenders of giving equal weight to utility across times is Sidgwick [208] who states:[37]

> If we are rational, our concern for a moment of our conscious experience won't be affected by the moment's position in time; so when a man is wondering whether to do x, thoughts and feelings that he expects to have later on should be given their due weight, and not discounted because they are off in the future. [208, p. 52]

A related class of principles that capture this component of aggregative ethical theories have appeared under different names in the literature.[38] I will refer to these as 'anonymity' principles. First we can consider the Strong Anonymity principle. Let the 'utility profile' of a world be all of the utility levels that exist at the basic locations of a world (i.e. if a world contains one location at utility 2 and infinitely many locations at utility 1, then this utility profile of this world consists of one 2 and infinitely many 1s). Strong Anonymity says that if we keep the utility profile of a world the same but we change *which* locations have those utility levels, then the resulting world is equally as good as the original:

Strong Anonymity (locations)

If the utility of each basic location of value in w_2 is a permutation of the utility of finitely many or infinitely many basic location of value in w_1, then world w_1 and w_2 are equally good.

[37] Similar sentiments in favor of equity are expressed by Ramsay [188] and Smart [211] regarding times and Bentham [34], Mill [163], and Pigou [182] regarding agents.

[38] These include 'equity', 'anonymity' (this is the near universal term used in the intergenerational anonymity literature), 'impartiality' and 'neutrality' (Vallentyne [232, p. 414-5]).

Suppose the utility levels at $t_1, t_2, t_3, ...$ in world w_1 are $(1, 2, 1, 2, 1, 2, ...)$. And suppose we permute the utility of each basic location of value in w_1 so that the utility levels at $t_1, t_2, t_3, ...$ in world w_2 are $(1, 2, 2, 1, 2, 2, ...)$. Worlds w_1 and w_2 are equally good by Strong Anonymity. It doesn't follow that two worlds are equally good if we *transfer* utility between basic locations of value. For example, if we transfer utility from the 1-util agents to the 2-util agents so that the utility levels at $t_1, t_2, t_3, , ,$ are $(0, 3, 0, 3, 0, 3, ...)$ then the world that results has a distinct utility profile from w_1.[39] Strong Anonymity is intended to capture the equity desideratum. If Strong Anonymity holds then it doesn't matter which particular locations bear the utility levels of a world as long as the profile of utility levels remains the same.

Just like the Sensitivity principle, Strong Anonymity is relative to whatever our basic locations of value are. If our basic locations of value are times, for example, then Strong Anonymity says that it doesn't matter *when* a given utility level is being experienced as long as the utility profile of the world remains the same. If our basic locations of value are people, then Strong Anonymity says that it doesn't matter *who* a given utility level is being experienced by, as long as the utility profile of the world remains the same. We can formulate Strong Anonymity principles for each of the basic locations of value discussed above:

Strong Generational Anonymity

If the utility of each generation in w_2 is a permutation of the utility of finitely many or infinitely many generations in w_1, then world w_1 and w_2 are equally good.

Strong Temporal Anonymity

If the utility of each time in w_2 is a permutation of the utility of finitely many or infinitely many times in w_1, then world w_1 and w_2 are equally good.

[39] We might find 'utility transfer' principles plausible in their own right. I will not discuss transfer principles here, but will return to them in Chapter 4.

Strong Agent Anonymity

If the utility of each agent in w_2 is a permutation of the utility of finitely many or infinitely many agents in w_1, then world w_1 and w_2 are equally good.

In 'Should utilitarians be cautious about an infinite future?' Luc Van Liedekerke [236] presents a problem for those who want to adopt both Sensitivity and Strong Anonymity about the same basic locations of value. He shows that a time-first ranking of worlds w_1 and w_2 cannot satisfy both Temporal Sensitivity and Strong Temporal Anonymity by asking us to consider the following utility streams:[40]

	t_1	t_2	t_3	t_4	t_5	t_6	...	t_n	t_{n+1}	...
u_1	1	0	1	0	1	0	...	1	0	...
u_2	1	1	1	0	1	0	...	1	0	...

Figure 3: An infinite temporal permutation

We can turn the first utility stream u_1 into the second utility stream u_2 by permuting the utility levels of u_1 – we just need to move utility 1 from t_3 to t_2, utility 1 from t_5 to t_3, utility 1 from t_7 to t_5, and so on, and keep all other utility levels at their original locations. Since these two utility streams have the same utility profiles and u_2 can be obtained from u_1 by means of an infinite permutation, if we accept Strong Temporal Anonymity then we must conclude that u_1 and u_2 are equally good. But if we take times to be the basic locations of value then u_2 is strictly better than u_1 by Temporal Sensitivity. This shows that we cannot satisfy both Temporal Sensitivity and Strong Temporal Anonymity in infinite worlds.

The same problem arises if we believe that agents are the basic locations of value. To show this, let us consider a variation of a case given in Hamkins and Montero [99]. Suppose that

[40] Many of the papers in infinite ethics prior to Vallentyne and Kagan [232] discuss time rather than spacetime. I follow the conventions of these papers here. In Chapter 2 I show that similar problems arise if we treat spatiotemporal regions rather than temporal regions as basic locations of value.

w_1 contains an infinite population of agents, where each person is named after an integer: $..., p_{-2}, p_{-1}, p_0, p_1, p_2, ...$ and in world w_1 each agent has a utility level that corresponds with their name, so p_1 has utility 1, p_2 has utility 2, and so on. We can permute the utility levels of the agents of w_1 by moving each utility level i to the agent with the name p_{i-1} to get world w_2. If we do this, the populations of w_1 and w_2 are as follows:

	...	p_{-3}	p_{-2}	p_{-1}	p_0	p_1	p_2	p_3	...
w_1	...	-3	-2	-1	0	1	2	3	...
w_2	...	-2	-1	0	1	2	3	4	...

Figure 4: An infinite permutation of agent utilities

We obtained the second world from the first world by permuting the utility levels of the agents in w_1, and the utility profiles of w_1 and w_2 is the same. This means that w_1 and w_2 are equally good by Strong Agent Anonymity. But w_2 is better than w_1 by Pareto. The conflict between Strong Anonymity principles and Sensitivity principles over the same basic locations of value therefore arises regardless of our choice of basic locations of value.[41]

If we want to retain at least one of these axioms for some basic location of value then we can either weaken our Sensitivity axiom and retain our Strong Anonymity axiom, or we can retain our Sensitivity axiom and weaken our Strong Anonymity axiom. Sensitivity axioms are generally taken to be the more plausible of the two. Theorists may deny Sensitivity for locations of value that they do not take to be basic, since we have seen that it is not always possible to satisfy Sensitivity across multiple candidates for basic locations of value, such as time and agents. But it is difficult to accept that we could add utility to the locations of value that we consider to be basic and decrease utility at none of these locations without thereby making the world better. For this reason, most theorists have concluded that since

[41] Hamkins and Montero [99] argue that we should reject Sensitivity because it is inconsistent with the 'Isomorphism Principle'. I will discuss this problem, Hamkins and Montero's isomorphism principle, and their objections to Sensitivity more in Chapter 2.

Strong Anonymity conflicts with Sensitivity, we must reject Strong Anonymity.[42]

Yew-Kwang Ng [172] argues that aggregative ethicists should retain their Sensitivity principle but adopt a principle weaker than Strong Anonymity: a principle that I will call Finite Anonymity.[43] This principle says that two worlds are equally good if we swap the utility levels of *finitely* many basic locations of value, but worlds may not be equally good if we swap the utility levels of infinitely many basic locations of value:[44]

Finite Anonymity (locations)

If the utility of each basic location of value in w_2 is a permutation of the utility of finitely many basic locations of value in w_1, then world w_1 and w_2 are equally good.

For any given basic location of value, it is always possible to jointly satisfy Sensitivity and Finite Anonymity, as Vallentyne (p. 415-6, 1995) shows. To see why, suppose we permute only finitely many utilities in a world. If one location has lower utility than it did prior to the permutation then, since this is a finite permutation, another location must have more utility than it did prior to the permutation. Therefore neither world can be better by Sensitivity. And so we cannot generate the kind of examples given by Van Liedekerke and Hamkins and Montero if we reject Strong Anonymity and accept Finite Anonymity instead.

In 'Infinite utility: Insisting on strong monotonicity', Luc Lauwers [136] argues that Finite Anonymity is too weak to guarantee that the utilities of each basic location of value will be given equal weight in infinite worlds. In defense of this, he asks us ([136, p. 224]) to

[42] There are exceptions, however. Nelson [169], Garcia and Nelson [91], Hamkins and Montero [97] [99] argue that Sensitivity is false because adding more value at locations does not increase overall utility at a world and (in the case of the latter) because it is inconsistent with the view that worlds are not as good as their 'isomorphic copies'. I discuss these objections in chapters 2.

[43] Svensson [219] calls this principle 'finite equitableness', but it is generally called 'finite anonymity' or just 'anonymity' elsewhere in the literature.

[44] Ng [172, p. 411] argues in favor of this principle on the grounds that 'infinity times zero need not equal zero' in cases where $x = \frac{y}{c}$ where c is a positive constant and so the value of x approaches zero as y approaches infinity. But even if we think that this is a plausible reason to reject *some* applications of Strong Anonymity, but in Van Liedkerke's example no such dynamic is present.

consider a contrived aggregation rule that says the value of a utility stream is the limit of its even-indexed utilities. He then asks us to consider the following two utility streams:

u_1: 0, 1, 0, 1, 0, 1, ... , 0, 1, ...

u_2: 5, 1, 5, 1, 5, 1, ... , 5, 1, ...

Figure 5: Difference at even-indexed and odd-indexed utilities

The contrived aggregation rule will say that the value of u_1 and u_2 are the same: they both have value 1. Even though this contrived rule completely ignores the odd-indexed utilities (and therefore violates Sensitivity with respect to whatever our basic locations of value are in the two utility streams above), the rule satisfies Finite Anonymity because finite permutations of this sequence will not change the limit of its even-indexed utilities.

Lauwers contends that a principle that ignores infinitely many locations entirely surely cannot be equitable. In light of this, Lauwers proposes an alternative 'Fixed-Step Anonymity' principle. In order to formulate this principle, we need the concept of a 'fixed step permutation' of a utility stream. Suppose that we cut an infinite utility stream into segments of equal finite length.[45] If there is some length of segment such that we can transform utility stream u_1 into utility stream u_2 by permuting only the utilities within each of these finite-length segments, then we say that u_2 is a fixed step permutation of u_1. With this concept of a fixed step permutation, we can formulate a 'Fixed-Step Anonymity' principle that is stronger than Finite Anonymity but weaker than Strong Anonymity:

[45] It is important that the segments are of equal length. If they are not then the permutation is a 'variable step' permutation. The Variable Step Anonymity principle conflicts with Sensitivity (Fleurbaey and Michel [87]) but not the $\succ$ condition of Sensitivity (Sakai [195]).

Fixed-Step Anonymity (locations)

If w_1 and w_2 share the same basic locations of value and the utilities of the basic location of value of w_2 are a fixed step permutation of the utilities of the basic locations of value of w_1, then world w_1 and w_2 are equally good.

In the case above, the aggregation rule does not satisfy Fixed Step Anonymity because we could cut u_1 and u_2 into segments of length 2 and switch all of the even-indexed and odd-indexed utilities in u_1 and u_2. If we did this, then u_2 would be better than u_1 by our contrived aggregation rule, which violates Fixed-Step Anonymity. Unlike Strong Anonymity, Fixed-Step Anonymity is consistent with Sensitivity over the same basic locations of value.[46]

Since Fixed Step Anonymity entails Finite Anonymity, however, any results that appeal to Finite Anonymity will also affect theories that satisfy Fixed Step Anonymity. Within the literature on intergenerational equity, several important negative results have been established for aggregative ethical theories that satisfy both Sensitivity and Finite Anonymity over the same basic locations of value.[47] In order to appreciate these results, we must first introduce a third desideratum to our list, namely *Completeness*:

Completeness of $\succcurlyeq$

For all worlds w_1, w_2, either $w_1 \succcurlyeq w_2$ or $w_2 \succcurlyeq w_1$

If the 'at least as good as' relation is complete then there are no pairs of worlds that are incomparable, where being incomparable means that neither world is better than, worse than, nor equal to the other.[48] Completeness seems like a desirable property for the 'at least

[46] The cases that demonstrate a conflict between Sensitivity and Strong Anonymity require permutations over segments that are infinitely long. Van Liedkerke and Lauwers [237] propose a stronger notion of anonymity, 'Bounded Anonymity', that is not consistent with Sensitivity. I will discuss anonymity principles in more detail in Chapter 3, where I will also argue against the fixed-step anonymity principle.

[47] I cannot summarize all of the work in this field here. For an excellent overview, see Asheim [8].

[48] As Chang [53] points out, it is possible for the 'at least as good as' relation to not hold between two worlds and for this worlds to still ethically comparable if we accept some additional ethical relation. However, I will focus on whether the 'at least as good as' relation holds between worlds.

as good as' relation to have. If an ethical relation is incomplete then there may sometimes be no fact of the matter about which of two outcomes is better and, as I will argue at the end of this chapter, we currently have no satisfactory framework for making ethical decisions in which the outcomes of our actions are incomparable.

If our ethical relation is complete then we are able to rank all worlds: to put them in an order from best to worst. It would be even better to know now *how much* better or worse each world is than every other world.[49] The intergenerational equity literature distinguishes between these. In this literature a 'social welfare relation' is a relation that compares utility streams across generations and that satisfies transitivity and reflexivity.[50] A 'social welfare ordering' is a social welfare relation that is also complete. And a 'social welfare function' is an order-preserving function from the set of utility streams to the real numbers.[51] Social welfare functions can therefore tell us how much better one utility stream is than another.

If we accept that infinite worlds can be incomparable but continue to hold that the $\succcurlyeq$ relation is transitive, reflexive and anti-symmetric, then the the $\succcurlyeq$ relation is a weak partial order rather than a non-strict total order. And if we continue to hold that the $\succ$ relation is transitive and irreflexive then the $\succ$ relation is a strict partial order rather than a strict total order. We can represent partial orders of worlds diagrammatically, where $w_1 \longrightarrow w_2$ indicates that $w_1 \succ w_2$ and $w_1 \longleftrightarrow w_2$ indicates that $w_1 \sim w_2$ and a lack of connector between w_1 and w_2 indicates that w_1 and w_2 are incomparable. Consider:

[49] In other words, we would prefer for the value of worlds to be measurable on an interval scale rather than just an ordinal scale. If we are able to say how much better outcome O_1 is than outcome O_2, for example, then we are also able to say how much an agent should value a gamble that involves probability n of O_1 and probability $1 - n$ of O_2. Being able to measure worlds, actions, and outcomes on an interval scale is therefore particularly useful for ethical decision making.

[50] I will discuss these axioms in more detail in Chapter 3.

[51] These definitions are given by Asheim [8, p. 203].

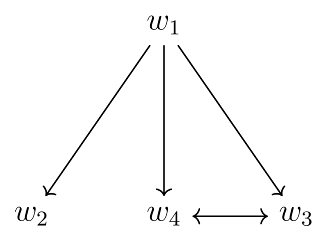


Figure 6: A partial order of worlds

In this example $\langle w_2, w_3 \rangle$ and $\langle w_2, w_4 \rangle$ are incomparable outcome pairs. Note that there is no incomparability if two outcomes are comparable by transitivity. For example, claiming that w_1 and w_4 are incomparable would be inconsistent with transitivity since $w_1 \succ w_3$ and $w_3 \sim w_4$. If any two worlds are incomparable then transitivity entails the following:[52]

(i) If $w_1 \not\gtrless w_2$ and $w_2 \succ w_3$ then $w_3 \not\succcurlyeq w_1$

(ii) If $w_1 \not\gtrless w_2$ and $w_2 \sim w_3$ then $w_1 \not\gtrless w_3$

We can see that the example of a partial order above is consistent with both (i) and (ii).

In 'The evaluation of infinite utility streams', Diamond [68] expands on an earlier result by Koopmans [126][53] by showing that there is no complete social welfare ordering that satisfies Sensitivity, Finite Anonymity, and continuity with respect to the same basic locations of value. Since the literature on intergenerational equity focuses on generations as basic locations of value, I will use Sensitivity and Finite Anonymity to refer to Generational Sensitivity and Finite Generational Anonymity in the remainder of this section. The third axiom in Diamond's result – continuity – says that we can alter a utility stream by an arbitrarily small amount without this giving rise to a large difference in the ranking of that utility stream.[54] If

[52] To show (i), suppose that $w_1 \not\gtrless w_2$, $w_2 \succ w_3$ and $w_3 \succcurlyeq w_1$. It follows that $w_2 \succ w_3 \succcurlyeq w_1$ and so $w_2 \succcurlyeq w_1$ by transitivity, contrary to our assumption that $w_1 \not\gtrless w_2$. To show (ii), suppose that $w_1 \not\gtrless w_2$ and $w_2 \sim w_3$ and $w_1 \succcurlyeq w_3$ or $w_3 \succcurlyeq w_1$. Then either $w_1 \succcurlyeq w_3 \sim w_2$ and so $w_1 \succcurlyeq w_2$ by transitivity, contrary to our assumption that $w_1 \not\gtrless w_2$, or $w_2 \sim w_3 \succcurlyeq w_1$ and so $w_2 \succcurlyeq w_1$, contrary to our assumption that $w_1 \not\gtrless w_2$.

[53] Koopmans shows that ethical relations cannot satisfy Sensitivity, continuity, non-complementarity (if we change all but a single period then we can ignore the periods we do not change) and stationarity (moving a positive change forward or backward should not make a difference) without violating Finite Anonymity.

[54] As Koopmans [126, p. 289] notes, in infinite spaces the concept of continuity depends on what distance

a social ordering does not satisfy continuity then we cannot find a numerically representable social welfare ordering – i.e. a social welfare function – that satisfies Sensitivity and Finite Anonymity. Basu and Mitra [21] extend this result by showing that there is no social welfare function that satisfies Sensitivity and Finite Anonymity even if we don't assume continuity. In other words, if a social welfare relation satisfies Sensitivity and Finite Anonymity then it is not numerically representable.

Svensson [219] shows that the concepts of Sensitivity and Finite Anonymity do not fundamentally conflict with one another, however. He proves that there are finer topologies than those appealed to in Diamond [68] that admit a social welfare ordering that satisfies Sensitivity, Finite Anonymity, and Continuity. These orderings are extensions of the 'Suppes-Sen grading principle'.[55] The Suppes-Sen grading principle says that two utility streams u_1 and u_2 are equally good if one can be obtained from the other by a finite permutation, and that stream u_1 is strictly better than stream u_2 if there is a finite permutation of u_1 that is strictly better than u_2 by Sensitivity.[56] Svensson therefore establishes that the Diamond's result only holds if we assume a more coarse-grained topology.

Svensson's possibility result does not do much to offset the restrictions on social welfare relations imposed by the Diamond-Basu-Mitra result, however. First, the topologies that Svensson appeals to have been the subject of powerful critiques.[57] Second, Svensson's proof requires the Axiom of Choice, and it has been argued that the the Axiom of Choice and non-

measure one selects. Stronger distance measures result in larger topologies, since the topological space is more fine grained, and it is easier for continuity to be satisfied in larger topologies. See Heal, Ed. (1997).

[55] This principle originates in Sen [203] and Suppes [218].

[56] The Suppes-Sen grading principle is the least demanding social welfare relation that satisfies Sensitivity and Finite Anonymity (Asheim, Buchholz and Tungodden [10]). In his paper, Svensson shows that there exists an extension of the 'Overtaking Rule' (due to von Weizsäcker [239]) that satisfies Sensitivity, Finite Anonymity and continuity. Neither principle is complete on its own.

[57] Svensson first appeals to the discrete topology and then to a topology that is weaker than the discrete topology but stronger than the sup topology that Diamond uses. As Campbell [49] points out, continuity is easier to satisfy as one's topology expands but larger topologies are less useful because they contain fewer compact sets. Lauwers [134] argues against the Svenssson topology for infinite utility streams and provides a comprehensive overview of these issues.

constructive proofs should have no place in ethics. The main objections are usually twofold. First, non-constructive objects like nonprincipal ultrafilters introduce an unacceptable level of arbitrariness into ethics, as I will show in the next section.[58] Second, non-constructive existence proofs demonstrate the existence of an object without offering an explicit description of that object. As a quote attributed to Weyl states, non-constructive existence proofs 'inform the world that a treasure exists without disclosing its location'. Objects that lack any explicit description can be of little practical use within ethics and economics.[59]

It is worth emphasizing that the Diamond and Basu-Mitra results show that there is no social welfare *function* that can satisfy Pareto and Finite Anonymity, but as others have noted, it may be sufficient to have a social welfare ordering – a reflexive, transitive, and complete relation – to rank infinite utility streams even if this ranking cannot be represented by a real-valued function.[60] Numerical representability may not be possible given the Diamond and Basu-Mitra results. But even in finite cases social welfare ordering can sometimes be complete without being numerically representable. For example, as Debreu [66, p. 164] points out, the lexicographic ordering is complete but not numerically representable.

Although it seems that, for the purposes of ethics, a social welfare ordering over infinite worlds is sufficient. Lauwers [143] and Zame [247] proved a conjecture first articulated by Fleurbaey and Michel [87], namely that we cannot offer a constructive definition of a social welfare relation that satisfies Sensitivity, Finite Anonymity and completeness: i.e. a social welfare ordering that satisfies Sensitivity and Finite Anonymity. They show that the existence of such a social welfare ordering requires a non-constructive axiom like the Axiom of Choice. Therefore, although we can prove that there exists a social welfare ordering that satisfies Sensitivity, Finite Anonymity, and completeness using non-constructive axioms, we

[58] The arbitrariness of free ultrafilters is also discussed by Lauwers and Liedkerke [149, p. 223-6].

[59] This is quoted in Kline [125, p. 239] and Lauwers [142]. See Weyl [243] on the topic of non-constructive mathematics.

[60] For example, this point is made by Dutta [75, p. 137].

can never offer an explicit description of such orderings.[61] Lauwers concludes that we must reject Sensitivity or Finite Anonymity or completeness.

A final important set of results are the Zame-Lauwers results. Zame shows that for any social welfare ordering that satisfies Finite Generational Anonymity, the set of infinite utility streams that are not strictly better than or strictly worse than one another has outer measure 1. Zame and Lauwers offer independent proofs that no social welfare ordering that satisfies Finite Generational Anonymity and Generational Sensitivity is explicitly describable. I will return to the Zame-Lauwers results in the next section.

Sensitivity, Equity, and completeness are desirable features for any ethical relation to have, and important tensions between these three desiderata have already been highlighted by some philosophers and economists. However, I believe that many of the results discussed above are limited by the impoverished models employed in this literature. In Chapter 2 I will argue that we are better able to capture the equity desideratum by combining Pareto with an axiom stating that haecceitistic facts should have no bearing on the ethical ranking of worlds, rather than with anonymity principles. Many of the important theorems, such as the Basu-Mitra theorem and the Zame-Lauwers theorem assume that agents have utilities within the $[0, 1]$ interval and that there are infinitely many agents with utilities at each end of the interval. In Chapter 3 I will show that we can generalize incomparability results to worlds in which agents can have utility levels outside of the $[0, 1]$ interval. Many of these theorems also assume that ethical rankings cannot be grounded in facts about the world

[61] In a related result, Chichilnisky [55] shows that a social aggregation rule cannot Continuity, Anonymity and Unanimity (which says that if everyone prefers p over q then the social choice rule prefers p over q) unless the preference space is contractible. A rule that satisfies these axioms is sometimes referred to as a Chichilnisky rule. Candeal et al. [51] and Efimov and Koshevoy [76] extend the Chichilnisky approach to cases that involve infinite populations. However, Lauwers [132] shows that infinite Chinchilnisky rules do not exist in any Hausdorff space of preferences or in any Tychonoff (product) topologies of preferences. Lauwers [132] does show that Chichilnisky rules exist in some box topologies, but box topologies are arguably less appropriate topologies of preferences than the coarser product topology since the compactness of Cartesian products whose components are compact is guaranteed in the product topology and not the box topology, i.e. the preference space is not always contractible.

other than utility levels at basic locations of value. I will show that my incomparability results hold regardless of what qualitative features we use to compare worlds.

Further problems with the intergenerational equity literature arise primarily because they assume that 'generations' are basic locations of value even though, as I have shown, the temporal interpretation of generations conflicts with Pareto. The assumption that basic locations of value come in a natural order has led to a focus on order-dependent rather than order-invariant principles for ranking infinite worlds. And whether the agents of different possible futures are identical is often not specified.[62] In order to avoid these problems we could simply adopt Pareto rather than Generational Sensitivity. If people are basic locations of value then we will naturally focus on order-invariant principles because people, unlike times, do not seem to come in any natural order. Since Pareto only applies if the agents in both worlds are identical it will also be important to specify whether populations contain the same people or not. There is not a large literature focusing specifically on Pareto rather than Generational Sensitivity, but in the next section I will offer a survey of the existing literature on the Pareto principle in infinite ethics.

The problems I discuss in this thesis arise for orderings of infinite worlds. Since an ordering is just a transitive and complete relation, it has fewer constraints a social welfare functions, which must be numerically representable. As we have seen, the Diamond-Basu-Mitra results present problems for the claim that a social welfare function can satisfy Generational Sensitivity and Finite Generational Anonymity. Since I will be focused on the difficulties of trying to construct an ordering of infinite worlds, these results are less relevant than the Zame-Lauwers results, which present problems for the claim that a social welfare ordering can satisfy Generational Sensitivity and Finite Generational Anonymity. The Diamon-Basu-Mitra results present problems for social welfare functions that satisfy Sensitivity and Finite

[62]Tungodden and Vallentyne [229] consider anonymous Pareto principles for variable population cases but they assume that populations are finite and are focused on egalitarian, person-affecting views.

Anonymity while the Zame-Lauwers results present problems for social welfare orderings that satisfy Sensitivity and Finite Anonymity. I will therefore explore the Zame-Lauwers results for orderings that satisfy Pareto rather than Generational Sensitivity.

1.4 Pareto Principles in Infinite Ethics

I introduced the Pareto principle in the previous section. Pareto is the Sensitivity principle that is endorsed by those who take agents to be basic locations of value. It says that if w_1 and w_2 contain the same agents and every agent has at least as much utility in w_1 as they do in w_1 then w_1 is at least as good as w_2, and if it's also the case that some agents have strictly greater utility in w_1 than they do in w_2 then w_1 is better than w_2.

There are many principles that are similar to Pareto whose generational equivalents have been explored in the intergenerational equity. Let us assume that x and y are agents and that the two worlds being ranked have identical populations. In the intergenerational equity literature the $\succ$ condition of the Pareto axiom formulated in this chapter is called 'Strong Pareto' the following axioms have also been identified (here interpreted as principles about agents and worlds that contain the same populations):[63]

Weak Pareto *If w_1 and w_2 contain the same agents and each agent in w_1 has strictly greater utility in w_1 than they do in w_2, then w_1 is strictly better than w_2.*

Monotonicity *If w_1 and w_2 contain the same agents and each agent in w_1 has at least as much utility in w_1 than they do in w_2 and some agent has strictly greater utility in w_1 than they do in w_2, then w_1 is at least as good as w_2.*

These weaker principles have generally been employed to prove stronger results not as alter-

[63]These are modeled on principles of the same name in Asheim [8, p. 203-4, 211] but I have removed order-dependent elements from the principles and formulated them as agent-based principles.

natives to Strong Pareto.[64] As I noted above, most of those working in the intergenerational equity literature have taken Strong Pareto to be a more fundamental axiom than anonymity principles, and have scaled back anonymity principles to Finite and Fixed Step Anonymity in response to the inconsistencies between Strong Anonymity and Pareto. This is also true in the literature on infinite ethics within philosophy. In response a conflict between 'temporal neutrality' (akin to Strong Anonymity) and Paretian principles presented by Van Liedekerke and Lauwers [149], Ng [172] notes that Paretian principles are 'so compelling that one is likely to be inclined to reject the time neutrality that necessitates their rejection.' and Vallentyne (p. 414, 1995) describes Pareto as a 'core commitment of utilitarianism' that, in infinite worlds, captures the spirit of utilitarianism better than the sum total rule.[65]

Within the literature on infinite ethics in philosophy, however, some have started to question the plausibility of the Pareto principle in infinite worlds. The most common challenges are based on the inconsistency of Pareto with other principles such as Strong Agent Anonymity. Some of those who question Pareto do so in a rather lukewarm way, acknowledging that it will strike many as a fundamental principle of ethics. For example, after noting that the theory that he puts forward (which I will discuss in the next chapter) is inconsistent with a Paretian principle – the 'weak people criterion' – Arntzenius says the following:

> I expect that most people will think that the verdict of the weak people criterion is better on the grounds that ethics is fundamentally concerned with people, or at least, with utility bearing living creatures, not with locations. If one can make some people happier and none less happy, who cares if some regions are less happy while the others are just as happy. I myself have no strong views on this matter, and I would prefer the view that it is indeterminate which act is better when the weak people criterion yields a different verdict from the weak location criterion. But I am happy to leave it to readers to make up their own

[64] See, for example, Dubey and Mitra's [74] on Monotonicity and Strong Anonymity.

[65] This is in contrast with Nelson and Garcia [91, p. 188] who prefer to treat the sum total in infinite cases as equal, even though this conflicts with the Pareto prinicple.

> minds, especially since I doubt that we are likely to be faced with a situation in which it matters which of the two criteria we use [6, p. 56]

In the next chapter I will show that the claim that we if the Pareto principle does not say that one world is better than another, we can employ a principle that sometimes conflicts with Pareto to compare those worlds, we will violate a plausible axiom about the 'at least as good as relation'.[66] We can see that Arntznius is disinclined to reject Pareto outright.

Others have been more willing to at least weaken the Pareto axiom on the basis of its conflict with other principles in infinite ethics. For example, Van Liedkerke and Lauwers [237] argue that we should replace Pareto with a principle they call 'Infinite Sensitivity'. Infinite Sensitivity says that w_1 is strictly better than w_2 if all agents have at least as much utility in w_1 as they do in w_2 and there is a group of agents with positive measure that has strictly more utility in w_1 than they do in w_2. In finite worlds, any group composed of a positive number of agents will have positive measure. Therefore the verdicts of Infinite Sensitivity coincide with the verdicts of Pareto in finite worlds. In infinite worlds, however, only infinite groups of agents will have positive measure. So in infinite worlds Infinite Sensitivity does not entail that w_1 is strictly better than w_2 if all agents have at least as much utility in w_1 as they do in w_2 and finitely many agents have strictly more utility in w_1 than they do in w_2. Van Liedkerke and Lauwers offer two main arguments for accepting Infinite Sensitivity over Pareto:

> Ultimately, we do not need strong sensitivity because in this article we have set ourselves the task of comparing tails of utility streams, or to put it more bluntly, of comparing infinite groups in a utilitarian manner. Consequently the strong form of sensitivity is no longer a necessity. [237, p. 170]

> First, in an infinite world strong Pareto is an excessive demand. Loosely speaking it demands from us that, looking upon an infinite crowd, we still see the

[66]This axiom – the 'Qualitativeness of $\succcurlyeq$' – will be discussed in more detail in the next chapter.

> face of every single person - a difficult task indeed. Secondly, because it is such a strong demand, accepting it leads to a slippery slope. When you start from strong Pareto you might get forced into the discounting rule. And, on the normative level, discounting is very unattractive indeed. [237, p. 173]

Lauwers argues that it is not necessary to satisfy Pareto when comparing infinite groups of agents. But even when comparing infinite groups of agents we should surely reject principles that entail that are indifferent to (potentially very large) increases in utility across a (potentially very large) finite set of agents. If this is the case then we will be inclined to think that Infinite Sensitivity is an inadequate alternative to Pareto in infinite worlds.

Lauwers argues that we might be forced into the discounting rule because, as argued in Lauwers [133], 'time neutrality' can be characterized by continuity, and linearity, and a bounded anonymity principle. Therefore if we accept Pareto, which is inconsistent with the bounded anonymity principle, discounting is the only decision rule available to us. The strength of these arguments against Pareto depends on the relative plausibility of these alternative principles. In the next chapter, I will argue for an alternative conception of equity than is offered by Lauwers' bounded neutrality principle. If we wish to reject temporal discounting we could, of course, reject the continuity or linearity axioms instead.

A further argument against Pareto in infinite worlds has been offered by Hamkins and Montero [98]. They use the term 'Basic Idea' to refer to the Weak Pareto principle formulated above and the term 'Fundamental Idea' to refer to the $\succcurlyeq$ condition of Pareto, which says that if w_1 and w_2 contain all the same agents and all agents have at least as much utility in w_1 as they do in w_2, then w_1 is at least as good as w_2. Hamkins and Montero accept the the $\succcurlyeq$ condition of Pareto but reject the Weak Pareto principle. They argue that Weak Pareto judges certain worlds to be better than others even though they are equally good. The example they offer in defense of this is as follows:

> Suppose that, in some vast grassy field, there is a soccer team with infinitely many players, some of who are very talented, and some of who are, frankly, dreadful athletes. The talent levels of the players accords with the following pattern:
>
> Team A: ...3, 2, 1, 0, -1, -2, -3, ...
>
> The players with positive talent are very good, while the players with negative talent are terrible, and actually hurt the team by occasionally passing the ball to the opposing team or, worse, scoring goals on the wrong side. The numbers are meant to capture each player's overall contribution to the success of the team. The players wear numbered jerseys, and as it happens, conveniently enough, the talent of each player is identical with the number on his or her shirt. Thus, the player wearing shirt number 27 coincidentally has talent 27, and he is more talented than player number 5, and certainly better than player number -3, who sometimes has problems controlling the ball. Imagine now another team in another field with exactly the same spectrum of talent:
>
> Team B: ...3, 2, 1, 0, -1, -2, -3, ...
>
> And as on Team A, the jersey numbers of the Team B players just happen to match their talent. Now, which is the better team? If we judge the teams in the utilitarian manner, paying attention only to the spectrum of talent on each team, then the answer is obvious: because the teams have exactly the same pattern of talent, we judge them to be equally good. [98, p. 237]

In order for Weak Pareto to be applicable to this example, we must also suppose that the two teams are composed of the same agents and that the agent with talent level 0 on Team A is identical to the agent with talent level 1 on Team A, and so on. Hamkins and Montero ask us to suppose that this is indeed the case: that Team B is actually Team A after some training that increases each agent's talent level. They point out that, after this training, everyone on Team A could pass their shirt down to the person at one talent level below them so that everyone's shirt still reflected their talent levels. They could do this without having to order any new shirts. Hamkins and Montero contend that if Team B is truly better than

Team A, we would expect that the team would have to buy new shirts.[67]

This case shows that two worlds can have the same number of agents and the same 'utility profiles' – the same distribution of utilities across agents – and yet every agent can be better off in one world than they are in the other. If this is the case then we have to decide whether a world must have a different utility profile than another in order to be strictly better than that world. If we believe that the project of ethics is about improving the lives of people then, if this doesn't entail a change in the utility profile of a world, the lesson that we could take away from this case is that one world can be better than another even if they have the same utility profile. After all, if all we care about is how much soccer talent we have then we would clearly think that it's better to be in Team B than in Team A because we know that, no matter who we are, we will have more talent if we are in Team B than if we are in Team A. If everyone agrees that they would prefer to exist in world w_1 than in world w_2, this seems like a strong reason to believe that world w_1 is strictly better than world w_2.

I will return to these and other objections to the Pareto principle in Chapter 2 and again in Chapter 5. In the remainder of this section, however, I want to offer a more detailed discussion of the Zame-Lauwers results using Pareto rather than Generational Sensitivity.

The results that have the most similarity to those that I formulate in this thesis are Zame's theorems. In his 2007 paper, 'Can intergenerational equity be operationalized?' [247], Zame proves four theorems. First, he shows that any complete and transitive relation across infinite utility streams that is invariant to he permutation of the utility of two locations must be indifferent between almost all possible infinite utility streams. Streams that can be *strictly* ordered by an irreflexive, transitive relation that is invariant to the permutation of the utility

[67] This is only true because there is an isomorphism between the talent levels of the agents in the two worlds. There is not always such an isomorphism between worlds that are strictly ranked by Weak Pareto: for example, a world in which the players are all at talent level 2 and a world in which they are all at talent level 1. Hamkins and Montero would prefer to make such comparisons using an Isomorphism Principle, which I will discuss in the next chapter.

of two locations has measure zero and so if the relation is complete then the streams that it is indifferent between must have measure one (Theorem 1). Zame's remaining theorems show that no complete and transitive relation across infinite utility streams that satisfies Generational Sensitivity and is invariant to the permutation of the utility of two locations is measurable (Theorem 2) and therefore the existence of such a relation is independent of Zermelo–Fraenkel set theory with the Axiom of Choice (ZFC) (Theorem 3). If such a relation exists, it cannot be explicitly described (Theorem 4).[68]

I refer the reader to Zame's paper for the proofs of each of these theorems, but it is worth trying to consider the affect that using Pareto rather than Generational Sensitivity has on Zame's results. First, let us use $\succ$ to refer to an irreflexive, transitive relation that satisfies Pareto and is invariant to the permutation of the utility of finitely many agents.[69] If some worlds are not strictly better than others (e.g. if they are equally good) then $\succ$ will not be complete. Assume that agents can have utility levels within a bounded interval [0,2].[70] The claim that $\succ$ is invariant to the permutation of the utility of finitely many agents means that $w_1 \succ w_2$ if and only if there is a world $w_{1'}$ such that the utilities of the agents in $w_{1'}$ are a finite permutation of the utilities of the agents of w_1 and $w_1' \succ w_2$.[71]

Although Zame does not discuss the identities of the basic locations of value across worlds because these are assumed to be generations, their identities are important if we are appealing to Pareto. I believe we must restrict the result to cases in which w_1 and $w_{1'}$ contain the same agents. First I will show that if w_1 and $w_{1'}$ contain different agents then we have no reason to accept the account of $\succ$ given above. Suppose that we do not restrict the account

[68]Lauwers [142] established a similar result showing that a reflexive and transitive relation over infinite streams of 0s and 1s that satisfies Generational Sensitivity and Finite Anonymity is not explicitly describable. The domains in which such relations are non-constructive is explored in Dubey [73].

[69]I appeal to finite permutations rather than permutations of the utilities of just two agents. I don't believe much rests on this. I will often use examples in which only two utility levels are permuted.

[70]Zame only assumes that infinite utility streams contain the [0,1] interval. He offers analogous proofs of his theorems under the assumption that utilities lie in a finite set.

[71]See Zame [247, p.188-9].

of the $\succ$ to cases in which w_1 and $w_{1'}$ have identical populations. Let w_1 and $w_{1'}$ be two worlds, where a_1, a_2, a_3,... are the agents in w_1 and and b_1, b_2, b_3,... are the agents agents in w_2 and the utilities of each agent in w_1 and w_2 are as follows:

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	...
w_1	1	2	1	2	1	2	1	2	1	...
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	...
$w_{1'}$	2	1	1	2	1	2	1	2	1	...

Figure 7: A finite utility permutation with disjoint populations

If just the first two utility levels have been switched, the utilities of the agents of $w_{1'}$ are a finite permutation of the utilities of the agents of w_1. Now consider w_1' and w_2:

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	...
$w_{1'}$	2	1	1	2	1	2	1	2	1	...
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	...
w_2	1	1	1	2	1	2	1	2	1	...

Figure 8: A world worse than $w_{1'}$ by Pareto

World $w_{1'}$ is strictly better than w_2 by Pareto and so $w_1 \succ w_2$ if we do not restrict the account of the $\succ$ to cases in which w_1 and $w_{1'}$ contain the same agents. Finally, consider world w_2 and $w_{2'}$:

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	...
w_2	1	1	1	2	1	2	1	2	1	...
	a_3	a_5	a_1	a_7	a_9	a_2	a_{11}	a_4	a_{13}	...
$w_{2'}$	1	1	2	1	1	2	1	2	1	...

Figure 9: A finite utility permutation of w_2

The utilities of the agents of $w_{2'}$ are a finite permutation of the utilities of the agents of w_2. Therefore $w_1 \succ w_{2'}$ by transitivity. But compare the utilities of the agents in w_1 and $w_{2'}$:

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	...
w_1	1	2	1	2	1	2	1	2	1	...
	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	...
$w_{2'}$	2	2	1	2	1	2	1	2	1	...

Figure 10: Comparing worlds w_1 and $w_{2'}$

We can see that $w_{2'} \succ w_1$ by Pareto. This contradicts the claim that $w_1 \succ w_{2'}$ and so we must restrict the account of the $\succ$ to cases in which w_1 and $w_{1'}$ have identical populations.

If w_2 can have a different population from w_1 and $w_{1'}$ then w_1 and $w_{1'}$ will never be strictly better than (or equal to) w_2 by Pareto. But w_1 or $w_{1'}$ may be better than w_2 by a principle other than Pareto. Therefore we need not assume that w_2 must have the same population as w_1 and $w_{1'}$. To most neatly draw out the analogue of Zame's result using Pareto, however, I will focus on a set of worlds with identical populations.

If we accept Pareto rather than Generational Sensitivity, the best analogue to Zame's account of the $\succ$ relation is the claim that $w_1 \succ w_2$ if and only if there is a world $w_{1'}$ such that w_1 and $w_{1'}$ contain the same agents and the utilities of the agents in $w_{1'}$ are a finite permutation of the utilities of the agents of w_1 and $w_1' \succ w_2$. This claim is not ethically trivial. For example, we might reject it because we think that it is better for more utility to one agent over another because they are more deserving. In other words, this principle seems to be inconsistent with the idea that one world can be better than another based on factors other than it's distribution of utility. Let us simply grant this assumption for now, however.

If we focus on worlds with identical populations and accept the assumptions listed above, Zame's theorem 1 shows that if $\succ$ an irreflexive, transitive relation that is invariant to the permutation of the utility of finitely many agents, then the set of identical population worlds that are comparable by $\succ$ will have measure zero. It does not seem implausible that this result will also hold across worlds with variable populations. Zame's first theorem does not rely on the Pareto principle. The set of identical population worlds that are comparable by $\succ$ will have measure zero because it is an irreflexive relation that is invariant to the permutation of the utility of finitely many agents. As Zame notes:

> 'even if we abandon completeness and respect for the Pareto ordering, every irreflexive preference relation that displays intergenerational equity has the property that almost all pairs of utility streams are incomparable (not strictly ranked). [247, p. 188]

That the $\succ$ relation is not complete should not be particularly troubling, since some worlds may be equally good. However, Zame's final three theorems show that if we find an extension $\succcurlyeq$ of $\succ$ that is a complete, transitive relation that satisfies Generational Sensitivity and invariance to the permutation of the utility of finitely many agents then the existence of this relation is independent of ZFC cannot be explicitly described. Let us suppose that this proof can be extended to worlds that are comparable by Pareto rather than Generational Sensitivity.[72] It would follow from this that the existence of a complete, transitive relation that satisfies Pareto and invariance to the permutation of the utility of finitely many agents is possible is independent of ZFC and cannot be explicitly described. We have already encountered objections to appeals within ethics to objects that cannot be explicitly described when discussing Svensson's possibility proof in the previous section.

As I noted in the previous section, the Diamond-Basu-Mitra results generate problems for the claim that there is a numerically representable social welfare function that satisfies Gen-

[72] The key differences Pareto introduces are that it only holds between identical population worlds and it is a principle across basic locations (people) that come in no natural order. We can simply restrict our discussion to worlds with identical populations. We must show that Theorems 2-4 can be established without the need to assume a privileged ordering of agents, however.

erational Sensitivity and Finite Generational Anonymity. We may be able to extend these results using Pareto and Finite Agent Anonymity rather than generation-based principles. Since I am not concerned with the numerical representability requirement, however, the Zame-Lauwers results are of greater relevance here. I believe that we may be able to extend the Zame-Lauwers result to worlds using the Pareto principle rather than Generational Sensitivity. If we do then we can establish that the existence of a complete, transitive relation that satisfies Pareto and invariance to the permutation of the utility of finitely many agents is possible is independent of ZFC and cannot be explicitly described.

In this thesis I will establish results that go beyond the problems generated by the Zame-Lauwers result for ethical orderings. First, I will not assume that the $\succcurlyeq$, the at least as good as relation, must be invariant to the permutation of the utility of finitely many agents. As I noted above, any ethical theory that does not rely wholly on the distribution of utility levels to rank worlds may reject this principle. Even theories that rely entirely on the distribution of utility levels to rank worlds need not endorse it. Within the intergenerational equity literature it has generally been assumed that anonymity principles ensure that an ethical theory is 'equitable'. In the next chapter I will argue that this we can derive a more plausible equity constraint from Pareto and the claim that if the world pair $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ in all qualitative respects, then w_1 is at least as good as w_2 if and only if w_3 is at least as good as w_4. I believe that it is much more difficult to deny the claim that $\succcurlyeq$ should hold between qualitative duplicates than the claim that $\succcurlyeq$ should be invariant to the permutation of the utility of finitely many agents.

I will also show that if we can permute the populations of worlds then the existence of a transitive and complete $\succcurlyeq$ relation that satisfies that satisfies Pareto and the 'invariance across qualitative duplicates' principle mentioned above is not merely independent of ZFC: such a relation does not exist. I therefore show using an axiom that is much more plausible than finite permutation invariance that it is not possible to find a transitive and complete ordering of worlds that satisfies Pareto. I also show that an ordering that satisfies Pareto plus these auxiliary axioms fails to infinite worlds whenever the utilities of the populations of those worlds that will typically hold between infinite worlds. I conclude that if we accept Pareto then incomparability between infinite worlds will be the norm rather than the exception.

Finally, I will show that these results apply not only to worlds with identical populations but also to worlds with disjoint and overlapping populations. I also show that they generate novel problems for objective and subjective theories of permissibility.

I will begin to lay the foundations for these results in the next chapter. Before concluding this chapter, however, I will survey the principles that have been developed within economics and philosophy to rank worlds that contain infinitely many 'basic locations of value'.

1.5 Infinite Aggregation Principles

Within the infinite ethics literature and intergenerational equity literature, many principles have been proposed for comparing worlds that contain infinite utility.[73] In this section I will attempt to offer an overview of the main solutions.[74] I divide the main solutions into four families. These families are: utility difference principles, utility density principles, alternative formalizations of infinity, and discounting principles.

[73]Decision theorists working on the problems generated by the possibility of infinite streams of utility for a single agent, such as Pascal's wager, the St Petersburg paradox, and the Pasadena game, have also formulated methods for comparing infinite utility streams that could be applied to the problems in infinite ethics. I will discuss the analogueies between solutions proposed in decision theory and ethics in chapter 6.

[74]In this chapter I will focus on existing solutions that attempt to preserve Sensitivity. In chapter 7 I will discuss some solutions that are inconsistent with Sensitivity and have not been explored in much depth.

1.5.1 Utility Difference Principles

To make it easier to discuss utility difference principles, it will be helpful to introduce some new terms. First, let $\delta_{1\text{-}2}(l_i)$ be the utility at world w_1 at location l_i minus the utility at world w_2 at location l_i. So if agents are locations and agent a_4 has utility 7 in w_1 and the same agent has utility 2 in w_2, then $\delta_{1\text{-}2}(a_4) = 5$. Next, let $\sigma_{1,2}$ be an enumeration of all of the locations of w_1 and w_2 – i.e. an ordering of locations such as $l_1, l_2, l_3, l_4, ...$ or $l_2, l_1, l_4, l_3, ...$ (for now, I will only consider pairs of worlds that share all of the same locations). Let $S_{1,2}$ be the set of all possible enumerations of the locations of w_1 and w_2: the set of all $\sigma_{1,2}$.[75] If locations come in a privileged order, such as the natural ordering of times, then I will use $\sigma^*_{1,2}$ to refer to this one privileged enumeration of the locations of w_1 and w_2.

As we have seen, one of the key problems in infinite worlds is that simply summing the utilities in two infinite worlds will often fail to take into account important differences between those streams. For example, the sums of $2, 2, 2, 2, ...$ and $1, 1, 1, 1, ...$ both diverge to positive infinity, but presumably we believe that a world in which each location has utility 2 is better than a world in which each location has utility 1.[76] Expressed in our new terminology, Sensitivity says that if, for all locations l_i in w_1 and w_2, $\delta_{1\text{-}2}(l_i) \geq 0$ and for some location l_j in w_1 and w_2, $\delta_{1\text{-}2}(l_j) > 0$, then w_1 is at least as good as w_2.

We have noted that reasonable aggregation principles should entail Sensitivity, but Sensitivity is an extremely weak principle: if some locations have more utility in w_1 and some locations have more utility in w_2 then Sensitivity is silent about w_1 and w_2. But consider

[75] Here I use symmetric group notation $S_{1,2}$ to refer to the set of all possible permutations of these populations, where the subscript $1, 2$ refers to the shared locations of w_1 and w_2 (which comprise all the locations of w_1 and w_2) and not to the worlds themselves. Since the locations in w_1 and w_2 are countably infinite, there are uncountably many possible enumerations of these locations in $S_{1,2}$ because the cardinality of the symmetric group on an infinite set is equal to the cardinality of the power set of that set.

[76] Hamkins and Montero [98] [100] argue against this principle, while others have argued that it ought to be weakened. I will consider these objection in Chapter 2 and 4, but here I will simply assume that it is correct.

the utilities at each location of worlds w_1 and w_2, which both have all the same loctions:

	l_1	l_2	l_3	l_4	l_5	l_6	...
w_1	1	2	2	2	2	2	...
w_2	2	1	1	1	1	1	...

Figure 11: World w_2 is better at a single location

World w_1 has one location – l_1 – that is at utility level 1 and all of the other locations of w_1 are at utility level 2. World w_2 has the same locations as w_1, but location l_1 is at utility 2 in w_2 and all of the other locations of w_2 are at utility level 1. World w_1 seems better than w_2 since there is a set of locations $l_2, l_3, ...$ that are jointly better off in w_1 by infinitely many utils and there is no set of locations that are jointly better off in w_2 by infinitely many utils.[77] But Sensitivity does not entail that w_1 is better than w_2.

What I am calling 'utility difference principles' are principles that solve this problem by using the *differences* in the utilities at locations to find the total value of a world. The Catching-Up principle is a well-known principle of this sort developed within the economics literature for comparing infinite streams of utility, an extension by Gale [89] of a principle first developed by Atsumi [15] and von Weizsäker [239]. As we have seen, many in the economics literature assume that generations are basic locations of value and that they come in a natural temporal order. Since the Catching-Up principle first developed by Atsumi and von Weizsäcker assumes a privileged ordering of locations (generations), I will call it 'Ordered Catching-Up'. The principle is as follows:[78]

[77]Dutta [75] calls a similar principle for generations 'Veto Power of Infinitely Many Generations'.

[78]Here and below I will mirror the formulation of Catching-Up given in Lauwers and Vallentyne [148]. Their formulation is slightly atypical: usually the Catching-Up principle says that w_1 is at least as good as w_2 if w_1 and w_2 share all the same locations and there is some location $l_j \geq 1$ in w_1 and w_2 such that for all locations at least as great as l_j, the sum of utilities in the locations of w_1 and is greater than or equal to the sum of their utilities in w_2 (i.e. lor all locations $l_k \geq l_j, \sum u(l_k), w_1 \geq \sum u(l_k), w_2$). One unfortunate feature of Lauwers and Vallentyne's formulation is that, unlike the standard formulation, it does not entail total utilitarianism in finite worlds. Setting aside comparisons between infinite utility and finite utility worlds, we would need to replace the limit inferior with the infumum if both worlds in a pair contain only finite utility.

Ordered Catching-Up (identical locations)

If w_1 and w_2 contain the same locations in the order $\sigma^*_{1,2}$ and the lower limit, as n approaches infinity, of the sum of $\delta_{1\text{-}2}(l_i)$ at locations 1 to n is at least as great as [strictly greater than] zero, w_1 is at least as good as [strictly greater than] w_2.

Let $p(w_1)$ be all of the basic locations of w_1, so that $p(w_1) = p(w_2)$ if and only if w_1 and w_2 have the same basic locations. Given this, we can express the Ordered Catching-Up principle more formally as follows:[79]

Ordered Catching-Up: $w_1 \succcurlyeq w_2$ if $p(w_1) = p(w_2)$ and, under $\sigma^*_{1,2}$, $\lim_{n\to\infty} \inf \sum_{i=1}^{n} \delta_{1\text{-}2}(i) \geq 0$

$w_1 \succ w_2$ if $p(w_1) = p(w_2)$ and $w_1 \succcurlyeq w_2$ and $w_2 \not\succcurlyeq w_1$

We can demonstrate how the Ordered Catching-Up principle works by looking at what it says in the example given above. To simplify things, let's assume that the privileged enumeration $\sigma^*_{1,2}$ of the locations of w_1 and w_2 is just $l_1, l_2, l_3, ...$, i.e. the order that they are presented in above. We can add the values of $\delta_{1,2}(l_i)$ for each of the locations in w_1 and w_2 and the sum of these values to our diagram as follows:

	l_1	l_2	l_3	l_4	l_5	l_6	...	
w_1	1	2	2	2	2	2	...	
w_2	2	1	1	1	1	1	...	
$\delta_{1\text{-}2}$	-1	1	1	1	1	1	...	$\lim\inf = 1$
$\sum \delta_{1\text{-}2}$	-1	0	1	2	3	4	...	$\lim\inf = \infty$

Figure 12: Applying Ordered Catching-Up to w_1 and w_2

Since the limit inferior of the sum of all $\delta_{1,2}(l_i)$ diverges to positive infinity, w_1 is strictly

[79] Generally, a strict better than condition for the Catching-Up principle is not explicitly formulated, with the exception of Dutta [75]. This definition of strictly better than seems to be implicit in much of the literature, however. For example, the verdict in the case in Asheim and Tungodden [13, p. 58] that I discuss below does not follow unless we assume something like the definition I have given here.

better than w_2 according to the Ordered Catching-Up principle. The Ordered Catching-Up principle also entails Sensitivity for the (ordered) locations being treated as basic locations of value. If all locations have at least as much utility in w_1 as they do in w_2, then it is clear that the limit inferior of the sum of $\delta_{1,2}(l_i)$ for all locations must be at least as great as zero. And if some of these locations have strictly greater utility in w_1 than they do in w_2, then the limit inferior of the sum of $\delta_{1,2}(l_i)$ for all locations must be strictly positive. Unlike Sensitivity, however, Ordered Catching-Up can compare worlds in which finitely many of the locations of w_1 and w_2 have strictly lower utility in w_1 than in w_2 and infinitely many of the locations of w_1 and w_2 have strictly greater utility in w_1 than in w_2.

The Ordered Catching-Up principle says that w_1 is strictly better than w_2 if the limit inferior of the sum of all values of $\delta_{1,2}(l_i)$ is positive and not just that the limit inferior of all values of $\delta_{1,2}(l_i)$ is positive. This is because if finitely many locations of w_1 and w_2 have more utility in w_1 than in w_2 and all other locations of w_1 and w_2 have the same utility in both worlds then the limit inferior of all values of $\delta_{1,2}(l_i)$ will be zero. So if we claimed that w_1 is at least as good as w_2 if the limit inferior of all values of $\delta_{1,2}(l_i)$ is at least as great as zero, then we would be forced to conclude that w_2 is at least as good as w_1 in this kind of case, which violates Sensitivity. But the limit inferior of the sum of all values of $\delta_{1,2}(l_i)$ will be some strictly positive finite value in this kind of case. Therefore Ordered Catching-Up will say that w_1 is strictly better than w_2, which is consistent with Sensitivity.

Not all cases in which infinitely many agents are better off in w_1 than in w_2 and finitely many agents are better off in w_2 than w_1 will be such that $w_1 \succ w_2$ according to Ordered Catching-Up. For example, suppose that the first agent of w_1 has utility 2 and the remaining agents of w_1 have utility 0, while each agent l_n of w_2 has utility equal to $\frac{1}{n^2}$ as follows:[80]

[80] Finding the exact sum of this sequence was also known as the 'Basel problem' and was solved by Euler who demonstrated that the exact value of the sum of this sequence is $\frac{\pi^2}{6} \approx 1.645$. Given this, the sum of $\delta_{1,2}(l_i)$ for all locations l_i is $2 - \frac{\pi^2}{6} \approx 0.355$.

	l_1	l_2	l_3	l_4	l_5	l_6	...	
w_1	2	0	0	0	0	0	...	$= 2$
w_2	1	$\frac{1}{4}$	$\frac{1}{9}$	$\frac{1}{16}$	$\frac{1}{25}$	$\frac{1}{36}$	...	≈ 1.645
$\delta_{1\text{-}2}$	1	$-\frac{1}{4}$	$-\frac{1}{9}$	$-\frac{1}{16}$	$-\frac{1}{25}$	$-\frac{1}{36}$	...	≈ 0.355
$\sum \delta_{1\text{-}2}$	1	$\frac{3}{4}$	$\frac{23}{36}$	$\frac{83}{144}$	$\frac{1931}{3600}$	$\frac{79}{144}$	...	$\liminf \approx 0.355$

Figure 13: Lower limit inferior with infinitely many agents better off

In this case there are infinitely many agents – $l_2, l_3, l_4, ...$ – that are better off in w_2 than in w_1 but the total utility that those agents are better off by in w_2 is approximately 0.645 utils. There is only one agent – l_1 – that is better off in w_1 than in w_2, but this agent is better off by 1 util. So even though infinitely many agents are better off in w_2 than in w_1 and finitely many agents are better off in w_2 than in w_1 the total utility of w_1 is strictly greater than the total utility of w_2. The limit inferior of the sum of $\delta_{1,2}(l_i)$ for all locations in w_1 and w_2 is strictly greater than zero, and so w_1 is strictly better than w_2 by Ordered Catching-Up.

The Ordered Catching-Up principle does have some questionable results, however. Consider the following case given by Asheim and Tungodden [**?**, p. 58]:[81]

	l_1	l_2	l_3	l_4	l_5	l_6	...	
w_1	1	1	1	1	1	1	...	
w_2	2	0	2	0	2	0	...	
$\delta_{1\text{-}2}$	-1	1	-1	1	-1	1	...	
$\sum \delta_{1\text{-}2}$	-1	0	-1	0	-1	0	...	$\liminf = -1$
$\delta_{2\text{-}1}$	1	-1	1	-1	1	-1	...	
$\sum \delta_{2\text{-}1}$	1	0	1	0	1	0	...	$\liminf = 0$

Figure 14: A single location makes a difference to Ordered Catching-Up

[81]Similarly, as Crespo, Núñez, and Rincón-Zapatero [64, p. 8] point out, the stream $(1, 0, 1, 0, 1, 0, ...)$ is strictly better than the stream $(0, 1, 0, 1, 0, 1, ...)$ according to the Ordered Catching-Up principle.

World w_2 does not strike us as better than w_1. After all, every two locations in world w_2 have 2 utils and so do every 2 locations in w_1. But according to the Ordered Catching-Up principle, w_2 is at least as good as w_1 (since the lim inf of $\sum \delta_{2\text{-}1}$ is 0) and w_1 is not at least as good as w_2 (since the lim inf of $\sum \delta_{1\text{-}2}$ is -1). Therefore w_2 is strictly better than w_1 by Ordered Catching-Up. Now consider what would happen if we were to remove location l_1 from this sequence, e.g. if times are the basic locations of value, we wait until the first time period has passed. The limit inferior of $\sum \delta_{1\text{-}2}$ if l_1 is removed is 0 and so, by Ordered Catching-Up, w_1 is at least as good as w_2. Surely adding or removing a single location at the beginning of the sequence should not have this impact on world rankings.

To avoid this kind of result we can retreat to the subrelation[82] of the Ordered Catching-Up principle called the Overtaking principle, which is the principle originally formulated by von Weiszäcker [239]. We can formulate the Ordered Overtaking principle as follows:

Ordered Overtaking:

$$w_1 \sim w_2 \text{ if } p(w_1) = p(w_2) \text{ and, under } \sigma^*_{1,2}, \exists l_j \geq l_1 \text{ such that } \forall l_k \geq l_j, \sum_{k=1}^{n} \delta_{1\text{-}2}(l_k) = 0$$

$$w_1 \succ w_2 \text{ if } p(w_1) = p(w_2) \text{ and, under } \sigma^*_{1,2}, \exists l_j \geq l_1 \text{ such that } \forall l_k \geq l_j, \sum_{k=1}^{n} \delta_{1\text{-}2}(l_k) > 0$$

To see how Ordered Overtaking works, suppose that w_1 and w_2 share the same locations and the utilities of these locations in w_1 are $(-1, -1, 2, 2, 2, ...)$ while the utilities of these locations in w_2 are $(1, 1, 1, 1, 1, ...)$. This means that the sequence $\delta_{1,2}$ is $(-2, -2, 1, 1, 1, ...)$ and so $\sum \delta_{1,2}$ is $(-2, -4, -3, -2, -1, 0, 1, 2, ...)$. We can see that $\sum \delta_{1,2}$ is strictly positive after location l_6 in this case, and so w_1 is better than w_2 by Ordered Overtaking.[83]

[82] If R_1 is a subrelation of R_2 then $x \succcurlyeq_{R_1} y \rightarrow x \succcurlyeq_{R_2} y$. Dutta [75] notes that the Overtaking principle is a subrelation of the Catching-Up principle: if $w_1 \succcurlyeq w_2$ by the Overtaking principle than $w_1 \succcurlyeq w_2$ by the Catching-Up principle (but the reverse is not always true).

[83] I have formulated the 'utilitarian' version of the Ordered Overtaking principle here, but Aumann and Shapley [16] formulate a 'Long run averages' principle that arguably plays the role of an average utilitarian overtaking principle, while Bossert, Sprumont and Suzumura [40] formulate a version of the leximin criterion

Unlike the Ordered Catching-Up principle, the Ordered Overtaking principle does not say that w_2 is strictly better than w_1 in the case above. Since $\sum \delta_{2\text{-}1}$ is $(-1, 0, -1, 0, -1, 0, ..)$ there is no location l_j such that, for all locations after l_j, $\sum \delta_{1\text{-}2}$ is greater than zero and no location l_j such that, for all locations after l_j, $\sum \delta_{1\text{-}2}$ is equal to zero. The same is true of $\sum \delta_{1\text{-}2}$. Even though $(1, 0, 1, 0, 1, 0, ..)$. is always at least as great as zero, there is no location after which it is equal to zero or strictly greater than zero. Therefore worlds w_1 and w_2 are not comparable by Ordered Overtaking.[84]

This illustrates a more general problem that affects both the Ordered Overtaking principle and the Ordered Catching-Up principle: namely, that there are many world pairs that are not comparable by either principle. Consider worlds w_1 and w_2:

	l_1	l_2	l_3	l_4	l_5	l_6	...	
w_1	1	3	0	1	3	0	...	
w_2	2	0	2	2	0	2	...	
$\delta_{1\text{-}2}$	-1	3	-2	-1	3	-2	...	
$\sum \delta_{1\text{-}2}$	-1	2	0	-1	2	0	...	$\liminf = -1$
$\delta_{2\text{-}1}$	1	-3	2	1	-3	2	...	
$\sum \delta_{2\text{-}1}$	1	-2	0	1	-2	0	...	$\liminf = -2$

Figure 15: A world pair that Catching-Up and Overtaking do not rank

These worlds are not comparable by either the Ordered Catching-Up rule or the Ordered Overtaking rule. The limit inferior of $\sum \delta_{1\text{-}2}$ and $\sum \delta_{2\text{-}1}$ are both negative and so neither world 'catches up' with the other.[85] And there is no location such that, for every location

for infinite worlds. Others like Asheim and Tungodden [13, p. 227] have formulated leximin versions of both Ordered Catching-Up and Ordered Overtaking. I won't discuss these non-utilitarian principles in much detail here, but I will return to such principles in Chapter 5.

[84] This point is argued by Asheim and Tungodden [13, p. 227].

[85] Worlds w_1 and w_2 are incomparable even if we adopt the standard formulation of the Catching-Up principle described in note 78.

$l_k \geq l_j$, the utility of $\sum \delta_{1\text{-}2}(l_k)$ or $\sum \delta_{2\text{-}1}(l_k)$ is equal to zero or strictly greater than zero. So neither world 'overtakes' the other. The Ordered Catching-Up principle and Ordered Overtaking principle are both consistent with Sensitivity and Finite Anonymity.[86] As we can see, however, neither principle satisfies completeness. Neither relation can compare ordered utility streams unless there is some finite location l_j in the sequence such that after with the sum of the difference in utilities across all future locations is strictly better in one of the streams. As the example above shows, this will not always be the case.

A further limitation of the Ordered Catching-Up principle and the Ordered Overtaking principle that they cannot be applied to basic locations that come in no privileged order, such as agents. I will discuss ways that we can (and cannot) extend ordered utility difference principles to basic locations of value that come in no natural order in Chapter 5.

1.5.2 Utility Density Principles

In 'Sacrificing the Patrol: Utilitarianism, Future Generations and Infinity' Luc Van Liedekerke and Luc Lauwers [237] offer an alternative to utility difference principles that appeals to the *density* of utility levels within a given world. Van Liedkerke and Lauwers are not the only authors to have proposed utility density principles for comparing infinite worlds[87] but theirs is arguably the most sophisticated version of the utility density approach. It also makes use of concepts like accumulation points and natural limits that I will refer back to in later chapters, and so I will focus on their view here.

The first concept that Van Liedkerke and Lauwers' utility density approach requires is that of an 'accumulation point' of a set of utility levels, defined as follows:[88]

[86] Banerjee [17] and Kamaga and Kojima [122] show that, unlike the Ordered Overtaking principle, the Ordered Catching-Up principle is not consistent with Fixed-Step Anonymity.

[87] See, for example, Lauwers [144], Bostrom [42], and Petri [181].

[88] Here I largely follow Van Liedkerke and Lauwers' definition of accumulation points. They use the term

Accumulation Points

Let $U(w_1)$ be the set of bounded utility levels $(u_1, u_2, u_3, ..., u_n, ...)$ of every location in world w_1. Then v is an accumulation point of the set $U(w_1)$ if, for all $\epsilon > 0$ and for all n, there exists an $m > n$ such that $|u_m - v| < \epsilon$.

So utility level v is an accumulation point of $U(w_1)$ if there are always utility levels that are epsilon-close to v.[89] As Van Liedkerke and Lauwers note, accumulation points can be thought of as utility levels shared by infinitely many locations in w_1. So if a world w_1 contains infinitely many locations at utility 1 and infinitely many locations at utility 2 and no other locations, then the accumulation points of this world are 1 and 2, no matter what order these locations come in (if any). The Bolzano-Weierstrass theorem entails that every infinite sequence whose components are bounded has at least one accumulation point.[90]

Van Liedkerke and Lauwers argue that knowing the accumulation points alone is not enough to be able to compare infinite worlds. In defense of this, they ask us (p. 16, *ibid.*) to consider utility streams $y : (1, 0, 1, 0, ..., 1, 0, ...)$ and $z : (1, 1, 0, 1, 1, 0, ..., 1, 1, 0, ...)$. These two streams have the same accumulation points – 1 and 0 – but Van Liedkerke and Lauwers claim that more people hold utility level 1 in utility stream z than in utility stream y.

Whether this claim is true or not will depend on what our basic locations of value are. Van Liedkerke and Lauwers assume that generations are the basic locations of value. As we saw in the previous section, if we think that the same agents can appear in different generations without altering those generations (i.e. we have a time-based account of generations) then it is possible for a world w_1 to be better than another world w_2 according to Sensitivity for generations, but for world w_2 to be better than world w_1 by Sensitivity for people. So Van

'cluster point', but this is merely a terminological difference. I will return to accumulation points and related concepts in Chapter 4, where I will discuss methods for ranking worlds that appeal solely to the accumulation points of their utility profiles.

[89] Note that v can be an accumulation point even if no locations have utility equal to v.

[90] Bartle and Sherbet [20, p. 78].

Liedkerke and Lauwers' claim that more people have utility 1 in z than in y seems most plausible if we assume an agent-based account of generations: for example, if, for now, we simply treat their 'utility streams' as a temporal sequence of individual agents.

It will also depend on what we mean when we say that 'more' people have utility 1 in the first world. The agents that have utility 1 in stream y can be put in one-to-one correspondence with the agents that have utility 1 in stream z,so by 'more' we must mean something other than having greater cardinality.[91] To see what is meant by 'more' here, consider the set of natural numbers $(1, 2, 3, 4, 5, ...)$ and the set of cube numbers $(1, 8, 27, 64, 125, ...)$. All cube numbers are natural numbers, but not all natural numbers are cube numbers. So there is a sense in which there are 'more' cube numbers than there are natural numbers even though the two can be put in one-to-one correspondence. Let's grant that if the set of agents with utility n in w_1 is a proper subset of the set of agents with utility n in w_2, then it is acceptable to say that there are more agents at utility n in w_1 than in w_2.[92]

Not only are cube numbers a proper subset of natural numbers; they also become more scarce as we progress along the natural number line: there seem to be even more natural numbers per cubed number as the natural numbers increase. We can imagine something similar happening when we compare two utility streams. For example, consider utility stream z : $(1, 1, 0, 1, 1, 0, ..., 1, 1, 0, ...)$ compared with a utility stream x in which every cube-numbered location contains utility 1 and all other locations contain utility 0. The utility 1 locations seem scarcer in x than they are in either y or z.[93]

Let us say that an accumulation point v is of more relative importance in w_1 than in w_2 if

[91] Hamkins and Montero [99, p. 239] call the claim that 'If a set U is obtained by adding elements to another set V, then U has a larger size than V' the Basic Mathematical Idea (BMI) and point out that despite being overwhelmingly intuitive, it is false. This is correct if we adopt a Cantorian account of size.

[92] I will return to this non-Cantorian notion of set size in Chapter 5.

[93] We can think of the scarcity of a number type as the probability that a dart thrown randomly at the natural number line will hit a number of that type.

the locations with utilities in the neighborhood of v are less scarce in w_1 than in w_2 (and therefore that there are 'more' of them). Van Liedkerke and Lauwers argue that we can measure the 'relative importance' of each accumulation point using a density function. The particular density function they appeal to is the 'natural' or 'asymptotic' density function. Let L be the entire set of locations of world w_1 ordered by σ and let G be some subset of these locations. The natural density d of G relative to L is given by:

$$d(G) = \lim_{n\to\infty} \frac{|G \cap \{l_1, l_2, ..., l_n\}|}{n}$$

G has a natural density whenever this limit exists.[94] As Van Liedkerke and Lauwers note, natural density has several desirable features: if G has density d then its complement $L\backslash A$ has density $1-d$ and if G and G' have natural densities d and d' then $G \cup G'$ has natural density $d+d'$. Its results also accord with our intuitions. For example, the natural density of the set locations H with utility 1 in streams y, and z is as follows:

$$y = (1, 0, 1, 0, 1, 0, ..., 1, 0, 1, ...), d(H) = \tfrac{1}{2}$$

$$z = (1, 1, 0, 1, 1, 0, ..., 1, 1, 0, ...), d(H) = \tfrac{2}{3}$$

Figure 16: Natural density of two utility streams

There are, however, some important limitations of natural densities. The first limitation is that the natural density of any finite set in an infinite set is zero. So if we are to use natural densities to compare infinite worlds, we will often end up being insensitive to increases in

[94]The upper natural density of G is the limit superior of $|G \cap \{l_1, l_2, ..., l_n\}|/n$ as n goes to infinity, while the lower natural density G is the limit inferior of $|G \cap \{l_1, l_2, ..., l_n\}|/n$ as n goes to infinity. A set may have an upper or lower natural density without having annatural density: it only has the latter if the upper and lower natural density exist and are equal. Since this limit does not always exist, natural densities are not defined for all subsets of the natural numbers.

utility at finitely many locations in a world.[95]

The more important limitation of natural densities is that they are order-dependent. The utility stream y can result from an infinite permutation of z and yet the set of agents at utility 1 in y and z have different natural densities.[96] If the basic locations of value in set L have no natural order then they can occur in any order $\sigma \in S_L$. The natural density of a set of locations G in L will therefore not be defined unless G has natural density zero (all but finitely many locations in L are in G) or G has natural density one (at most finitely many locations in L are in G). Remember that Liedkerke and Lauwers said that more *people* are at utility 1 in stream z than they are in stream y. But agents come in no natural order and if agents can appear in different generations then there could in fact be more agents at utility 1 in stream y than in stream z. Therefore theories that appeal to natural densities are only useful if we assume that the basic locations of value have some natural or privileged order.[97]

Van Liedkerke and Lauwers develop a rule for comparing infinite utility streams based on the 'medial limit' or 'generalized Cesaro limit' of each stream. To define this, we first consider the sequence of the average utility at the ordered locations of w_1:

$$Av_1 = avg(U_{\sigma^*}(w_1)) = \left(\frac{u(l_1)}{1}, \frac{u(l_1)+u(l_2)}{2}, ..., \frac{u(l_1)+u(l_2)+, ..., +u(l_n)}{n}, ...\right)$$

We are then asked to consider the accumulation points of the sequence Av_1. Since the components of $U_{\sigma^*}(w_1)$ are bounded, there exists a limit inferior and a limit superior of these accumulation points. The medial limit of Av_1 is a linear mapping $F(U_{\sigma^*}(w_1))$ between this limit inferior and limit superior.[98] In particular, if the limit inferior and limit superior

[95] Van Liedkerke and Lauwers [237, p.169-170] accept this result and a weaker Sensitivity principle.

[96] Natural densities remain constant under what Van Liedkerke and Lauwers (p. 168, 1997) call 'bounded permutations': a permutation σ such that $\lim_{l\to\infty} \frac{\sigma(l)-l}{l} = 0$.

[97] Or that they have a class of privileged orders that are all bounded permutations of one another.

[98] More precisely, $\lim_n \inf(A_1) \leq F(U_{\sigma^*}(w_1)) \leq \lim_n \sup(A_1)$, where linearity means that for all ordered utility sequences with bounded components $U(w_i), U(w_j)$ and all real numbers n, m, $F(n(U(w_i)) +$

of Av_1 both exist and are equal, then this is the medial limit of $U_{\sigma^*}(w_1)$ is equal to these limits. For example, consider the utilities of worlds w_1 and w_2:

	l_1	l_2	l_3	l_4	l_5	l_6	...	l_n	l_{n+1}	l_{n+2}	...
w_1	0	3	1	0	3	1	...	0	3	1	...
w_2	2	2	1	2	2	1	...	2	2	1	...

Figure 17: Utility sequences with distinct medial limits

The accumulation points of w_1 are 0, 1, and 3. The accumulation points of w_2 are 1 and 2. In w_1 the natural density of each accumulation point is $\frac{1}{3}$, while in w_2 the natural density of 2 is $\frac{2}{3}$ and the natural density of 1 is $\frac{1}{3}$. So the medial limits of these utility sequences are:

$$F(U_{\sigma^*}(w_1)) = \tfrac{1}{3}(0) + \tfrac{1}{3}(3) + \tfrac{1}{3}(1) = 1\tfrac{1}{3}$$
$$F(U_{\sigma^*}(w_2)) = \tfrac{2}{3}(2) + \tfrac{1}{3}(1) = 1\tfrac{2}{3}$$

So if $w_1 \succcurlyeq w_2$ iff $F(U_{\sigma^*}(w_1)) \geq F(U_{\sigma^*}(w_2))$ and $w_1 \succ w_2$ iff $F(U_{\sigma^*}(w_1)) > F(U_{\sigma^*}(w_2))$ then$w_1 \succ w_2$ in this case, which is what we might expect from this sort of rule.

I have already noted two key problems with using natural densities to compare infinity utility streams. The first is that they are insensitive to finite improvements and the second is that they are highly order-dependent. A third problem is that if the limit inferior and limit superior of Av_1 don't coincide, medial limits can only be defined non-constructively using the Axiom of Choice or non-principal ultrafilters, which I discuss below.

1.5.3 Alternative Formalizations of Infinity

So far I have assumed that infinite worlds consist of countably infinitely many locations, each with bounded utility. The utilities of these worlds have been represented as ordered

$m(U(w_j))) = nF(U(w_i)) + mF(U(w_j))$.

or unordered sequences of real numbers.[99]There are, however, alternative ways that we can respresent the utility across locations at infinite worlds. In this section I will consider two that have gained attention recently: the hyperreals and the surreals.

In 'Infinite Ethics', Nick Bostrom [42] argues that we could use 'hyperreal numbers' to represent utilities in infinite worlds. The hyperreals are an extension of the real numbers that let us represent different infinitesimal and infinite numbers. We can represent hyperreals using sequences of real numbers.[100] The hyperreal representation of a real number is just an infinite sequence of that real number. For example, the hyperreal representation of the real number 3 is $< 3, 3, 3, ... >$. Infinitesimal hyperreals are sequences that approach zero, such as $< 1, \frac{1}{3}, \frac{1}{9}, \frac{1}{27}, ... >$.[101] Infinite hyperreals are sequences that increase towards infinity, such as $< 1, 4, 8, 12, ... >$. Addition, subtraction, division, and multiplication of hyperreals are all pointwise operations. For example, $< 3, 3, 3, ... > \times < 1, 4, 8, 12, ... >=< 3, 12, 24, 36, ... >$.[102] If all of the components of one hyperreal are strictly greater than all of the components of another hyperreal, then the first hyperreal is strictly larger than the second. So the hyperreal $< 3, 12, 24, 36, ... >$ is strictly larger than the hyperreal $< 1, 4, 8, 12, ... >$.

Suppose that basic locations of value of come in a natural order, and that the utility levels of the locations of world w_1 are $(2, 3, 2, 3, 2, 3, ...)$. Bostrom [42] argues that we can represent this world using the hyperreal representation of the sum of this sequence. So this world can be represented by the hyperreal $< 2, 5, 7, 10, 12, 15, ... >$.[103] We can use these hyperreal

[99]I also implicitly assume a cardinality-based account of size: i.e. that if the locations of two worlds can be put in one-to-one correspondence then the two worlds are of equal size. For an alternative conception of the size of infinite sets, see Benci and Di Nasso [29].

[100]This is known as the 'ultrapower construction' of the hyperreals (Hurd and Loeb [110, 1.1]).

[101]Similarly, hyperreals that are infinitesimally close to a given real number appraoch that real number.

[102]The fact that hyperreal multiplication is well-defined may appear to make them an attractive choice for subjective ethical principles, where we need to multiply the value of the outcome of an act in a world state by an agent's credence that she is in that world state. At the end of this section, however, I will show that we can just as easily construct subjective versions of principles like Ordered Overtaking.

[103]Bostrom uses the hyperreal representation of the sum of the sequence rather than the sequence itself

representations to rank worlds: world w_1 is strictly better than another world w_2 if the hyperreal representation of w_1 is strictly larger than the hyperreal representation of w_2, and world w_1 is equally as good as world w_2 if their hyperreal representations are equal.

Crucial to this approach, then, is how we define 'larger than' and 'equal to' for hyperreal numbers. Hyperreal numbers h_1 and h_2 are 'equal to' on another if and only if their components are the same at some 'large' set of locations (which we will define in a moment). The hyperreal h_1 is larger than the hyperreal h_2 if and only if the components of h_1 are strictly greater than the components of h_2 at a 'large' set of locations.

What does it mean for a set of locations to be 'large'? I will summarize the account given by Arntzenius [6]. First, we can think of each hyperreal as a sequence with a first entry, second entry, third entry and so on, so that each component of the hyperreal is associated with an integer. For all hyperreal numbers, we can first say that no finite number of components is large and that for any given set of components either that set or its compliment must be large. In order to distinguish 'large' sets beyond this, we must use a non-principal ultrafilter. A non-principal ultrafilter F is a set of the set of integers such that (i) if two sets of integers are in F then their intersection is in F, (ii) if a set is in F then so are its subsets, (iii) for any set, either that set or its complement is in F, and (iv) there is no set X such that F is the set of all A such that $X \subseteq A$ (i.e. F does not contain a least element).[104] Finally, can say that a set of integers is 'large' if that set is in our ultrafilter F.[105]

As Arntzenius points out, however, this means that the definition of 'larger than' is radically dependent on our choice of ultrafilter. He demonstrates this with the following example:

because hyprreals are of the same size if they differ at only finitely many locations. So the hyperreal $< 2, 3, 2, 3, 2, 3, ... >$ is not larger than the hyperreal $< 3, 3, 2, 3, 2, 3, ... >$. Therefore if we were to compare worlds using the hyperreal representation of their sequences and not their sums, we would violate Sensitivity.

[104] If an ultrafilter contains a least element then it is called a principal ultrafilter. All ultrafilters on finite sets are principal. For an overview see Jech [113, Ch. 7].

[105] Using the axiom of choice, one can show that all infinite sets have a non-principal ultrafilter. This is known as the 'ultrafilter lemma'. See theorem 7.5 in Jech [113].

> Consider a world with hyperreal utility $< 1, -2, 3, -4, 5, -6, ... >$. How big or small is this? Well, suppose that according to the ultrafilter in question the odd numbered locations are a large set. Then the hyperreal $< 1, -2, 3, -4, 5, -6, ... >$ is the same hyperreal as $< 1, 2, 3, 4, 5, 6, ... >= \alpha$, since they agree on a large set of locations. On the other hand suppose that according to our chosen ultrafilter the even numbered locations are a large set. Then $< 1, -2, 3, -4, 5, -6, ... >$ is the same hyperreal as $< -1, -2, -3, -4, -5, -6, ... >= -\alpha$. That is to say, the hyperreal utility of a given world can depend wildly on the choice of ultrafilter. And, either the evens or the odds have to be a large set, but there clearly is no unique correct choice here. So we have a problem. (p. 51, Arntzenius, 2014)

Our choice of ultrafilter is arbitrary, and yet in this case an arbitrary choice will determine whether a world is as good as the infinitely good world $< 1, 2, 3, 4, 5, 6, ... >$ or as good as the infinitely bad world $< 1, -2, 3, -4, 5, -6, ... >$. This is an unacceptable result.

A further problem with the hyperreal approach that Arntzenius [6, p. 49] points out is that hyperreals are highly order-dependent. If, according to our ultrafilter F, the set of odd numbered locations is large, then the hyperreal $< 3, 1, 3, 1, 3, ... >$ is not equal to the hyperreal $< 1, 3, 1, 3, 1, ... >$ even though the latter is just a permutation of the former.[106] This obviously poses a problem if basic locations of value come in no natural order.[107]

An alternative to representing infinite worlds as hyperreal numbers is to represent them using John Conway's surreal numbers [60]. Surreal numbers are constructed in stages using $x = \{L|R\}$ notation. Here x is a surreal number and L and R are sets of numbers, where L contains every value strictly less than N and R contains every value strictly greater than

[106]This is not merely an artifact of our choice of ultrafilter. There is no ultrafilter that can treat all possible permutations of all uncountable sequences of real-numbers as equal.

[107]Artnzenius points out that the order-dependence of the hyperreals is worse than it is for other order-dependent theories, since it also depends on how we carve up locations. For example, suppose that the seconds of world w_1 have utilities $(1, 3, 1, 3, ...)$. Then the hyperreal of temporal locations 2 seconds in length is $< 4, 8, 12, 16, ... >$ while the hyperreal of temporal locations 3 seconds in length is $< 5, 12, 17, 24, ... >$ and the latter is strictly larger than the former. Arnzenius also notes that hyperreals are not countably additive, which results in a 'countable Dutch book' problem (see p. 50, Arntzenius, 2014). A further argument against the application of hyperreals and non-constructive objects is given in Litak [157].

x. Following Conway's conventions [60, p. 4], if $x = \{L|R\}$ then we use x^L to refer to the typical member of L and x^R to refer to the typical member of R, and we use $\{x^L|x^R\}$ to refer to x itself.

We construct all numbers sequentially using the following construction rule and definitions of greater than or equal to ($\geq$), equality ($=$), and strictly greater than ($>$):

Construction: If L and R are sets of numbers and no member L is less than or equal to any member of R, then there is a number $\{L|R\}$

Definition ($\geq$)**:** $x = \{x^L|x^R\} \leq y = \{y^L|y^R\}$ iff there is no x^R such that $x^R \leq y$ and there is no y^L such that $x \leq y^L$

Definition ($=$)**:** $x = y$ iff $x \geq y$ and $y \geq x$

Definition ($>$)**:** $x > y$ iff $x \geq y$ and $y \ngeq x$

On day one no numbers have been defined and so the first object we define must use the empty set as its left set and it's right set, written $\{|\}$. We call this number 0. On the second day we use the number 0 to define $1 = \{0|\}$ and $-1 = \{|0\}$.[108] On the third day we define $2 = \{1|\}$, $\frac{1}{2} = \{0|1\}$, $-2 = \{|-1\}$, and $-\frac{1}{2} = \{-1|0\}$, and so on.[109] On the first countably infinite day, ω, we define the first infinite ordinals $-\omega = \{|-1, -2, -3, ...\}$ and $\omega = \{1, 2, 3, ...|\}$ as well as the first infinitesimals $\frac{1}{\omega} = \{0|1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, ...\}$ and $-\frac{1}{\omega} = \{-1, -\frac{1}{2}, -\frac{1}{4}, -\frac{1}{8}, ...|0\}$ and all of the real numbers e, π, and so on (p. 12-13, *ibid.*). In this way we can construct a totally ordered class that contains all real numbers, infinite numbers, and infinitesimal numbers. Conway ([60, p. 11]) represents when the first 'few' numbers are born as follows:

[108] The definition of a negative number $-x$ is $\{-x^L|-x^R\}$ ([60, p. 5]).

[109] We generally remove members of L and R that have no bearing on the value of the number in question, i.e. $2 = \{1|\} = \{-1, 0, 1|\}$ and $-2 = \{|-1\} = \{|-1, 0, 1\}$.

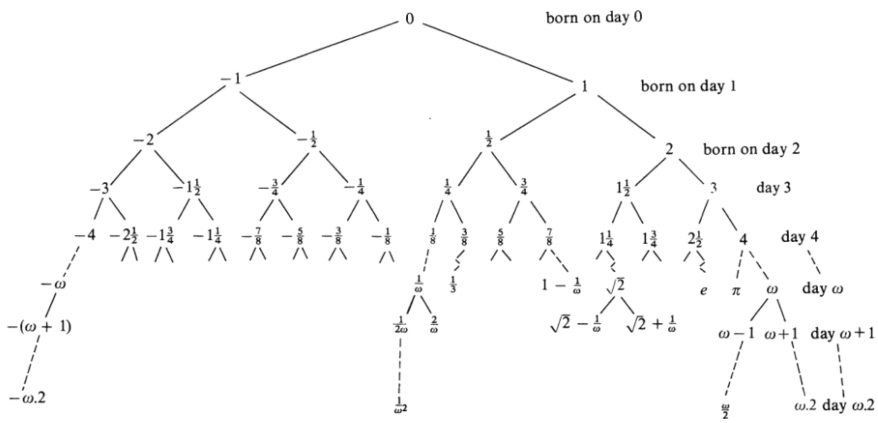


Figure 18: The surreal tree

One of the useful things about the surreal numbers is that the standard operations of addition, subtraction, multiplication, and division are all well-defined over all surreals and behave in the transfinite case much as they do in the finite case. For example, x is strictly less than $x + 1$ for all surreals, including infinite surreals.[110] The surreals therefore form an ordered field. In fact, they form the largest possible ordered field: the rationals, reals, hyperreals, and so on can all be embedded as subfields of the surreals.

We might hope that surreal representations of infinite worlds could be highly useful in infinite ethics. The surreals form an ordered field and therefore satisfy the completeness axiom (for all surreal numbers x, y, $x \leq y$ or $y \leq x$). If we could associate each infinite world with a unique surreal representation, we could simply order worlds by their surreal representations.

The problem with this is that surreals are insufficiently fine-grained for this to result in an ordering of worlds that is consistent with Sensitivity. Suppose that the basic locations of value in w_1 come in a natural order $(1, 1, 2, 4, 8, ...)$ and that the surreal representation of a world is just the surreal representation of the sum of this sequence, $\{1, 2, 4, 8, 16, ...|\}$. Compare this with a world w_2 that contains the same basic locations of value with utility levels

[110] For the definitions of surreal addition and multiplication, see Conway [60, p. 5].

$(0,1,1,1,1,...)$ and its representation $\{1,2,4,8,16,...|\}$. World w_1 is strictly better than world w_2 by Sensitivity. As Conway [60, p. 12] notes, however, both $\{1,2,4,8,16,...|\}$ and $\{1,2,4,8,16,...|\}$ are just different forms of the surreal number ω. Since $\{1,2,4,8,16,...|\} = \{1,2,4,8,16,...|\} = \omega$, any theory that identifies infinite worlds with their surreal representation and ranks infinite worlds using these surreal representations will violate Sensitivity.

An alternative strategy is to formulate existing principles for comparing infinite worlds within the framework of the surreals. For example, Chen and Rubio [54] have recently proposed using the surreal field in infinite decision theory with the decision rule from Colyvan's 'Relative Expectation Theory'.[111] One can easily imagine a similar proposal being made within infinite ethics. Appealing to surreals may offer modest extensions of principles in infinite ethics. However, using the surreals does not help us to avoid the problems described in this chapter. For example, if the basic locations of value come in no natural order then the 'surreal approach' must be supplemented with some further construct, such as limits (discussed above) or a numerosity function.[112] If it is not supplemented in this way, it cannot compare world pairs that are not already comparable by weaker utility difference principles like Catching-Up and Overtaking principles. The surreal approach therefore inherits the problems associated with whatever supplementation is used.

It is important that we do not unthinkingly default to one formal model of infinities when alternatives are available. If we do so, we run the risk of trying to solve problems that appear to arise from the phenomena we are exploring, but that disappear once a richer formal framework has been introduced. The problems of infinite ethics do not seem to be

[111]This rule is a subjective analogue of an order-invariant version of the Overtaking principle discussed above.

[112]Numerosities are measures for the size of infinite sets developed by Benci, Di Nasso, and Forti [32] that satisfies certain properties of finite cardinals that are not satisfied by infinite cardinals if we equate size with one-to-one mappings. In particular, the numerosity of a set is strictly greater than the numerosity of a proper subset of that set. Benci and Di Nasso prove that 'a numerosity function exists if and only if there exists a selective ultrafilter' [29, 51]. Since the existence of selective ultrafilters is independent of ZFC, the existence of a numerosity function is also independent of ZFC.

mere phantoms of an impoverished formal model, however, since they do not disappear if we appeal to hyperreal or surreal representations of infinite worlds.

1.5.4 Discounting Principles

The final set of principles I will describe here are discounting principles. According to the discounting approach to infinite ethics, the value of a world with the sum of the utility of its basic locations (agents, times, etc.), discounted by some discount rate r that gives less weight to basic locations that are more distant from the location of the ethical decision maker.

Koopmans [127] argues that five axioms characterize the discount utilitarian rule: Continuity, Sensitivity, Independence, Stationarity, and a Sensitivity axiom. We have met some of these axioms above, but can summarize them informally here. Continuity states that a sufficiently small difference in the utility stream should not result in a large difference in the rankings of those streams.[113] Koopmans' Sensitivity axiom (not to be confused with the Sensitivity principles given above) states that you can make a stream worse by changing the utility of the first period in the stream. Independence says that the parts of two utility streams that are identical should not change the ranking of those utility streams.[114] This is a weaker version of an important property called 'strong separability'. The relation $\succcurlyeq$ is strongly separable if and only if the ranking of any subset of the locations of two worlds is independent of the utilities at the remaining locations of those worlds.[115] Stationarity states that if two

[113] As we noted above (54), the definition of continuity relies on a distance measure that is, in infinite spaces, relative to our se lection of distance measure. This distance measure induces a topology on the infinite space in question. See Lauwers [134] for further discussion of this and a defense of the sup and product topologies.

[114] Koopmans (p. 292, 1960) calls this 'Postulate 3 (3a and 3b)' but it is now commonly referred to as Indeprendence. We can summarize this postulate informally but in more detail. First, let $b_1, b_2, b_3, ...$ and $d_1, d_2, d_3, ...$ be streams of utilities at infinitely many locations and let a and c both be utility levels at a single location. (3a) states that if $(a, b_1, b_2, b_3, ..) \succcurlyeq (c, b_1, b_2, b_3, ...)$ then $(a, d_1, d_2, d_3, ...) \succcurlyeq (c, d_1, d_2, d_3, ...)$. (3b) states that if $(a, b_1, b_2, b_3, ..) \succcurlyeq (a, d_1, d_2, d_3, ...)$ then $(c, b_1, b_2, b_3, ..) \succcurlyeq (c, d_1, d_2, d_3, ...)$. This is sometimes called anti-complementarity (two factors are complementary if the value of one affects the value of the other).

[115] For a more precise treatment of Separability, see Broome [45, Ch. 4].

streams have the same utility at the first location, then the first stream is better than the second if and only if the first stream after the first location is better than the second stream after the first location.[116] This means that 'postponing' some utility gain should make no difference to our preference over streams: i.e. if $(3,1,1,1,...) \succcurlyeq (2,1,1,1,...)$ then $(1,1,1,3,1,1,1,...) \succcurlyeq (1,1,1,2,1,1,1,...)$. Finally, the Sensitivity axiom – Monotonicity – says that if all locations in one infinite stream are at least as good as all locations in a second infinite stream, then the first infinite stream is at least as good as the second. This is equivalent to the $\succcurlyeq$ condition of Sensitivity.

Let the locations of w_1 be ordered temporally and let the discount rate α be in the open interval $(0,1)$. Koopmans [127] shows that if the at least as good as relation $\succcurlyeq$ satisfies the five axioms above, then it can be represented by the following function:[117]

$$d : \sigma^*(w_1) = (l_1, l_2, ..., l_i, ...) \longmapsto (1-\alpha)\sum_{i-1}^{\infty} \alpha^{i-1} u(l_i)$$

The discount utilitarian rule gives less weight to more temporally distant locations. For example, if we suppose that $\sigma^*(w_1) = (1,1,1,..)$ and that α is a steep discount rate of 0.5 then, since $0.5^0(1) + 0.5^1(1) + 0.5^2(1) + 0.5^3(1) = 1 + 0.5 + 0.25 + 0.125, ...$ and the sum of this sequence is 2, $d : \sigma^*(w_1) \mapsto 0.5 \cdot 2 = 1$. As long as the utility levels of $\alpha^{i-1}u(l_i)$ do not diverge, d will return a real-numbered value of world w_1.[118]

One problem with Koopman's discount utilitarian rule in particular is that, as Chichilnisky [56] shows, the rule entails 'Dictatorship of the Present': if w_1 is better than w_2 then there is some finite i such that what happens at locations after l_i makes no difference to the ranking

[116] In other words, $(a, b_1, b_2, b_3, ..) \succcurlyeq (c, d_1, d_2, d_3, ...)$ if and only if $(b_1, b_2, b_3, ...) \succcurlyeq (d_1, d_2, d_3, ...)$.

[117] I model this on the formulation given in Lauwers [146, p. 236].

[118] The utility levels will diverge if they grow faster than the discount rate. For example, if our discount rate is 0.5 and our utility stream is $(1, 2, 4, 8, ...)$ then d will diverge to positive infinity. See Dolmas [69] for a discussion of utility streams that grow without bound.

of w_1 and w_2. It seems implausible that we should ignore the utility at locations after some finite time has passed, but this is entailed by the discount utilitarian rule.[119] Consider the following illustrative example given by Lauwers:[120]

> Consider an economy in which trees are a necessary input to production or consumption. The dynamics of tree reproduction are as follows. If n out of $2n$ subsequent generations cut the forest at a maximal rate, the species become extinct after the $2n$'th generation, in which case there is zero utility at every period from then on. Assume this strategy results in utility streams of the form $u^n = (0.1, 0.1, \ldots, 0.1; 1, 1, \ldots, 1; 0, 0, \ldots)$ with the first (resp. last) 1 at the $n + 1$ (resp. $2n$)'th place, in which generations $n + 1, \ldots, 2n$ cut at a full capacity and exhaust the forest. When the consumption of the forest is delayed and n becomes larger, the forest slightly expands and more generations can bene- fit. Alternatively, generations can invest in the forest and only cut at an equilibrium rate which allows the forest to survive. This strategy results in the utility stream $u^\infty = (0.25, 0.25, \ldots, 0.25, \ldots)$ in which each generation reaches the same utility level. Optimization with respect to a discounted utilitarian rule leads to the elimination of the forest. [146, 238]

In a case like this, discount utilitarianism says that we should prefer some finite number of nearby generations to receive greater utility, even if they can only do so at the expense of infinitely many future generations. This degree of 'impatience' seems ethically unjustifiable.

More generally, all discounting approaches to infinite ethics are committed to the claim that we should give more ethical weight to the wellbeing of agents that are temporally closer to us simply because they are not temporally close to us.[121] This violates a core ethical

[119] In illustrative example of this is given in Lauwers [146, 237-8]. Chichilnisky proposes an alternative discount rule that does not entail Dictatorship of the Present but, as Lauwers notes, Chichilnisky's rule violates Stationarity. In essence, we must choose whether to ignore the tail of the infinite stream (Koopmans' discount utilitarianism) or the head of the infinite stream (Chichilnisky's sustainable preferences).

[120] This kind of examples shows that the discount utilitarian rule violates the principle that Asheim [8, p. 215] calls 'Hammond equity for the future', which states that we should not be willing to sacrifice an infinite gain in the future for some finite gain in the present.

[121] This kind of 'pure temporal discounting' can be contrasted with merely instrumental temporal discounting. The latter discounts future consumption for instrumental reason.

commitment to treating all agents' utility as being of equal moral worth regardless of when it occurs.[122] Ethicists have generally found any violation of neutrality to be unacceptable, as have many economists.[123] Ramsey called temporal discounting 'a practice which is ethically indefensible and arises merely from the weakness of the imagination' [188, p. 541], while Sidgwick notes that 'the time at which a man exists cannot affect the value of his happiness from a universal point of view' [208, p. 414]. The primary reasons for adopting temporal discounting in cases that involve infinite populations is that the alternative is perceived to be too demanding: we could require that present generations sacrifice n units of wellbeing so that better off generations in the future can experience m units of wellbeing, where $m \gg n$.[124] As Cowen and Parfit [62] note, however, this is not a reason to discount utility to *future* agents. Rather, it is only a reason to discount utility to *better off* agents.

This failing of the discount approach to infinite ethics is even worse than has been generally acknowledged. Those who adopt the discounting approach have typically assumed that agents or generations are given less weight the more *temporally* distant from the ethical decision-maker but, as Bostrom [42] points out, if the world is temporally infinite then the discounting approach must discount across both time *and* space. In other words, the wellbeing of agents decreases in value the further away they are from us in space. Cowan and Parfit use this kind of spatial discounting view as a reductio for temporal discounting:

> Remoteness in time roughly correlates with a whole range of morally important facts. So does remoteness in space. Those to whom we have the greatest obligations, our own family, often live with us In the same building. We often live close to those to whom we have other special obligations, such as our clients, pupils, or patients. Most of our fellow citizens live closer to us than most aliens. But no one suggests that, because there are such correlations, we should adopt

[122] The discounting approach fails to treat agents neutrally regardless of whether agents or times are taken to the basic locations of value, since the utility at a given time is experienced by agents.

[123] See note 37 for references.

[124] See Rawls [189, p. 297].

a spatial discount rate. No one thinks that we would be morally justified if we cared less about the long-range effects of our acts, at some rate of n percent per yard. The temporal discount rate is, we believe, as little justified. [62, p. 159]

Spatial discounting produces highly counterintuitive results. For example, if we spatially discount utility then we can make the world better by clustering happy agents closer to us or by moving unhappy agents further away from us without changing their utility levels.[125] This seems absurd. Therefore even though the discounting approach allows for greater comparability between infinite worlds than some of the other approaches mentioned thus far, it is extremely difficult to motivate this view on independent grounds.[126]

Summary

According to some of the most successful theories in modern cosmology, we live in an infinite world: a world that contains infinitely many agents whose lives may be of moral value. There are various principles that we want to satisfy when we rank infinite worlds ethically: Sensitivity, Equity, and completeness. None of the solutions that I considered – utility difference principles, utility density principles, alternative formalizations of infinity, and discounting principles – can satisfy all of these desiderata. In the next chapter I will defend Pareto, the agent-based Sensitivity principle, and an equity principle that is distinct from Anonymity. I show that these axioms, in combination with a Permutation Principle, jointly entail that any transitive ethical 'at least as good as' relation must be incomplete.

[125] Temporal discounting must be relative to a moving present: i.e. the utility of agents becomes less valuable the further they exist from us into both the future and the past relative to our present temporal location. If we do not assume a moving present view then, even if we adopt a relatively modest temporal discount rate, one of the most important actions that we could focus our energies on is to attempt to invent some kind of time travel device, which seems absurd. Thanks to Carl Shulman for pointing this out.

[126] I have not offered a comprehensive survey of all discounting views here. For recent examples of discounting views, see Jonsson and Voorneveld's [115] 'Limit Discount Utilitarianism'.

Chapter 2

Pareto, the $\succcurlyeq$ Relation, and the Permutation Principle

In this chapter I will outline several core principles and show that they cannot be jointly satisfied in infinite worlds. In section 2.1 I show that the 'Expansionist' view in infinite ethics delivers intuitively correct verdicts about the ranking of many infinite world pairs. I then show that this view violates the agent-based Pareto principle introduced in the previous chapter. I argue that we should accept the Pareto principle and that we must therefore reject Expansionism. In section 2.2 I defend two further axioms. First, I argue that the 'at least as good as' relation ($\succcurlyeq$) is transitive. Second, that it is a 'necessary qualitative relation'. In section 2.3 I argue that for any world pair we can find a qualitatively identical world pair that contains entirely different agents. I call this the 'Permutation Principle' and compare my axioms to the Anonymity principles introduced in the previous chapter. I show that these principles jointly entail that certain world pairs are ethically incomparable.

2.1 Pareto and Expansionism

In this section I will consider three different pairs of infinite worlds in which one world seems clearly better than the other before offering an example of a theory that can vindicate these intuitions. The world pairs are divided into three types: *Archipelagos* are worlds that are spatially but not temporally infinite, *Infinity Houses* are worlds that are temporally but not spatially infinite, and *Cubelands* are worlds that are both spatially and temporally infinite.

Case 1: Archipelagos

Archipelagos are worlds that contain infinitely many islands on an infinite sea. Four agents live on each of the islands. All of them were born at exactly the same time, and they will all live for exactly fifty years before dying. You have been offered a chance to decide what the islands of an Archipelago are like. You can decide that the islands will be full of insects and their weather will consist mostly of hurricanes and rain. If you do, then only one person on each island will lead a happy life: the rest will all be miserable. This Archipelago is called 'Stormy'. Alternatively, you can decide that each island will be full of fruit and its weather will always be clement. If you do, then an entirely different population of agents will exist on the islands of the Archipelago and only one person on each island will be miserable: the rest will all be happy. This Archipelago is called 'Clement'.

Let's suppose that there are only two possible happiness levels for the agents of an Archipelago: happy and miserable. I will use smiley faces ☻ to represent happy agents and sad faces ☹ to represent miserable agents. The agents on the islands of Clement and Stormy are as follows:

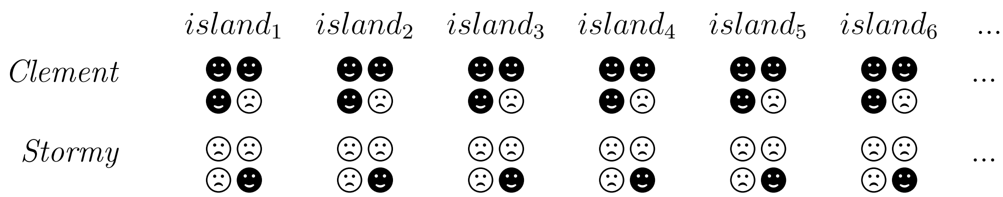


Figure 19: Clement and Stormy

Every island of Stormy is made up of mostly miserable agents while every island of Clement is made up of mostly happy agents. Given this, it seems like you should bring about Clement rather than Stormy in this case. Indeed, bringing about Stormy when you can choose Clement instead strikes many as morally impermissible.[127]

Case 2: Infinity House

Infinity Houses are worlds that contain only a single house in which one agent is born. This agent lives in the house for fifty years and then they die. When they die they are replaced by another agent who lives in the house for fifty years and then dies, and so on ad infinitum. You have the choice about whether the world contains a house that is a mansion or a shack. If the house is a mansion then it will contain two generations that are happy followed by one generation of a sad agent. Call this world 'Mansion'. If the house is a shack then it will contain an entirely different set of agents consisting of two generations that are sad followed by one generation that is happy. If we use t_i to represent a fifty-year period, we can depict the worlds Mansion and Shack as follows:

[127] One might worry that our intuition that Archipelago_1 is better than Stormy is simply based on a poor understanding of infinities. For example, one could object that any two worlds that both contain infinitely many happy agents and infinitely many miserable agents must be equally good because they contain the same amount of happy and miserable agents. But this kind of view overlooks important differences between these worlds and, as we will see, there are theories that produce verdicts in line with our intuitions in these cases. The defenders of these theories do not simply misunderstand infinities.

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	...
Mansion	☻	☻	☹	☻	☻	☹	☻	☻	☹	...
Shack	☹	☹	☻	☹	☹	☻	☹	☹	☻	...

Figure 20: Mansion and Shack

In this case, it seems that we are morally required to bring about Mansion rather than Shack.

Case 3: Cubeland

Cubelands are worlds that contain infinitely many agents that each live forever. The spend their whole lives within a 10m by 10m cube and these cubes are spread uniformly across space. You have a choice to either bring about a world in which all agents are optimists by default, but in which one agent is born a pessimist. This pessimist manages to convince each agent agent adjacent to her along the x, y and z axes from her location l_0 to become pessimists too. These pessimists then convert a single agent adjacent to them along these axes at a rate of one conversion per year. Call this world 'Optimistic' since the agents are optimistic by default. Alternatively, you can bring about a world that contains an entirely different set of agents exists. These agents are all pessimists by default but there is one agent that is born an optimist. This optimist manage to convince each agent in a straight line along the x, y and z axes from her location l_0 to become optimists too at a rate of one conversion per year. Call this world 'Pessimistic' since the agents are pessimistic by default.

If we assume that all optimistic agents are happy and that all pessimistic agents are sad, we can represent Optimistic Cubeland and Pessimistic Cubeland as follows:

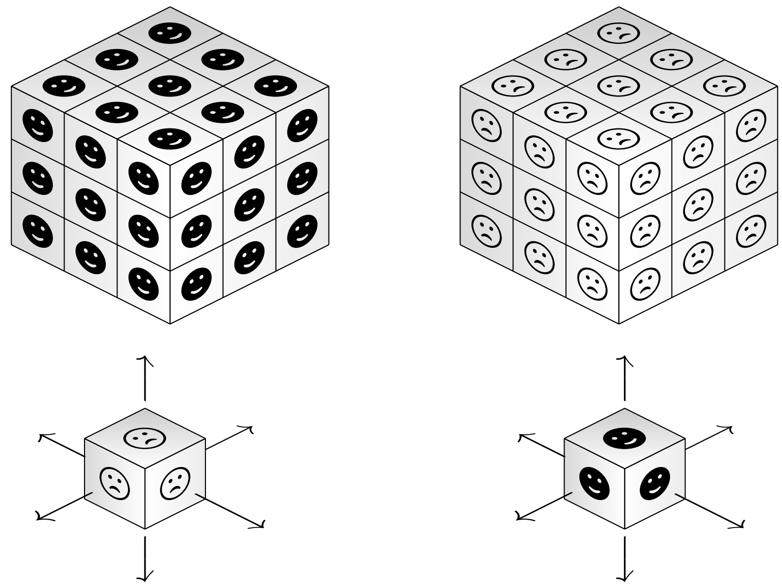


Figure 21: Optimistic Cubeland and Pessimistic Cubeland

So if t_i is an interval of a year then at t_1 there is just one pessimistic agent in Optimistic Cubeland and at t_2 there are 7 pessimistic agents in Optimistic Cubeland and at t_3 there are 13 pessimistic agents in Optimistic Cubeland, and so on. We can represent the cubed directly adjacent to location l_0 at every time after t_1 by first showing the cube itself and then each of the three segments that comprise this cube:

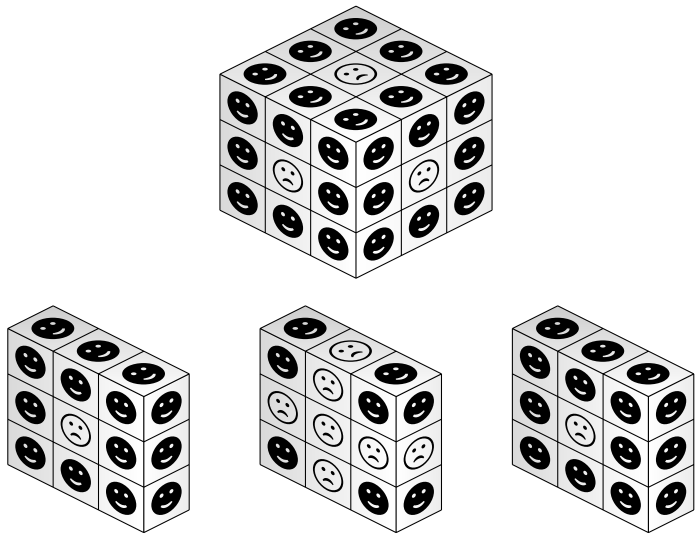


Figure 22: Segments of Optimistic Cubeland

In this case, it seems that we are required to bring about Optimistic Cubeland rather than Pessimistic Cubeland, since the sad agents in Optimistic Cubeland make up a smaller and smaller proportion of the agents we will encounter as we evaluate locations more distant from l_0. By contrast, the happy agents in Pessimistic Cubeland make up a smaller and smaller proportion of the agents we will encounter as we evaluate locations more distant from l_0.

So the Clement Archipelago seems better than the Stormy Archipelago, the Infinite Mansion seems better than the Infinite Shack, and the Optimistic Cubeland seems better than the Pessimistic Cubeland in the three cases above.

The most notable theory that vindicates all three of these intuitions is the 'Expansionist' theory defended initially by [233] and more recently by Arntzenius [6]. The key concept we need to state this theory is that of an *allowable expansion* from a given spatiotemporal location l.[128] An allowable expansion is a sequence of spatiotemporal regions r_1, r_2, r_3,... in which (i) every spatiotemporal region in the sequence is finite in size, (ii) each region

[128] In the cases I consider, the starting point for the expansion can be arbitrary.

is contained in the one that came before it: so r_1 is a proper subregion of r_2, which is a proper subregion of r_3 and so on, and (iii) the expansion expands at a uniform rate across spacetime. We can follow Arntzenius's (pg. 39, 2014) suggested interpretation of clause (iii): to get a *spatially* uniform expansion of r we consider a band of width d from the edge of region r (or location l if it is the first region in the sequence). Region r and this band are the next region in the sequence. To get a *temporally* uniform expansion in a non-relativistic setting, we consider a temporal length t from r, such as one minute. The minute before and after r are then the next temporal region in the sequence. So in a non-relativistic setting, a uniform temporal expansion is one that moves as fast into the future as it does into the past.[129] The union of these regions is the whole of spacetime.

By way of illustration, consider an allowable expansion r_1, r_2, r_3,... of our two Archipelagos, where d is the distance from the center of a single island to its perimeter:

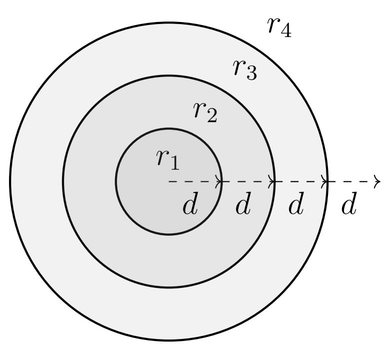


Figure 23: An allowable expansion

Assume that the sea between each island is of negligible size. This means that within region r_1 of the Clement Archipelago there is a single island containing three happy agents and one sad agent, and within region r_1 of Stormy Archipelago there is a single island containing three sad agents and one happy agent. There will be more miserable agents in each subsequent

[129]Arntzenius extends the Expansionist theory to both relativistic worlds. For simplicity I will assume that we can use the non-relativistic account of a temporally uniform expansion in the cases given above.

region of the Stormy Archipelago than there are in the Clement Archipelago:

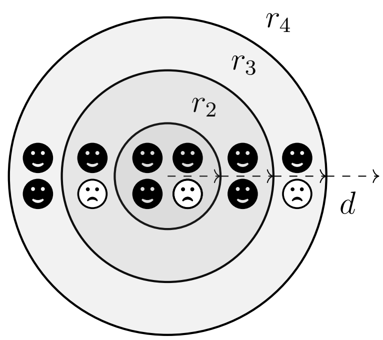


Figure 24: Allowable expansions of Clement and Stormy

Therefore the following Expansionist principle can explain and generalize our intuition that the Clement Archipelago is better than the Stormy Archipalgo:[130]

Expansionism

($\succcurlyeq$) *If w_1 and w_2 share the same spacetime with the same metric, and for all allowable expansions r_1, r_2, r_3, ... in this metric there exists an integer n such that for every $k > n$, the total utility in r_k at w_1 is greater than or equal to the total utility in r_k at w_2 then w_1 is at least as good as w_2*

($\succ$) *And if there exists an integer n such that for every $k > n$, the total utility in r_k at w_1 is strictly greater than the total utility in r_k at w_2 then w_1 is strictly better than w_2*

Since the Clement Archipelago contains more happy agents than the Stormy Archipelago does from the very first region r_1 in each world and, in all allowable expansions of region r_1, the Clement Archipelago is better than the Stormy Archipelago according to Expansionism.[131] Expansionism therefore delivers the result that is in line with our intuitions in the case of the Clement and Stormy Archipelagos.

[130]This principle is closely modeled on the principle defended by Arntzenius (p. 53, 2014). A key difference is that Arntzenius's principle ranks actions rather than worlds and replaces utility with expected utility.

[131]In this case this will be true even if we select different starting points in each world and even if the regions were of different sized in each world, but in general we need to identify regions across worlds using a counterpart relation (p. 53-5, Arntzenius, 2014).

Expansionism also delivers the intuitively correct result in the Mansion and Shack case. Suppose that our temporal distance is at least as great as a single temporal period t. Then for any t_i in the temporal sequence that we start our expansion, there is strictly greater happiness in Mansion than in Shack at every time from t_{i+1} at the latest.[132]

Finally, in the case of Optimistic and Pessimistic Cubeland suppose, for simplicity, that the spatiotemporal region expands at a spatial rate of one cube in every direction and at a temporal rate of one year from location l_0.[133] Let's suppose we can represent the utility of happy agents with 1 and the utility of sad agents with 0. If we start our expansion from location l_0 (the location containing the atypically optimistic or pessimistic agent) then the utility of each new spatiotemporal band in Optimistic Cubeland is $(0, 20, 51, 106, 191, ...)$. Meanwhile,the utility of each new spatiotemporal band in Optimistic Cubeland is $(1, 7, 13, 19, 25, ...)$. So the difference in total of utility at each region of these worlds is as follows:[134]

	r_1	r_2	r_3	r_4	r_5	...
Optimistic (O)	0	20	71	177	368	...
Pessimistic (P)	1	8	21	40	65	...
$u(O_{r_i}) - u(P_{r_i})$	-1	12	50	137	303	...
$\sum u(O_{r_i}) - u(P_{r_i})$	-1	11	61	198	501	...

Figure 25: Utility difference of Cubeland regions

Since the total utility of Optimistic Cubeland is strictly greater than the total utility of Pessimistic Cubeland at every region after region r_1, and since there will exist such a region under all allowable expansions, Optimistic Cubeland is strictly better than Pessimistic

[132]If the expansion starts at a location containing a sad agent in Mansion and a happy agent in Shack, the first region r_1 may not be such that r_1 of Mansion contains strictly more happiness than r_1 of Shack.

[133]This is an idealization: a regional expansion growing a constant rate will generally not be cube shaped.

[134]Since regional expansions include the region before them, these bands are r_1, r_2-r_1, r_3-r_2-r_1, and so on. The happiness at each region r_1, r_2, r_3 and so in is the sum of the sequence of happiness at these bands.

Cubeland according to Expansionism.[135]

Expansionism is reminiscent of the Ordered Overtaking principle from the last time. Unlike that principle, however, Expansionism applies to spatiotemporal regions and not just to ordered temporal sequences. This can make an important difference to how it ranks world pairs. Consider Cain's Sphere of Suffering case from the previous section. Suppose that the number of agents that fall within the regions of our expansion – $r_1, r_2, r_3, \ldots$ – correspond with the squares of odd numbers: $(1, 9, 25, 49, 64, \ldots)$, and that our the expansion expands at uniform rate in time, e.g. each region moves one year into the future.[136] Let's also suppose that the sphere of suffering starts with the agent in r_1 and takes two years to envelop the 8 agents around the central agent, two more years to envelop the 16 agents around those agents, and so on. In other words, it can be represented using the following square of suffering:

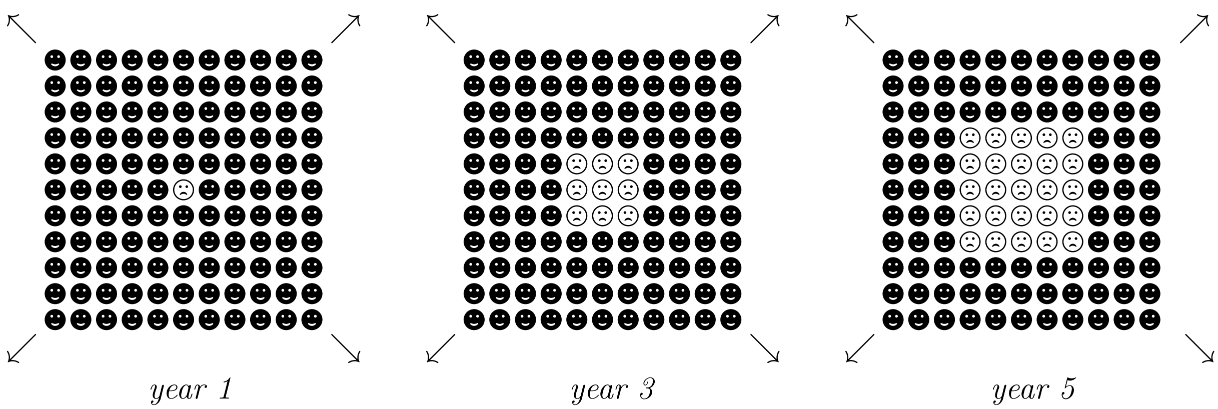


Figure 26: Squares of suffering across time

This means that spatiotemporal region r_1 of the Sphere of Suffering world contains a single agent who is sad for one year. Region r_2 contains the original sad agent plus eight happy

[135]Note that in this case it doesn't matter how positive utility happy agents compared with the negative utility that sad agents receive, and that if we started at any region other than l_0 then there would strictly more utility in Optimistic Cubeland than in Pessimistic Cubeland from region r_1 onwards.

[136]Since this world has a temporal starting point we don't need to worry about the temporal expansion moving into the past.

agents for one year (since the sphere of suffering has not reached them yet). Region r_3 contains the original 9 agents who are all now sad (since the sphere of suffering has now reached them) plus 16 happy agents for one year, and so on. If we again assume that happy agents have utility 1 and sad agents have utility 0, we can represent the cardinality and the utility of the agents in this allowable expansion as follows:

	y_1	y_2	y_3	y_4	y_5	y_6	...
$\|a\|$ in r	1	9	25	49	81	121	...
$u(a)$ in r	0	8	16	40	54	94	...

Figure 27: Utilities in the regions of the sphere of suffering

By contrast, the spatiotemporal region r_1 of the Sphere of Happiness world contains a single agent who is happy for one year. Region r_2 the original happy agent plus 8 sad agents for one year (since the sphere of happiness has not reached them yet). Region r_3 contains the original 9 agents who are all now happy (since the sphere of happiness has now reached them) plus 16 sad agents for one year, and so on, represented as follows:

	y_1	y_2	y_3	y_4	y_5	y_6	...
$\|a\|$ in r	1	9	25	49	81	121	...
$u(a)$ in r	1	1	9	9	27	27	...

Figure 28: Utilities in the regions of the sphere of happiness

The difference in total of utility at each region of these worlds is as follows:

	r_1	r_2	r_3	r_4	r_5	r_6	...
Sphere of Suffering (S)	0	8	16	40	54	94	...
Sphere of Happiness (H)	1	1	9	9	27	27	...
$u(S_{r_i}) - u(H_{r_i})$	-1	7	7	31	27	67	...
$\sum u(S_{r_i}) - u(H_{r_i})$	-1	6	13	44	71	138	...

Figure 29: Utility differences between sphere of suffering and sphere of happiness

Therefore the total utility in the Sphere of Suffering world is greater than the total utility in the Sphere of Happiness in every region from r_2 onwards. But now consider an allowable expansion that expands at the same spatial rate as the previous expansion every two years instead of every year.[137] Then the amount of utility in each region of the expansion will correspond with the amount of utility in the spheres, since both the expansion and the spheres are 'expanding' at the same rate across spacetime:

	r_1	r_2	r_3	r_4	r_5	r_6	...
Sphere of Suffering (S)	0	0	0	0	0	0	...
Sphere of Happiness (H)	1	9	25	49	81	121	...
$u(S_{r_i}) - u(H_{r_i})$	-1	-9	-25	-49	-81	-121	...
$\sum u(S_{r_i}) - u(H_{r_i})$	-1	-10	-35	-84	-165	-286	...

Figure 30: Utility differences between the spheres under a different allowable expansion

So for some allowable expansions that expand at a faster rate than the spheres do, there is a region r_k such that the Sphere of Suffering world has more utility than the Sphere of Happiness world at every region after r_k. But for some allowable expansions that expand at a slower rate than the spheres do, there is a region r_k such that the Sphere of Happiness

[137]Both of the expansions I have mentioned are allowable since an allowable expansion is just one that 'at each time expands at the same rate in each direction in space, and at each location in space expands at the same rate in each direction of time' (p. 53, Arntzenius, 2014).

world has more utility than the Sphere of Suffering world at every region after r_k. Therefore Expansionism does not say that either world is better than or equal to the other. This at least seems more plausible than the time-first Sensitivity principle or a similar 'space-first' Sensitivity principle, which both entail that the Sphere of Suffering world is better than the Sphere of Happiness world, contrary to the agent-based Sensitivity principle.[138]

My principal objection to Expansionism is that, as Arntzenius himself notes (pg. 55-6, 2014), it is inconsistent with the widely accepted Pareto principle.[139] The Pareto principle is the agent-based Sensitivity principle introduced in the previous chapter:

Pareto

($\succcurlyeq$) *If w_1 and w_2 contain the same agents and every agent has at least as much utility in w_1 as they do in w_2, then w_1 is at least as good as w_2.*

($\succ$) *If w_1 and w_2 contain the same agents and each agent has at least as much utility at w_1 as they do at w_2 and some agent has strictly greater utility in w_1 than they do in w_2, then w_1 is strictly better than w_2.*

Notice that Pareto can only rank worlds that contain exactly the same agents. Going forward, I will distinguish between world pairs with *identical populations*, world pairs with *disjoint populations*, and world pairs with *overlapping populations*, each defined as follows:

Identical Populations

A pair of worlds w_1 and w_2 have identical populations if and only if every agent in w_1 is also in w_2 and every agent in w_2 is also in w_1

[138] Artnzenius (pg. 41, 2014) argues that if whether one world dominates the other depends on the speed of the expansion then it is not so implausible to conclude that the utility of one world is not greater or smaller than the other.

[139] Pareto entails what Arntzenius (p. 35-6, 2014) calls 'Permutation of Times' and 'Dominance'. And so it is inconsistent with the 'Permutation of People' principle that is entailed by Expansionism. Although Expansionism does not deliver the same verdict as Pareto in the Sphere of Suffering case above, it does not actually say that a world that is strictly worse by Pareto is better than a world that that is strictly better by Pareto, which is a more troubling result.

Disjoint Populations

A pair of worlds w_1 and w_2 have disjoint populations if and only if there is no agent in w_1 that is also in w_2 (i.e. w_1 and w_2 contain no shared agents)

Overlapping Populations[140]

A pair of worlds w_1 and w_2 have overlapping populations if and only if there is at least one agent in w_1 that is also in w_2 and there is at least one agent that (i) is in w_1 but not w_2 or (ii) is in w_2 but not w_1

We can show that Expansionism is inconsistent with Pareto by considering two Archipelago worlds that are geographically identical to Clement and Stormy in structure but that contain exactly the same agents (so that they can be compared by Pareto). I will call these worlds 'Balmy' and 'Blustery'. To depict these Archipelagos it will be helpful to group the agents of world pairs into sets of agents, X, Y, Z, and so on. In the diagrams below I will use 🅧 to represent a happy agent in set X and Ⓧ to represent a miserable agent in set X. And if two worlds contain all and only agents in the same sets then those worlds contain identical populations. We can now depict both the happiness levels and identities of the agents that live on the islands of Balmy and Blustery as follows:

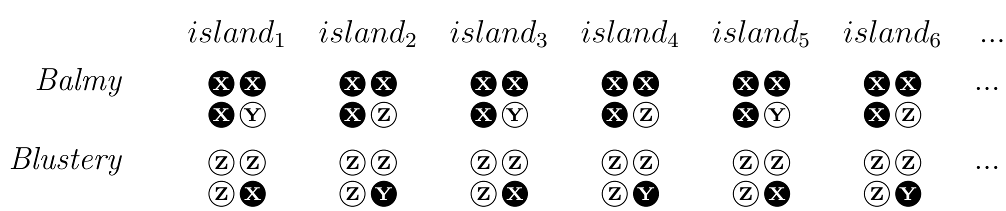


Figure 31: Balmy and Blustery

Balmy is clearly better than Blustery according to Expansionism, since these worlds have the same spatiotemporal distribution of happy and sad agents that Clement and Stormy do. But, unlike Clement and Stormy, Balmy and Blustery are Pareto comparable. In Balmy

[140] Normally we would want to treat identical population worlds as a special case of overlapping worlds, but it will be useful to keep the two kinds of populations distinct here.

the agents in set X are all happy but the agents in Y and agents in Z are all miserable. In Blustery the agents in X and agents in Y are all happy and the agents in Z are all miserable:

	X	Y	Z
Balmy	☻	☹	☹
Blustery	☻	☻	☹

Figure 32: Agent happiness in Balmy and Blustery

Balmy is better than Blustery by Expansionism, but we can see that Blustery is better than Balmy by Pareto. I take this as a convincing refutation of Expansionism. Ethics is concerned with the wellbeing of people. It is not particularly concerned with the pattern of utility across spacetime. The fact that the happy agents of Balmy come in one particular spatiotemporal order rather than another is not a good reason to think that Balmy is better than Blustery when we can see that infinitely many agents in Y are better off in Blustery than in Balmy and no agent is better off in Balmy than in Blustery.

Although the Pareto principle has generally been taken to be one of the most fundamental axioms in ethics, the fact that it conflicts with principles like Expansionism in infinite ethics have caused some to question whether we really need to retain it.[141] There are, however, relatively few independent arguments against Pareto. And in order to weaken or reject the Pareto principle in favor of some other axiom, we must try to establish that, despite its prima facie plausibility, Pareto is less plausible than at least one of these axioms.

As I noted in the previous chapter, Hamkins and Montero [99] offer an argument against the Weak Pareto principle, which is entailed by Pareto. In one example, they ask us to consider two agents, Dan and Daniela, who both spend an eternity traveling from hell to heaven. The sequence of daily utility that Dan and Daniela experience is identical:

[141]For example, Van Liedkerke and Lauwers (p. 169-170, 1997) propose a possible weakening of Pareto that is consistent with their utility density principle. It is quite rare for people to reject or weaken the principle on independent grounds. One exception are those who accept principles like Maximin, which is inconsistent with the $\succ$ condition of Pareto, but I believe this is a strong reason to reject such principles.

$(..., -3, -2, -1, 0, 1, 2, 3, ...)$. Hamkins and Montero argue that we should therefore be indifferent between whether we were to go through life as Dan or Daniela. We are asked to suppose that Dan crosses the zero point one day before Daniela. Where t_i is one day, we can represent this as follows:

	t_{-3}	t_{-2}	t_{-1}	t_0	t_1	t_2	t_3	...
Dan	-3	-2	-1	0	1	2	3	...
Daniela	-4	-3	-2	-1	0	1	2	...

Figure 33: Utility streams of Dan and Daniela

Hamkins and Montero argue that a Paretian principle would require that we prefer Dan's life to Daniela's because Dan's life is one day ahead of Daniela's and so on any given day Dan's life is better than Daniela's. They argue that since we should be indifferent between which of these two lives we live, we should reject Paretian principles in infinite cases.

Of course, the Pareto principle formulated above does not entail that Dan's life is better than Daniela's, since it is a principle that compared worlds and not lives. Hamkins and Montero respond to this by arguing that 'if we assume that one life is better than another just in case the world containing only that life is better than the world containing only the other life, as it seems reasonable to suppose, then this distinction evaporates.' (p. 234-5, 2000). But it is only true that a world containing Dan's life is better than a world containing Daniela's life if we adopt the time-based Sensitivity principle outlined in the previous chapter.[142] The Pareto principle does not entail that a world containing Dan is better than a world containing

[142] What about Expansionism? Expansionism does not commit to time-based Sensitivity (it commits to Sensitivity about allowable spatiotemporal expansions). However, if we think that the counterpart relation between worlds is such that t_0 in Dan's world must correspond with t_0 in Daniela's world, then Expansionism entails that the world containing Dan is better than the world containing Daniela. Of course, if we think that this need not be the relevant counterpart relation then Expansionism could entail that these two worlds are equally good (if the relevant counterpart relation maps t_0 in Dan's world to t_{-1}) in Daniela's or that neither is better than the other (if there are multiple admissible counterpart relations).

Daniela since neither life has strictly more utility than the other.[143]

A very different objection to Pareto comes from those who believe that even if we care about the utility being experienced by agents and not about the distribution of utility across spacetime, agents and facts about agent identities should be irrelevant to the ethical ranking of worlds. On this view, which was mentioned in Chapter 1, subjective experiences are the basic locations of value, and both the identity of the agent experiencing them and the spatiotemporal location at which they occur are ethically irrelevant.

The subjective experience view is inconsistent with Pareto. If w_1 is better than w_2 according to Pareto then although w_1 will not be worse than w_2 according to the subjective experience view, since improving agents' wellbeing cannot *reduce* the utility of subjective experiences, it may not be better either. We can make all agents in an infinite world better without changing the sum of good and bad experiences happening in the world. Therefore, in whatever form agents exist, it seems likely that the subjective experience view will fail to satisfy Pareto with respect to these agents. The key problem for the subjective experiences view, which I will discuss in chapter 5, is that it will be able to strictly rank almost no infinite worlds.

In this section I have argued that Expansionism conflicts with Pareto and that we therefore ought to reject Expansionism, despite its initial plausibility. Indeed, we should be inclined to reject any ethical axioms that violate Pareto. Of course, the falsity of the general Expansionism principle does not by itself undermine the judgments that Clement is better than Stormy, Mansion is better than Shack, and Optimistic Cubeland is better than Pessimistic Cubeland. For example, we might think that if two worlds contain disjoint populations then we can use a restricted version of Expansionism to compare them. Or we might think that these world pairs can be ranked by some principle other than Expansionism. But in the next

[143] I will generally assume that the utility at a life is bounded. Extending these results to worlds that include unbounded utilities and uncountably many agents will be the subject of future work.

section I show that if we accept Pareto and a plausible account of $\succcurlyeq$, then we must conclude that none of these three world pairs are ethically comparable.

2.2 The $\succcurlyeq$ Relation

In this section I am going to argue that, along with Pareto, we should accept two claims about the ethical at least as good as' ($\succcurlyeq$) relation: that it is transitive and that it is a 'necessary qualitative relation'. I show that these principles entail that all three of the world pairs discussed in the previous section are ethically incomparable: contrary to our initial intuitions, neither world is better than or equally as good as the other.

We can begin with the claim that $\succcurlyeq$ is a transitive relation, since this is the least controversial of the two. Transitivity is defined as follows:

Transitivity of $\succcurlyeq$ *If* $w_1 \succcurlyeq w_2$ *and* $w_2 \succcurlyeq w_3$ *then* $w_1 \succcurlyeq w_3$

Transitivity prevents $\succcurlyeq$ from producing cyclic rankings of worlds, where w_1 is better than w_2 and w_2 is better than w_3 and w_3 is better than w_1. Although transitivity was once an entirely uncontroversial principle, it has come under some criticism in recent decades.[144] It is still widely accepted as a property of the $\succcurlyeq$ relation, however. I will therefore assume transitivity here and will discuss objections to this principle in Chapter 5.

As is standard, I define 'strictly better than' as follows (I take this to be uncontroversial):

Strictly Better Than $w_1 \succ w_2 \equiv w_1 \succcurlyeq w_2$ *and* $w_2 \not\succcurlyeq w_1$

Note that Pareto entails that $\succcurlyeq$ is reflexive: for all worlds w_1, $w_1 \succcurlyeq w_1$. From the reflexivity of $\succcurlyeq$ and this definition of strictly better than, it follows that the strict better than is

[144] See Rachels [187] and Temkin [226, Ch. 7-8]. For a critique of Temkin, see Voorhoeve [240].

irreflexive: for all worlds w_1, $w_1 \not\succ w_1$.[145] In addition, transitivity and this definition of $\succ$ entail that if w_1 is strictly better than w_2 and w_2 is at least as good as w_3 then w_1 is strictly better than w_3 (if $w_1 \succ w_2$ and $w_2 \succcurlyeq w_3$ then $w_1 \succ w_3$).[146]

The second claim I will make, in addition to transitivity, is that $\succcurlyeq$ is a *necessary qualitative relation*. To say that a property or relation is qualitative is to say that it involves no specific individuals. For example, the property 'loving Obama' and the binary relation 'being closer to Obama than' are both non-qualitative since they involve the particular Obama. By contrast, the property of 'being happy' and the relation of 'being closer to some politician than' are both qualitative relations since they involve no particulars. This distinction extends to properties of and relations among possible worlds. For example, the property of 'being a world in which Obama is happy', and the relation 'Obama is happier at w_1 than at w_2' are non-qualitative since they involve the particular Obama but the property of 'being a world in which every person is happy', and the relation 'everyone who exists at w_1 also exists at w_2 and is happier at w_2 than at w_1' are qualitative relations since they involve no particulars.

To say that the $\succcurlyeq$ relation is to say that if the relation holds between two objects then it holds necessarily. The numerical identity relation is a typical example of a necessary relation: if x and y are numerically identical then there is no possible world in which x and y are non-identical. We can contrast this with a contingent relation like 'is to the left of'. An object x may be to the left of an object y but it could have been the case that object x was to the right of object y and so this relation is not a necessary relation. If one world is at least as good as another then it seems that this cannot be only contingently true. If it's true that w_1 is at least as good as w_2 then even though a different world – a world like w_1 but different in some small respect – may not be at least as good as w_2, it does not seem

[145] If $w_1 \succ w_1$ then $w_1 \not\succcurlyeq w_1$ but by reflexivity $w_1 \succcurlyeq w_1$.

[146] Since $w_1 \succ w_2$ and $w_2 \succcurlyeq w_3$ it follows that $w_1 \succcurlyeq w_3$ by transitivity. Suppose $w_3 \succcurlyeq w_1$. Then, $w_2 \succcurlyeq w_3$ and $w_3 \succcurlyeq w_1$ and so $w_2 \succcurlyeq w_1$ by transitivity. But since $w_1 \succ w_2$ it follows from the definition of strictly better than that $w_2 \not\succcurlyeq w_1$. Therefore $w_3 \not\succcurlyeq w_1$. Since $w_1 \succcurlyeq w_3$ and $w_3 \not\succcurlyeq w_1$, it follows that $w_1 \succ w_3$.

possible that w_1 could have not been at least as good as w_2. I will discuss the domain of the $\succcurlyeq$ relation in Chapter 5.

One world pair is a 'qualitative duplicate' of the other if and only if it has all of the same qualitative properties and relations as the original world pair. If a relation R between a and b is a necessary qualitative relation then whenever the pair $\langle a, b \rangle$ is a qualitative duplicate of the pair $\langle c, d \rangle$, Rab if and only if Rcd. We can use this concept of a qualitative duplicate of a world pair to formulate the qualitativeness of $\succcurlyeq$ axiom as follows:

Qualitativenesss of $\succcurlyeq$ *If the pair $\langle w_3, w_3 \rangle$ is a qualitative duplicate of the pair $\langle w_1, w_2 \rangle$, then $w_3 \succcurlyeq w_4$ if and only if $w_1 \succcurlyeq w_2$.*

The claim that $\succcurlyeq$ is a necessary qualitative relation should strike us as highly plausible. The alternative would be to allow that non-qualitative facts can make a difference to the ethical ranking of two worlds. If this were true then I could describe every qualitative fact about world w_1 and w_2: how many agents are in each world, how happy they are, how rich their lives are, and so on. But I could not always tell you whether w_1 is at least as good as w_2 if I don't also know who the particular agents in w_1 and w_2 are: for example, that the president in both worlds is Obama and not merely some agent that is qualitatively identical to Obama. This seems wrong. Whether one world is ethically at least as good as another is not the sort of thing that should depend on the identity of its agents.[147]

The claim that $\succcurlyeq$ is a qualitative relation should not be confused with the claim that $\succcurlyeq$ is a qualitative *internal* relation. If relation R is a qualitative internal relation then if Rab and c is a qualitative duplicate of a and d is a qualitative duplicate of b then Rcd. Not all necessary qualitative relations are qualitative internal relations. For example, the relation 'being adjacent to' is qualitative but not qualitative internal: a world could contain objects a, b, c such that a is adjacent to b, and b is a qualitative duplicate of c, but c is on the

[147]Relatedly, Hare (pg. 156-164, 2013) argues benevolence should not be sensitive to haecceities.

other side of the universe from b and thus not adjacent to a. Similarly, the relation 'every person who exists at w_1 also exists at w_2' is a necessary qualitative relation but it is not a qualitative internal relation: this relation could hold between w_1 and w_2 while failing to hold between w_1 and a qualitative duplicate w_2 at which a different collection of people exist.[148]

The Pareto principle is inconsistent with the claim that $\succcurlyeq$ is a qualitative internal relation. We can see this by considering the world Balmy from the previous section and a qualitative duplicate of Balmy, aptly named 'Duplicate Balmy', which is as follows:

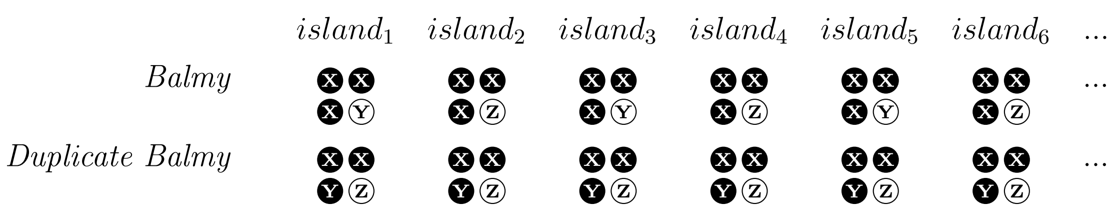


Figure 34: Balmy and Duplicate Balmy

Balmy is at least as good as itself by the Pareto principle. Duplicate Balmy is a qualitative duplicate of Balmy: it has all of the same qualitative properties as Balmy, including – as we can see above – the distribution of happy and sad agents on each island. So if $\succcurlyeq$ is a qualitative internal relation then it follows that Balmy is at least as good as Duplicate Balmy. But Pareto entails that Duplicate Balmy is strictly better than Balmy, since all agents in Balmy are at least as happy in Duplicate Balmy, but infinitely many agents in set Y are strictly happier in Duplicate Balmy than they are in Balmy. Therefore Pareto is inconsistent with the claim that $\succcurlyeq$ is a qualitative internal relation.

This brings us to Hamkins and Montero's [98] second argument against Paretian principles. They argue that we should reject Pareto because it conflicts with what they call the 'Isomorphism Principle', which says any world is as good as an isomorphic copy of itself. An 'isomorphic copy' of a world is a copy that is preserves 'the topological structure of locations

[148] I adopt a roughly Lewisian view of internal and external relations. See Lewis [152, p.62].

and the amount of local goodness at those locations' of a world (p. 235, *ibid.*). I disagree with this objection to Pareto. Given transitivity, the Isomorphism Principle entails that $\succcurlyeq$ (and hence $\succ$ also) is a qualitative internal relation. To show this, suppose that w_1 is at least as good as w_2 and w_3 is a qualitative duplicate of w_2. Then w_3 is an isomorphic copy of w_2 so by the Isomorphism Principle, w_3 is equally as good as w_2 and therefore w_2 is at least as good as w_3. So by transitivity, w_1 is at least as good as w_3. Similarly, if w_4 is a qualitative duplicate of w_1 then w_4 is an isomorphic copy of w_1 and so is equally as good as w_1 by the Isomorphism Principle, and so w_4 is at least as good as w_2 by transitivity.

Although I believe it is highly plausible that $\succcurlyeq$ is a necessary qualitative relation for the reasons given above, I see no reason to believe that it is a qualitative internal relation. The very fact that the claim that $\succcurlyeq$ is a qualitative internal relation conflicts with Pareto in cases like the one above gives us strong reasons to reject the claim that $\succcurlyeq$ is a qualitative internal relation. After all, it seems plausible that if infinitely many agents are better off and no agents are worse off in one world than in the other, then the two worlds are not equally good.[149] The claim that $\succcurlyeq$ is a necessary necessary qualitative relation but not a qualitative internal relation is consistent with Pareto but also does not commit us to the implausible view that whether one world is better than another can depend on non-necessary necessary qualitative relations at or between those worlds.

Pareto is not consistent with the claim that $\succcurlyeq$ is a qualitative internal relation. But Pareto *is* consistent with the claim that $\succcurlyeq$ is a necessary qualitative relation: the claim that if one world pair $\langle w_1, w_2 \rangle$ is a qualitative duplicate of another world pair $\langle w_3, w_4 \rangle$, then $w_1 \succcurlyeq w_2$ if and only if $w_3 \succcurlyeq w_4$. In the case above, Balmy is at least as good as itself by Pareto. Let us

[149] An additional argument against the claim that better than is a qualitative internal relation comes from Hare (p. 184-6, 2013). Suppose in world w_1 each agent has a neighbor that has more utility than she does. Consider a sequence of worlds in which each agent becomes more qualitatively like her original neighbor. Eventually we will reach a world in which each agent is qualitatively identical to her neighbor. This world is qualitatively identical to the original world but is strictly better than w_1 by Pareto and transitivity.

use $w_1 \sim w_2$ to mean $w_1 \succcurlyeq w_2$ and $w_2 \succcurlyeq w_1$. It follows that Balmy $\sim$ Balmy. Therefore if $\succcurlyeq$ is a qualitative relation then the $\sim$ relation must hold between any qualitative duplicate of the pair ⟨Balmy, Balmy⟩. But the pair ⟨Balmy, Duplicate Balmy⟩ is not a qualitative duplicate of the pair ⟨Balmy, Balmy⟩. For example, the necessary qualitative relation 'has no shared agents that are worse off in' holds between Balmy and itself but does not hold between Balmy and Duplicate Balmy because they have disjoint populations, and so the claim that $\succcurlyeq$ is a necessary qualitative relation does not entail that Balmy is at least as good as Duplicate Balmy.[150] I will therefore assume that $\succcurlyeq$ is a necessary qualitative relation but not that it is a qualitative internal relation.

I will call worlds w_1 and w_2 'ethically incomparable' (or just 'incomparable') if w_1 is not at least as good as w_2 and w_2 is not at least as good as w_1. Therefore if world w_1 is not better than or equal to or worse than world w_2 then the two worlds are incomparable.[151] I am going to demonstrate that if $\succcurlyeq$ is transitive and qualitative, then Clement and Stormy are ethically incomparable. To show this, suppose that we group the agents of Clement into sets A, B, and C, and we group the agents of Stormy into sets D, E, and F, where the agents in each of these sets come in the following patterns:

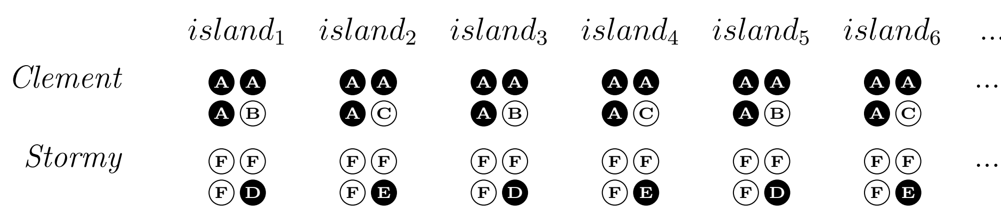


Figure 35: Sets of agents in Clement and Stormy

Set A comprises agents a_1, a_2, a_3 on island$_1$, agents a_4, a_5, a_6 on island$_2$, and so on. Set B

[150]The claim that $\succcurlyeq$ is a necessary qualitative relation *does* entail that the 'at least as good as' relation holds between ⟨Duplicate Balmy, Duplicate Balmy⟩, but this is consistent with Pareto.

[151]Those who accept the fourth relation 'on a par' (Chang, 2002) may wish to contend that Archipelago$_1$ and Stormy could be on a par rather than incomparable. But this case has none of the usual features of parity such as different evaluative considerations and so I won't discuss the parity relation here.

comprises agent b_1 on island$_1$, agent b_2 on island $_3$, agent b_3 on island$_5$, and so on. Suppose for reductio that Clement is at least as good as Stormy. Consider a world pair ⟨ Duplicate Clement, Duplicate Stormy⟩ that is a qualitative duplicate of the pair ⟨Clement, Stormy⟩:

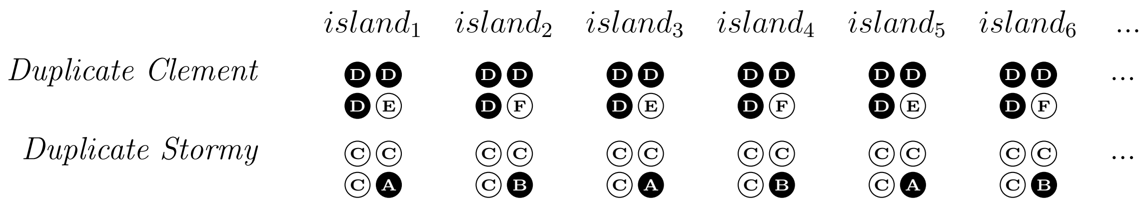


Figure 36: Sets of agents in Duplicate Clement and Duplicate Stormy

All of the qualitative roles that agents play in ⟨Clement, Stormy⟩ are played by different agents in ⟨Duplicate Clement, Duplicate Stormy⟩. The qualitative roles played by the agents in set A in Clement – agents $a_1, a_2, a_3, ...$ – are played by agents from set D in Duplicate Clement – agents d_1, d_2, d_3, ... (these are the very same agents in set A that exist in Clement and the very same agents in set D that exist in Stormy).

The agents of ⟨Clement, Stormy⟩ have been permuted. A permutation g of the population of the pair ⟨Clement, Stormy⟩ is a bijection of the joint populations of Clement and Stormy onto itself. Consider the set of numbers $\{1, 3, 5\}$ and the set of numbers $\{2, 4, 6\}$. A permutation of the union of these sets – like a permutation of the population of a world pair rather than a permutation of the populations of the individual worlds – is a bijection from the union of these sets onto itself. We can use Cauchy's two-line notation, which lists the original items on the top line and their image on the second line below, to illustrate a permutation h:

$$h(\{1,3,5\} \cup \{2,4,6\}) = \begin{pmatrix} 1 & 3 & 5 & 2 & 4 & 6 \\ 3 & 6 & 4 & 1 & 5 & 2 \end{pmatrix}$$

If the agents of ⟨Clement, Stormy⟩ have been permuted by some g then if a_1 plays a given qualitative role in the pair ⟨Clement, Stormy⟩ then an agent $g(a_1)$ plays that qualitative role in ⟨Duplicate Clement, Duplicate Stormy⟩. In this case, we can construct the permutation g of the population of ⟨Clement, Stormy⟩ onto itself from a bijection f from the population of Clement – denoted $pl(A_1)$ – to the population of Stormy – denoted $pl(A_2)$ – as follows:

$$
\begin{array}{ccc}
pl(A_1) & f & pl(A_2) \\
a_1 \bullet & \longrightarrow & \bullet\ d_1 \\
a_2 \bullet & \longrightarrow & \bullet\ d_2 \\
\vdots & & \vdots \\
b_1 \bullet & \longrightarrow & \bullet\ e_1 \\
b_2 \bullet & \longrightarrow & \bullet\ e_2 \\
\vdots & & \vdots \\
c_1 \bullet & \longrightarrow & \bullet\ f_1 \\
c_2 \bullet & \longrightarrow & \bullet\ f_2 \\
\vdots & & \vdots
\end{array}
$$

Duplicate Stormy contains all of the same agents as Clement, but each agent $a_i, b_i; c_i$ in Duplicate Stormy plays the role that the agent $f(a_i), f(b_i); f(c_i)$ plays at Stormy respectively. Duplicate Clement contains all of the same agents as Stormy, but each agent $d_i, e_i; f_i$ in Duplicate Clement plays the role that the agent $f^{-1}(d_i), f^{-1}(e_i); f^{-1}(f_i)$ plays in Clement respectively. Let g be a bijection from the population of ⟨Clement, Stormy⟩ onto itself such that for all agents that exist in Clement, $g(pl(A_1)) = f(pl(A_1))$ and for all agents that exist in Stormy, $g(pl(A_2)) = f^{-1}(pl(A_2))$. The world pair ⟨Duplicate Clement, Duplicate Stormy⟩ is the result of permuting the agents of ⟨Clement, Stormy⟩ by g.

Duplicate Clement is a qualitative duplicate of Clement but it has a population that is identical to that of Stormy (the agents in sets D, E and F). Duplicate Stormy is a qualitative duplicate of Stormy but it has a population that is identical to that of Clement (the agents in sets A, B and C). We can see above that the happiness levels of each of the agents in sets A – G in the four Archipelagos are as follows, where – denotes non-existence:

	A	*B*	*C*	*D*	*E*	*G*
Clement	☻	☹	☹	—	—	—
Stormy	—	—	—	☻	☻	☹
Duplicate Clement	—	—	—	☻	☹	☹
Duplicate Stormy	☻	☻	☹	—	—	—

Figure 37: Happiness levels of Clement, Stormy, and duplicates

Since all of the agents in set E are better off in Stormy than in Duplicate Clement and no agent is worse off in Stormy than they are in Duplicate Clement, Stormy is better than Duplicate Clement by Pareto. Since all of the agents in set B are better off in Duplicate Stormy than in Clement and no agent is worse off in Duplicate Stormy than in Clement, Duplicate Stormy is better than Clement by Pareto.

We assumed that Clement is at least as good as Stormy. We have shown that Stormy is strictly better than Duplicate Clement by Pareto. But since $\succcurlyeq$ is a necessary qualitative relation and ⟨Duplicate Clement, Duplicate Stormy⟩ is a qualitative duplicate of ⟨Clement, Stormy⟩, it follows that Duplicate Clement is at least as good as Duplicate Stormy. We have shown that Duplicate Stormy is strictly better than Clement by Pareto. This violates transitivity. To show this, let $w_1 \xrightarrow{Pareto} w_2$ mean that w_1 is strictly better than w_2 by Pareto, let $w_1 \longrightarrow w_2$ mean that w_1 is at least as good as w_2 by hypothesis (hyp.), by the qualitativeness of the 'at least as good as' relation $\succcurlyeq$ (q$\succcurlyeq$.) or by Pareto (Pareto) and let w_1 - - - w_2 mean that w_1 and w_2 are qualitative duplicates. If we use C to denote Clement, S to denote Stormy, DC to denote Duplicate Clement, and DS to denote Duplicate Stormy, the transitivity failure above is as follows:

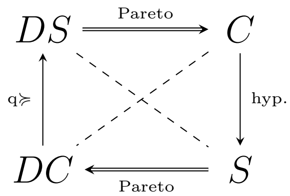


Figure 38: Transitivity violation showing Clement $\not\succcurlyeq$ Stormy

Therefore, if we accept the Pareto principle and that $\succcurlyeq$ is transitive and qualitative, then Clement cannot be at least as good as Stormy.

Note that Clement and Stormy have entirely disjoint populations. This means that we can rule out the possibility, hinted at in the last section, of using some kind of restricted Expansionist principle – for example, an Expansionist principle that only compare worlds that are not Pareto-comparable – to conclude that Clement is better than Stormy. Even though Clement and Stormy are not Pareto comparable, if $\succcurlyeq$ is transitive and qualitative then the claim that Clement is better than Stormy is inconsistent with Pareto.

Of course, in order to show that Clement and Stormy are incomparable – the neither world is better than or equal to the other – we also need to show that Stormy is not at least as good as Clement. We have no reason to believe that Stormy is at least as good as Clement, but we can also show that Stormy $\not\succcurlyeq$ Clement by the same reasoning.

Last time I grouped the agents of Clement and Stormy into the sets $A - G$. This time I will group the same agents of Clement and Stormy into the sets $U - Z$:

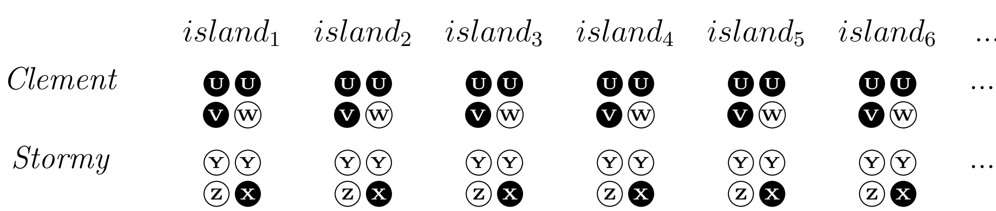


Figure 39: Alternative groupings of the agents of Clement and Stormy

This time, suppose for reductio that Stormy is at least as good as Clement. Consider a world pair ⟨New Duplicate Clement, New Duplicate Stormy⟩ that is a qualitative duplicate of ⟨Clement, Stormy⟩ and whose populations are as follows:

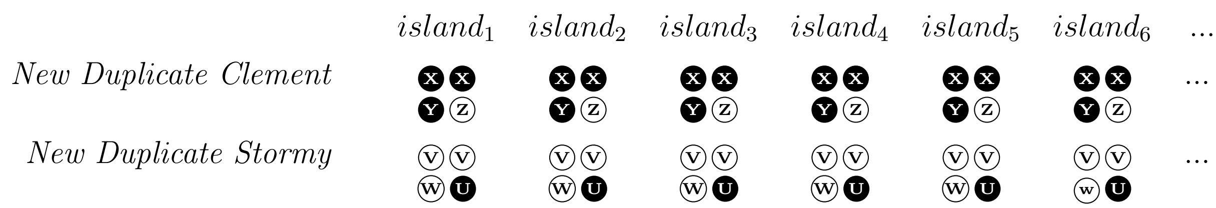
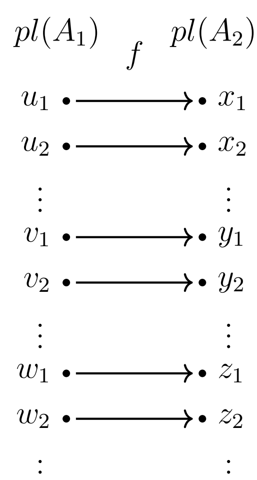


Figure 40: New Duplicate Clement and New Duplicate Stormy

As before, the agents of ⟨Clement, Stormy⟩ have been permuted. This time, the bijection f from the population of Clement to the population of Stormy is:



New Duplicate Stormy contains all of the same agents as Clement, but each agent $u_i, v_i; w_i$ in New Duplicate Stormy plays the role that $f(u_i), f(v_i); f(w_i)$ plays at Stormy respectively. Let g be a bijection from the population of ⟨Clement, Stormy⟩ onto itself such that $g(pl(A_1)) = f(pl(A_1))$ and $g(pl(A_2) = f^{-1}(pl(A_2))$. ⟨New Duplicate Clement, New Duplicate Stormy⟩ is the result of permuting the agents of ⟨Clement, Stormy⟩ by g. The happiness levels of each of the agents in $U - Z$ in these four worlds are as follows:

	U	V	W	X	Y	Z
Clement	☻	☻	☹	—	—	—
Stormy	—	—	—	☻	☹	☹
New Duplicate Clement	—	—	—	☻	☻	☹
New Duplicate Stormy	☻	☹	☹	—	—	—

Figure 41: Happiness levels of Clement, Stormy, and new duplicates

We assumed that Stormy is at least as good as Clement. But we can see that Clement is strictly better than New Duplicate Stormy by Pareto. New Duplicate Stormy is at least as good as New Duplicate Clement by the qualitativeness of $\succcurlyeq$. And New Duplicate Clement is strictly better than Stormy by Pareto. This results in the following transitivity violation:

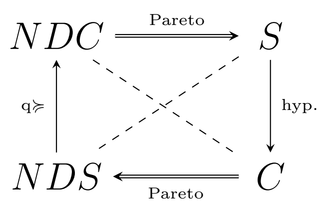


Figure 42: Transitivity violation showing Stormy $\not\succcurlyeq$ Clement

Therefore if we accept the Pareto principle and that $\succcurlyeq$ is transitive and qualitative, then Clement is not at least as good as Stormy and Stormy is not at least as good as Clement. These two worlds are ethically incomparable. It is easy to see that we can find duplicate world pairs that will show that if we accept the Pareto principle and that $\succcurlyeq$ is transitive and qualitative, then Mansion and Shack are incomparable, as are Optimistic Cube and Pessimistic Cube, contrary to our initial intuitions about these three cases.

2.3 The Permutation Principle

In the argument of the previous section I simply assumed that there exist world pairs like ⟨Duplicate Clement, Duplicate Stormy⟩ and ⟨New Duplicate Clement, New Duplicate Stormy⟩ that we can use to show that Clement and Stormy are incomparable. However, one might think that this assumption was questionable. Perhaps we won't always be able to find a pair of worlds $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ such that w_3 contains the same agents as w_2 and w_4 contains the same agents as w_1. In this section I will outline and defend the Permutation Principle. This principle entails that for any world pair ⟨Clement, Stormy⟩, world pairs like ⟨Duplicate Clement, Duplicate Stormy⟩ and ⟨New Duplicate Clement, New Duplicate Stormy⟩ exist.

As I noted above, a permutation of the population of a pair of worlds is a bijection from the population of that world pair to itself. For each agent x in $\langle w_1, w_2 \rangle$, a permutation g maps each agent x in $\langle w_1, w_2 \rangle$ onto some agent $g(x)$ in $\langle w_1, w_2 \rangle$. Suppose that we permute the agents of a world pair $\langle w_1, w_2 \rangle$ by some g such that the qualitative role of every agent x in $\langle w_1, w_2 \rangle$ is played by the agent $g(x)$ in $\langle w_3, w_4 \rangle$. This means that every qualitative property that an agent x has in w_1, $g(x)$ has this property in w_3. And every necessary qualitative relation that agents $x_1, ..., x_n$ have in w_1, agents $g(x_1), ..., g(x_n)$ have this relation in w_3. Therefore the pair $\langle w_3, w_4 \rangle$ is a qualitative duplicate of the pair $\langle w_1, w_2 \rangle$. In such cases I will say that the world pair $\langle w_3, w_4 \rangle$ is a 'qualitative duplicate of $\langle w_1, w_2 \rangle$ under g':

Qualitative Duplication Under a Bijection (world pairs)

A pair of worlds $\langle w_3, w_4 \rangle$ is a qualitative duplicate of a pair of worlds $\langle w_1, w_2 \rangle$ under bijection f iff for any necessary qualitative relation R, if relation $R(x_1, ..., x_n)$ holds at w_1 then relation $R(f(x_1), ..., f(x_n))$ holds at w_3, and if relation $R(x_1, ..., x_n)$ holds at w_2 then relation $R(f(x_1), ..., f(x_n))$ holds at w_4, and if $R(w_1, w_2)$ then $R(w_3, w_4)$

The domain of the bijection f does not have to be restricted to agents. We can permute any object x in w_1 so that in world w_2, the qualitative role that x plays in w_1 – all of the necessary qualitative relations that x has in w_1, denoted by $R(x)$ – are played by $f(x)$, such that $R(x)$ at w_1 and $R(f(x))$ at w_2. Moreover, the necessary qualitative relations preserved by f include all of the qualitative properties of objects at w_1, since these properties are 1-ary necessary qualitative relations.

It is worth emphasizing that permuting the population of a world pair will preserve all necessary qualitative relations that exist both within and *between* the two worlds of the original pair. For example, there will never be an agent a that exists in w_1 but not w_2 such that $g(a)$ exists in both w_3 and w_4. This is because for all agents x, the qualitative role that x plays in w_1 is played by $g(x)$ in w_3 and the qualitative role that x plays in w_2 is played by the same $g(x)$ in w_4. So if x exists in w_1 but not in w_2, then $g(x)$ exists in w_3 but not in w_4.

The claim that I will defend here is that for any world pair and any permutation of its population, we can find a world pair that is a qualitative duplicate of the original world pair under that permutation. I call this the Permutation Principle:[152]

Permutation Principle

For any world pair $\langle w_1, w_2\rangle$ and any bijection g from the population of $\langle w_1, w_2\rangle$ onto any population, there exists a world pair $\langle w_3, w_4\rangle$ that is a qualitative duplicate of $\langle w_1, w_2\rangle$ under bijection g.

The Permutation Principle is clearly true if we believe that an agent can have completely different qualitative properties while still remaining the same agent. For example, if we think

[152]The cases I consider in this chapter will all involve a bijection from the population of $\langle w_1, w_2\rangle$ onto itself, but in the next chapter I will consider bijections from the population of $\langle w_1, w_2\rangle$ onto a distinct equinumerous population. It is worth noting that we can establish a great deal of incomparability even if we assume more restricted versions of this principle, since all we really need is that there are qualitative duplicate worlds for permutations of the right kind. I see no reason to favor such restrictions, however.

that Bill Clinton can have the non-qualitative property 'being Bill Clinton' even if all of his qualitative properties change – he could be a different height and weight, he could have been a gardener instead of president, and so on – then we can clearly replace any agent in a world pair with Bill Clinton without making a qualitative change to that world pair.

It is perhaps worth pointing out that the kind of essentialism that one would need to accept in order to claim that worlds like Clement and Stormy are comparable is of a particularly strange variety. In this case we only permuted the agents so that their image had different spatial locations and different happiness levels. As I will argue in Chapter 5, it seems implausible that many ethically-relevant qualitative properties will be essential to identity.

The claim that we could permute the identities of agents without altering any of the qualitative properties of a world might nonetheless seem too strong, metaphysically speaking. Anti-haecceitists who believe that it is not possible for two worlds to be qualitatively identical but to differ non-qualitatively will clearly reject the Permutation Principle.[153] But even those with less extreme anti-haecceitist leanings may think that the principle is too strong. Suppose we are ancestor essentialists – we believe that certain necessary qualitative relations like 'being the son of' or 'being the granddaughter of' are essential to identity.[154] For example, Bill Clinton could not exist without being the son of William Jefferson Blythe Jr. Now consider a world w_1 that contains only Bill Clinton and William Jefferson Blythe Jr. and another world w_2 that contains only Abraham Lincoln and his father Thomas Lincoln. Consider permutation g of the population of this world pair:

[153] Here I adopt the definition of anti-haecceitism given in Skow (§2, 2008). Note that anti-haecceitists hold that all relations between worlds are qualitative internal relations. So anti-haecceitist will already have rejected my earlier claim that the better than relation is a qualitative but not a qualitative internal relation.

[154] See Lecture III in Kripke [131] for a defense of origin essentialism.

g

Bill Clinton • ⟶ • Thomas Lincoln
William Jefferson Blythe Jr. • ⟶ • Abraham Lincoln
Thomas Lincoln • ⟶ • Bill Clinton
Abraham Lincoln • ⟶ • William Jefferson Blythe Jr.

Figure 43: Permutation of worlds each containing two agents

If we are ancestor essentialists, however, then there is no possible world pair $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g. The permutation g would result in a world w_3 in which Thomas Lincoln is the father of Abraham Lincoln, and a world w_4 in which Bill Clinton is the father of William Jefferson Blythe Jr. But if we are ancestor essentialists then w_3 and w_4 are clearly not metaphysically possible worlds.

My response to this objection is that the world pair that exists and is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g need not be a metaphysically possible world pair. We can appeal to a broader form of logical possibility in which it is possible for agents to have completely different qualitative properties while remaining the same agent. After all, even if it is metaphysically impossible for Bill Clinton to be the father of William Jefferson Blythe Jr., it is not logically impossible for Bill Clinton to be the father of William Jefferson Blythe Jr.

One might worry, however, that any results that rely on a notion of possibility that is narrower than metaphysical possibility will not be the sort of results that we can derive ethical conclusions from. I believe this worry is unfounded. To see why, consider the set of worlds such that for any bijection g from the population of $\langle w_1, w_2 \rangle$ onto itself, there exists metaphysically possible world-pair $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g. Contrast this with the set of worlds such that for any bijection g from the population of $\langle w_1, w_2 \rangle$ onto itself, there exists a logically possible world-pair $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g. The first set might include all of the worlds in which agents lack ancestors while the latter set includes all of the worlds in which agents have ancestors.

Suppose that one world pair in the first set contains two agents that bear no relation to one another: one that was born at t_1 and another that was born at t_2. Suppose that another world pair in the second set contains two agents: one that was born at t_1 that is the mother of the agent born at t_2. The two world pairs are identical in all other respects. If we believe that we cannot derive ethical conclusions from logical possibility then we must conclude that, even if we can show that the first pair of worlds are incomparable, we cannot conclude from this that the second pair of worlds is incomparable. But it seems implausible that the latter pair of worlds can be shielded from this conclusion unless essential relations are also morally relevant. I discuss this more in Chapter 5.

It seems far more plausible to me that the ethical better than relation must hold between worlds that are logically possible even if those worlds are metaphysically impossible.[155] If this is the case then in order to show that w_1 and w_2 are incomparable it is sufficient to show that, if we accept Pareto and that $\succcurlyeq$ is qualitative, then the assumption that w_1 and w_2 are comparable results in a transitivity violation across logically possible world pairs.

The Permutation Principle and the qualitativeness of $\succcurlyeq$ might look superficially similar to the Anonymity principles of the previous section, but they are quite different. The Permutation Principle and the qualitativeness of $\succcurlyeq$ jointly prevent us from using facts about the particulars of a world to produce an ethical ranking of worlds. But the two axioms do not, by themselves, entail even the most basic forms of equity.

We remember from the last chapter that a theory is equitable if it does not prefer one distribution over another if the two distributions are identical but differ only in terms of who has a given utility level. This was supposedly captured by 'Anonymity' principles, which were formulated as follows:

[155]This notion of logical possibility must be narrower than metaphysical possibility but could be broader than'narrow logical possibility' under which any world that is logically consistent is possible.

Anonymity (locations)

If the utility of each basic location of value in w_2 is a (finite/fixed-step/infinite) permutation of the utility of the basic locations of value in w_1, then world w_1 and w_2 are equally good.

The weakest of these equity constraints is Finite Anonymity, which states that if the utility of each basic location of value in w_2 is a finite permutation of the utility of the basic locations of value in w_1, then world w_1 and w_2 are equally good. Consider, for example, the following world pair $\langle w_1, w_2 \rangle$ in which an agent a_1 is born on day 1, an agent a_2 is born on day 2, and so on. World w_2 is identical to w_1, except that in world w_2 the utilities of the first two agents – agents a_1 and a_2 – have been switched:

	a_1	a_2	a_3	a_4	a_5	a_6	...	a_n	a_{n+1}	...
w_1	1	0	1	0	1	0	...	1	0	...
w_2	0	1	1	0	1	0	...	1	0	...

Figure 44: Switching the utilities of two agents

The Finite Anonymity principle entails that w_1 and w_2 are equally good: switching the utilities of agents a_1 and a_2 should not result in one world being ranked over the other. But the Permutation Principle and the qualitativeness of $\succcurlyeq$ are consistent with the claim that w_1 is strictly better than w_2 or vice versa. These axioms jointly require that if w_1 is strictly better than w_2, and $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$, then w_3 is strictly better than w_4. To show that this does not entail Finite Anonymity suppose that the world pair $\langle w_3, w_4 \rangle$ is a qualitative duplicate of world pair $\langle w_1, w_2 \rangle$ under g, where $g(a_1) = a_2, g(a_2) = a_1$ and for all other agents in the population of $\langle w_1, w_2 \rangle$, $g(a) = a$:

	a_1	a_2	a_3	a_4	a_5	a_6	...	a_n	a_{n+1}	...
w_3	0	1	1	0	1	0	...	1	0	...
w_4	1	0	1	0	1	0	...	1	0	...

Figure 45: Permuting the agents whose utilities have been switched

Suppose we believed that w_1 is strictly better than w_2 whenever the agent born on day 1 in w_1 is happier than the agent born on day 1 in w_2. This means that we would rank w_1 strictly better than w_2. But we would also rank w_3 strictly better than w_4. This is consistent with the Permutation Principle and the qualitativeness of $\succcurlyeq$. So the requirement that our ethical rankings are not sensitive to what particular agents exist in a world does not entail equity: it is consistent with a denial of even the weakest anonymity principles.

Interestingly, adding Pareto to these axioms does entail something in the spirit of equity. In the case above, for example, these three axioms jointly entail that worlds $w_1 \not\succ w_2$ and that $w_2 \not\succ w_1$. Suppose that $w_1 \succ w_2$ and so, by the qualitativeness axiom, $w_3 \succ w_4$. We can see that $w_2 \succcurlyeq w_3$ by Pareto and $w_4 \succcurlyeq w_1$ by Pareto. This violates transitivity. Therefore neither of these worlds is strictly better than the other if we accept these three axioms.[156]

We can establish a more general result: namely that Pareto, the qualitativeness of $\succcurlyeq$, and the Permutation Principle entail that if the utility of each agents in w_2 is a finite permutation of the utility of each agent in w_1 then $w_1 \not\succ w_2$ and $w_2 \not\succ w_1$. To show this, suppose that the utility of each agents at w_2 is a finite permutation of the utility of each agent in w_1. This means that there are infinitely many (shared or unshared) agents in w_1 and w_2 with identical utility levels in w_1 and w_2. Let $A = \{a_1, ...a_n\}$ denote this set of agents. And let f be a bijection from each agent in A in w_1 to a unique agent in A in w_2 such that there

[156]Since we reject completeness, the fact that $w_1 \not\succ w_2$ and $w_2 \not\succ w_1$ does not entail that w_1 and w_2 are equally good: we cannot conclude that $w_1 \sim w_2$ from these three axioms alone. In the next chapter I will defend a plausible extension of Pareto that does entail that these two worlds are equally good.

are no agents in A in w_2 that are not the image of some agent in A in w_1 and for all $a \in A$, $u(a_i)$ in $w_1 = u(f(a_i))$ in w_2.

There are also finitely many (shared or unshared) agents in w_1 and w_2 that have distinct utility levels in w_1 and w_2. Let $B = \{b_1, ...b_n\}$ denote this set of agents. Since the utility of each agents at w_2 is a finite permutation of the utility of each agent in w_1, it follows that there exists a bijection h from the agents in B in w_1 to the same set of agents in w_2 such that $u(b_i)$ in $w_1 = u(h(b_i))$ in w_2 for all agents in B. Let the permutation g of the population of $\langle w_1, w_2 \rangle$ onto itself be such that, for all agents in A, $g(a_i) = f(a_i)$ and $g(f(a_i)) = f^{-1}(a_i)$ and for all agents in B, $g(b_i) = h(b_i)$ and $g(h(b_i)) = h^{-1}(b_i)$.

Let $\langle w_3, w_4 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under permutation g. Since g maps all agents in w_1 to an agent in w_2, world w_3 contains the same agents as world w_2 and since g maps all agents in w_2 to an agent in w_1, world w_4 contains the same agents as world w_1. And for all agents x in $\langle w_1, w_2 \rangle$, $u(x) = u(g(x))$. Therefore w_2 and w_3 are equally good by Pareto and w_4 and w_1 are equally good by Pareto. Therefore $w_1 \not\succ w_2$ and $w_2 \not\succ w_1$ by transitivity.

This establishes that Pareto, the qualitativeness of $\succcurlyeq$, and the Permutation Principle entail a weak equity principle: if we can permute the utility levels of finitely many agents in w_1 and this finite permutation results in world w_2, then this finite permutation does not result in a world that is strictly better than or strictly worse than the original world.

What about stronger equity principles involving more than just finite permutations? We can use the same method above to show that if we can perform any finite-length permutation (including a fixed-step permutation) on the utility levels of the agents in w_1 and this finite-length permutation results in world w_2, then world $w_1 \not\succ w_2$ and $w_2 \not\succ w_1$.[157] To show this,

[157] We can use this argument strategy to show that all 'permissible extensions' of Finite Anonymity, as defined in Basu and Mitra [**?**, 24], are such that Pareto plus the qualitativeness of $\succcurlyeq$ entail that permuting the utilities of one world by such a permutation results in a world that is neither strictly better than nor strictly worse than the original.

we can use fixed-step permutations as an example. Suppose that the utilities of w_2 can be derived from the utilities of w_1 by means of a fixed-step permutation. From the definition of fixed-step permutations it follows that there exists a bijection f from the population of w_1 to the population of w_2 such that f maps each agent in w_1 to a unique agent in w_2 and all agents in w_1 are the image of an agent in w_1 under f and for all agents x in w_1, $u(x)$ in $w_1 = u(f(x))$ in w_2. Let the permutation g of the population of $\langle w_1, w_2 \rangle$ onto itself be such that, for all agents x in w_1, $g(x) = f(x)$ and for all agents $f(x)$ in w_2, $g(f(x)) = f^{-1}(x)$. Let $\langle w_3, w_4 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under permutation g. It follows that $w_2 \sim w_3$ by Pareto and $w_4 \sim w_1$ by Pareto. Therefore $w_1 \nsucc w_2$ and $w_2 \nsucc w_1$ by transitivity.

Let us turn, however, to *infinite* permutations. In the previous chapter we considered the Strong Anonymity principle, which says that if the utility of each basic location of value in w_2 is a (possibly infinite) permutation of the utility of finitely many or infinitely many basic location of value in w_1, then world w_1 and w_2 are equally good. It was shown that Pareto is inconsistent with Strong Anonymity using cases like the following, which involve variable step permutations (here using agents rather than times are basic locations of value):

	a_1	a_2	a_3	a_4	a_5	a_6	...	a_n	a_{n+1}	...
w_1	1	0	1	0	1	0	...	1	0	...
w_2	1	1	1	0	1	0	...	1	0	...

Figure 46: An identical population world pair with variable step permuted utilities

As before, we can turn the first utility stream w_1 into the second utility stream w_2 if we permute the utility levels of w_1 by moving utility 1 from a_3 to a_2, utility 1 from a_5 to a_3, utility 1 from a_7 to a_5, and so on, and keep the same utility levels at all other agents agents. But the Pareto principle says that $w_2 \succ w_1$. Since the Permutation Principle permutes the populations of *world pairs* and not of individual worlds, it follows that if $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ then w_4 will be strictly better than w_3 by Pareto. The three axioms I have defended are therefore inconsistent with the Strong Anonymity principle.

These three axioms do, however, entail a new form of equity under infinite permutations. Consider a world pair that is just like the one above, except that the agents of the second world are entirely different from the agents of the first world:

	a_1	a_2	a_3	a_4	a_5	a_6	...	a_n	a_{n+1}	...
w_1	1	0	1	0	1	0	...	1	0	...
	b_1	b_2	b_3	b_4	b_5	b_6	...	b_n	b_{n+1}	...
w_2	1	1	1	0	1	0	...	1	0	...

Figure 47: A disjoint population world pair with variable step permuted utilities

The utility of the agents of world w_2 is an infinite permutation of the utility of the agents of world w_1. Therefore the finite and fixed-step Anonymity principles are both unable to rule out the claim that $w_1 \succ w_2$ or that $w_1 \succ w_2$. But Pareto, the qualitativeness of $\succcurlyeq$, and the Permutation Principle jointly entail that $w_1 \not\succ w_2$ and $w_2 \not\succ w_1$. To show this, consider a bijection f from the population of w_1 to the population of w_2 that maps all agents in w_1 to an agent in w_2 with the same utility level (e.g. $f(a_1) = b_1$, $f(a_2) = b_4$, $f(a_3) = b_2$, $f(a_4) = b_6$, and so on). Let g be a permutation of the population of $\langle w_1, w_2 \rangle$ such that for all agents x, $g(x) = f(x)$ and $g(f(x)) = f^{-1}(x)$. If $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g then $w_2 \sim w_3$ by Pareto and $w_3 \sim w_4$ by Pareto. So $w_1 \not\succ w_2$ and $w_2 \not\succ w_1$ by transitivity.

The difference between this case and the previous one is that in the case above there exists an infinite permutation of the agents of the *pair* of worlds $\langle w_1, w_2 \rangle$ that results in a qualitative duplicate pair $\langle w_3, w_4 \rangle$ such that $w_2 \sim w_3$ and $w_3 \sim w_4$. This is no such infinite permutation in the previous case, nor in any cases in which one world is better than another by Pareto.

I believe that the standard formulations of the Anonymity principles, including Strong Anonymity, rest on the same confusion that underlies Hamkins and Montero's second objection to Pareto: namely, they assume that the $\succcurlyeq$ relation is equitable only if it is a necessary qualitative relation, which seems correct, but they assume that if the $\succcurlyeq$ relation is a necessary qualitative relation then it must be a qualitative internal relation, which is incorrect. But Pareto is inconsistent with the claim that if the $\succcurlyeq$ is a qualitative internal relation, and so we should not be surprised that the Strong Anonymity principle conflicts with Pareto.

This confusion also means that Anonymity principles clearly fail to capture what they were intended to capture: whether an ethical ranking is equitable or not. If w_2 is better than w_1 by Pareto, then no equity principle should entail that w_1 and w_2 are equally good, as Strong Anonymity does. But Finite Anonymity and Fixed-Step Anonymity seem to be too weak to capture what we mean by equity. Permuting the utility levels of infinitely many agents should not result in a world that is strictly better if it does not result in a Pareto improvement. This is entailed by Pareto and the qualitativeness of $\succcurlyeq$.

The qualitativeness of $\succcurlyeq$ says that we should not favor agents based solely on *who* they are. This seems like a key component of our concept of an equitable ethical theory. When we combine this axiom with Pareto, it follows that we must not give different weight to utility that agents have based solely on qualitative properties like where they are located in spacetime. If we do give more weight to utility based on such properties then, since we can find a world pair in which different agents play those qualitative roles, our theory will come into conflict with Pareto. The idea that we should not give more weight to someone's utility on the basis of qualitative properties like where they are located in spacetime is also a key component of our concept of an equitable theory.[158] And, as I have shown, these two principles jointly entail that if the utility levels of w_2 is a finite, fixed-step, or – in some cases – an infinite permutation of the utility of w_1, then w_2 is not strictly better or worse than w_1.

[158]One might object that Pareto is itself an inequitable principle because it can give extra weight to the utility of agents on the basis of the qualitative property 'being a shared agent' even if it is not consistent with giving extra weight to other qualitative properties. The results of Chapter 3 entail that Pareto and the qualitativeness of $\succcurlyeq$ are jointly inconsistent with giving extra weight to the utility of shared agents.

I therefore believe that the qualitativeness of $\succcurlyeq$ and Pareto jointly constitute a novel kind of equity that is superior to the anonymity axioms outlined in the previous chapter:[159]

Paretian Equity

If (i) it's not the case that $w_1 \succ w_2$ or $w_2 \succ w_1$ by Pareto and (ii) the utility levels of the agents at w_1 is a (finite, fixed step, or variable step) permutation of the utility levels of the agents at w_2, then w_1 is not strictly better than w_2 and w_2 is not strictly better than w_1.

Consider a view like limit discounted utilitarianism [114]. This view satisfies Pareto and Fixed-Step Anonymity because any fixed-step permutations of utility will have no effect on the the limit of a sequence of discounted utility.[160] I contend that it is inequitable to spatiotemporally discount agents' utilities even if doing so is consistent with Finite and Fixed-Step Anonymity. Suppose that there are infinitely many agents with utility 1 lives and infinitely many agents with utility 2 lives in both w_1 and w_2. The agents in w_1 come in the temporal order $(2, 1, 2, 1, 2, 1, ...)$ while the utilities in w_2 come in the temporal order $(1, 1, 2, 1, 1, 2, ...)$. For any choice of discount rate, limit discount utilitarianism will rank w_1 over w_2. But w_1 and w_2 could contain exactly the same agents with exactly the same utility levels. The agents might simply appear in a different temporal order. Given this, any theory that says one of these worlds is strictly better than the other seems inequitable, since it gives more weight to agents' utility levels depending on when they occur. The verdict of limit discounted utilitarianism is consistent with finite and fixed-step anonymity, however, since the first world can only be derived from the second by means of an infinite variable step permutation of utilities. The verdict of limit discounted utilitarianism is inconsistent with Pareto and the qualitativeness of $\succcurlyeq$, however. This suggests that Pareto and the qualitativeness of $\succcurlyeq$ jointly place more plausible equity constraints on theories than finite or

[159] As we adopt extensions of Pareto, the qualitativeness of $\succcurlyeq$ will guarantee that our concept of equitableness will also be extended.

[160] For example, the limit of the sequence $(2, 1, 1, 1, 1, ...)$ is equal to the limit of the sequence $(1, 2, 1, 1, 1, 1...)$ regardless of our choice of discount rate.

fixed-step anonymity.

In this section I have argued that Pareto and the qualitativeness of $\succcurlyeq$ jointly entail a more plausible account of equity than either finite or fixed-step anonymity that is more in line with our intuitions.[161] The key reason that has been given for rejecting equity principles that are stricter than finite or fixed-step anonymity is that they conflict with Pareto. I have shown that this is only true of Strong Anonymity because it assumes that $\succcurlyeq$ is a qualitative internal relation. If we reject the assumption that $\succcurlyeq$ is a qualitative internal relation but maintain that $\succcurlyeq$ is a necessary qualitative relation, we can maintain that if the utility levels of agents at w_1 are a variable step permutation of the utility levels of agents at w_2 and neither of these worlds is strictly better than the other by Pareto, then neither of these worlds is strictly better than the other.

Summary

In this chapter I have shown that if we accept four axioms – Pareto, transitivity, the qualitativeness of $\succcurlyeq$, and the Permutation Principle, we must reject the completeness axiom: we must accept that some world pairs are such that $w_1 \not\succcurlyeq w_2$ and $w_2 \not\succcurlyeq w_1$. If we want to retain completeness then we must reject one of the four axioms I have formulated.

We could reject Pareto, but Pareto is a fairly fundamental ethical principle. If we are forced to choose between completeness and the claim that if no agents are better off in w_1 than they are in w_2 and some agents are strictly worse off in w_1 than they are in w_2 then w_1 is worse than w_2, we seem to be justified in choosing to reject completeness.

I believe that rejecting transitivity will be of little help in the cases I discuss since, as I show in Chapter 5, retaining completeness at the cost of transitivity requires accepting implausible

[161] For principles that satisfy Pareto and fixed-step Anonymity if basic locations of value come in a natural order, see the fixed-step anonymous overtaking and catching-up principles formulated by Asheim and Banerjee [9]. We can use the same argument to show that these are inconsistent with Pareto and Equity.

cyclic rankings of infinite worlds. Not only do such cycles create problems for subjective ethical decision making, but they also lack any independent motivation. I will argue that incompleteness generates similar difficulties for subjective ethical decision making but the verdict that Clement and Stormy are incomparable seems more independently motivated.

If we reject the Permutation Principle then we must claim that it is not always logically possible to find a qualitative duplicate of a world pair under a given permutation, which is highly implausible. Not only this, but we would still be able to generate incompleteness among at least some of the world pairs of the kind that I discuss above, unless we make the even stronger claim that it is never possible to find such qualitative duplicates of these pairs.

Finally, if we reject the claim that the $\succcurlyeq$ relation is a necessary qualitative relation, then we must conclude that haecceitistic facts – facts about whether Obama is in a given world or not, for example – can determine the ethical ranking of worlds. The idea that haecceitistic facts could play this kind of role in ethics seems highly objectionable. Even if it were to be adopted, however, it is not clear how much good it would do us. Suppose we were to claim that haecceitistic facts are essential for the ethical ranking of worlds. This would prevent us from claiming that if $\langle w_3, w_4 \rangle$ is qualitatively identical to $\langle w_1, w_2 \rangle$ then $w_1 \succcurlyeq w_2$ if and only if $w_3 \succcurlyeq w_4$. But it is not clear how we could use haecceitistic facts to construct a positive ethical theory that would allow us to rank infinite worlds like Clement and Stormy.

In light of this, I believe that we must retain all four of the axioms I have formulated and reject the completeness axiom. This means that we must accept that infinite world pairs like Clement and Stormy, Mansion and Shack, and Optimistic Cube and Pessimistic cube are genuinely incomparable. In this chapter I have used just a few illustrative examples of world pairs that are incomparable if we accept Pareto, the qualitativeness of $\succcurlyeq$, the Permutation Principle, and transitivity. In the next chapter I will generalize these results. I show that many types of world pairs are incomparable by the axioms formulated in this chapter.

Chapter 3

The Incomparability Results

In this chapter I formulate the general conditions that entail two worlds are incomparable by the kind of 'four world' argument given in the previous chapter. I in section 3.1 I demonstrate that many classes of world pairs with disjoint populations are incomparable by four world arguments. In section 3.2 I extend these results to world pairs with identical populations. In section 3.3 I offer a general four world result that applies to world pairs with disjoint, identical, and overlapping populations. In section 3.4 I show that certain world pairs cannot be shown to be incomparable by a four world argument but their incomparability is entailed by the axioms of the previous chapter via what I call a 'cyclic argument'. I then formulate the general conditions that entail two worlds are incomparable by a cyclic argument.

3.1 Incomparability in Disjoint Population Pairs

In the previous chapter I focused on world pairs in which agents had only two utility levels: happy or miserable. This is clearly an idealization since agents can be happy or sad to varying degrees. In this chapter I will assume that agents can have any real-valued level of utility and that we can measure agents' utility levels on a common interval scale.[162] I will focus on total lifetime utility levels rather than utility levels at a time. So an agent whose entire life comprises three days at utility 2 in world w_1 has a lifetime utility level of 6.

To begin with, we can continue to focus on infinite world pairs that have disjoint populations. Consider the following disjoint population world pair. World w_1 contains infinitely many agents in A at utility level 1 and infinitely many agents in C at utility 2, while w_2 contains infinitely many agents in B at utility level 3 and infinitely many agents in D at utility 2:

	A	B	C	D
w_1	1	–	2	–
w_2	–	3	–	2

Figure 48: Disjoint worlds with utilities outside the [0,1] interval

It will be helpful for us to introduce the concept of a bijection that is a 'strict upgrade' of the population of world w_1. Let $pl(w_1)$ denote the population of world w_1 and let $u_{w_1}(x)$ denote the utility that agent x has in world w_1. A strict upgrade from the population of one world to the population of another world can be defined as follows:

Strict Upgrade

A bijection g from $pl(w_1)$ to $pl(w_2)$ is a strict upgrade from $pl(w_1)$ to $pl(w_2)$ iff for all agents x in w_1, $u_{w_1}(x) \leq u_{w_2}(g(x))$ and for some agent x in w_1, $u_{w_1}(x) < u_{w_2}(g(x))$

[162] Here I will not assume that utility has a privileged zero point but I will return to this issue in the next chapter where I defend the existence of a privileged zero point for utility. Although I am most sympathetic to a welfarist account of utility in line with that defended by Ng [171] and others, I will try to remain agnostic about what grounds utilities, as long as it is consistent with the properties I ascribe to utilities here.

We can perform a strict upgrade of the population of w_1 above by mapping every agent in set A that has utility 1 in w_1 onto an agent in set B that has utility 3 in w_2 and mapping every agent in set C that has utility 2 in w_1 onto an agent in set D that has utility 2 in w_2. Let g be a bijection from the population of w_1 to the population of w_2 such that $g(a_i) = b_i$, so that g is a strict upgrade of the population of w_1. Let f be a permutation of the agents of $\langle w_1, w_2 \rangle$ such that $f(x) = g(x)$ if x exists at w_1 and $f(x) = g^{-1}(x)$ if x exists at w_2. By the Permutation Principle, we can find a world pair $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under the permutation f. We can depict this permutation as follows:

	A	B	C	D
w_1	1	–	2	–
w_2	–	3	–	2
w_3	–	1	–	2
w_4	3	–	2	–

Figure 49: A strict upgrade from w_1 to w_2

Notice that world w_2 is strictly better than world w_3 by Pareto and world w_4 is strictly better than world w_1 by Pareto. If $w_1 \succcurlyeq w_2$ then $w_3 \succcurlyeq w_4$ by the qualitativeness of $\succcurlyeq$. So $w_1 \not\succcurlyeq w_2$ by transitivity.

I will show that this generalizes: if w_1 and w_2 have disjoint populations and there is a strict upgrade from the population of w_1 to the population of w_2 then we can show that $w_1 \not\succcurlyeq w_2$.[163]

Result 1: Disjoint Pair $\not\succcurlyeq$ by a Four World Argument

If w_1 and w_2 have disjoint populations and there is a strict upgrade from $pl(w_1)$ to $pl(w_2)$ then $w_1 \not\succcurlyeq w_2$.

To show this, suppose that w_1 and w_2 have disjoint populations and there exists a strict upgrade g from the population of w_1 to the population of w_2. Since such a bijection exists, the

[163] These conditions are in fact necessary and sufficient for there to be a four world argument showing that $w_1 \not\succcurlyeq w_2$, but I will only establish sufficiency here.

populations of w_1 and w_2 must be equinumerous. Define a permutation f of the populations of w_1 and w_2 as follows: $f(x) = g(x)$ if x exists at w_1 and $f(x) = g^{-1}(x)$ if x exists at w_2. Since the populations of w_1 and w_2 are disjoint, this results in a permutation of $pl(\langle w_1, w_2 \rangle)$. By the Permutation Principle, we can find a pair of worlds $\langle w_3, w_4 \rangle$ that is a qualitative duplicate under f. But it follows from Pareto that $w_2 \succ w_3$ and $w_4 \succ w_1$. Given transitivity, this results in the following contradiction:

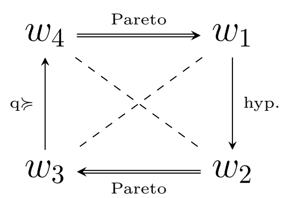


Figure 50: Transitivity violation in disjoint world pairs with a strict upgrade

To show that world w_2 is strictly better than w_3 by Pareto, note that for all agents x that exist in w_1, agent $g(x)$ exists at w_3 and so worlds w_2 and w_3 have identical populations. Since $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f, the qualitative properties of each agent x in w_1 are the same as the qualitative properties of $f(x)$ in w_3. In particular, the utility of agent x in w_1 is the same as the utility of $f(x)$ at w_3. Since there is some agent x in w_1 such that $u(x) < u(g(x))$ there will be some agent $g(x)$ in w_3 that contains strictly less utility than agent $g(x)$ in w_2 does. And since all other agents x in w_1 are such that $u(x) \leq u(g(x))$, there will be no agent $g(x)$ in w_3 that contains strictly greater utility than any agent $g(x)$ in w_2 does. Therefore world w_2 is strictly better than w_3 by Pareto.

To show that $w_4 \succ w_1$ by Pareto, note that for all agents x that exist in w_2, agent $g^{-1}(x)$ exists at w_4 and so worlds w_4 and w_1 have identical populations. Since $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f, the qualitative properties of each agent x in w_2 are the same as the qualitative properties of $f(x)$ at w_4. In particular, the utility of agent x in w_2 is the same as the utility of $f(x)$ in w_4. Since there is some agent x in w_2 such that $u(x) > u(g^{-1}(x))$

there will be some agent $g^{-1}(x)$ in w_1 that contains strictly less utility than agent x in w_4 does. And since all other agents x in w_2 are such that $u(x) \geq u(g^{-1}(x))$, there will be no agent $g^{-1}(x)$ in w_4 that contains strictly less utility than any agent $g^{-1}(x)$ in w_1 does. Therefore world w_4 is strictly better than w_1 by Pareto.

This establishes Result 1. Given this, it is easy to show the following:

Result 2: Disjoint Pair Incomparability by a Four World Argument

If w_1 and w_2 have disjoint populations and there is a strict upgrade from $pl(w_1)$ to $pl(w_2)$ and there is a strict upgrade from $pl(w_2)$ to $pl(w_1)$ then w_1 and w_2 are incomparable ($w_1 \not\gtrless w_2$).

If w_1 and w_2 have disjoint populations and there is a strict upgrade from $pl(w_1)$ to $pl(w_2)$ and there is a strict upgrade from $pl(w_2)$ to $pl(w_1)$ then $w_1 \not\succcurlyeq w_2$ by Result 1 and $w_2 \not\succcurlyeq w_1$ by Result 1. Therefore w_1 and w_2 are incomparable ($w_1 \not\gtrless w_2$).

I have now established that if w_1 and w_2 have disjoint populations and there exists a strict upgrade from the population of w_1 to the population of w_2 and there exists a strict upgrade from the population of w_2 to the population of w_1 then w_1 and w_2 are incomparable.[164]

Next I will show that certain properties are sufficient for there to exist both a strict upgrade from $pl(w_1)$ to $pl(w_2)$ and a strict upgrade from $pl(w_2)$ to $pl(w_1)$.

Result 3: A Sufficient Condition for Disjoint Pair Incomparability

If w_1 and w_2 have countable disjoint populations and both worlds have infinitely many agents at utility level n and infinitely many agents at utility level m, where $n < m$ and no agents with utility levels less than n or greater than m then w_1 and w_2 are incomparable.

To establish Result 3, it is sufficient to show that if w_1 and w_2 have countable disjoint populations and both worlds have infinitely many agents at utility level n and infinitely

[164] It should be clear from the argument above that if there is a weak but not a strict upgrade from the population of w_1 to the population of w_2 – i.e. there is no agent x that has strictly less utility in w_1 than $f(x)$ is in w_2 – then we can show by a four world argument that $w_1 \not\succ w_2$.

many agents at utility level m, where $n < m$ and no agents with utility levels less than n or greater than m, then there exists a strict upgrade g from the population of w_1 to the population of w_2 and there exists a strict upgrade h from the population of w_2 to the population of w_1.

To show that there will exist a strict upgrade g from the population of w_1 to the population of w_2, we're first going to split the lowest utility agents in w_1 into two infinite sets – A_1 and A_2. We're also going to split the highest utility agents in w_2 into two infinite sets – B_1 and B_2. Let g be a mapping from all of the agents in A_1 in w_1 onto all of the agents in B_1 in w_2. So $g(A_1) = B_1$ where, for all agents in $a \in A_1$ and $b \in B_1$, $u(a) < u(b)$.

Let C_1 be the remaining agents that are in w_1 that are not in A_1 or A_2. There are countably infinitely many agents in C_1 and their utilities are all strictly greater than the lowest level and less than or equal to the highest utility level. Let C_2 be the remaining agents that are in w_2 that are not in B_2. Their utilities are strictly less than the utility highest level and less than or equal to the lowest utility level. Let g map all of the agents in A_2 in w_1 onto all of the agents in C_2 in w_2. Let g map all of the agents in C_1 in w_1 onto all of the agents in B_2 in w_2. So $g(A_2) = C_2$ where, for all agents in $a \in A_2$ and $c \in C_2$, $u(a) \leq u(c)$. And $g(C_1) = B_2$ where, for all agents in $c \in C_1$ and $b \in B_2$, $u(c) \leq u(b)$.

Each agent in w_1 is now mapped onto an agent in w_2 by g and each agent in w_2 is mapped by g^{-1} onto an agent in w_1. Therefore g is a bijection from $pl(w_1)$ to $pl(w_2)$ such that for some agent x, $u_{w_1}(x) < u_{w_2}(g(x))$ and for all agents y, $u_{w_1}(y) \leq u_{w_2}(g(y))$. This shows that if w_1 and w_2 have countable disjoint populations and both worlds have infinitely many agents at utility level n and infinitely many agents at utility level m, where $n < m$ and no agents with utility levels less than n or greater than m, then there will exist a strict upgrade g from w_1 to w_2. Therefore $w_1 \not\succcurlyeq w_2$.

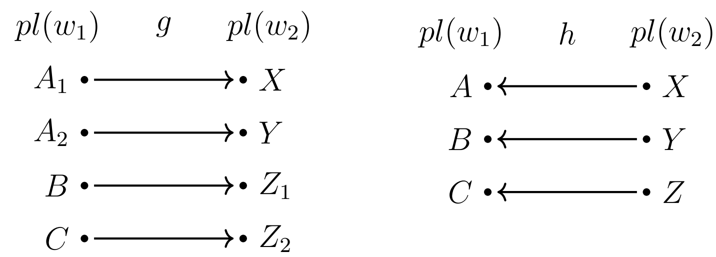
We can use the same method to find a strict upgrade h from the population of w_2 to the population of w_1. It follows that $w_2 \not\succcurlyeq w_1$. Therefore w_1 and w_2 are ethically incomparable.

This establishes Result 3. All disjoint population worlds in which infinitely many agents have the same upper level of utility and infinitely many agents have the same lower level of utility are incomparable. This includes disjoint population world pairs like Clement and Stormy that have two utility levels and infinitely many agent at each utility level. But it also includes world pairs in which agents have more than one utility level. Consider worlds w_1 and w_2 that both contain infinitely many agents at utility levels 1 and 5, and infinitely many agents with utility levels between 1 and 5 in the following infinite sets:

	A	B	C	X	Y	Z
w_1	1	4	5	–	–	–
w_2	–	–	–	1	3	5

Figure 51: A world pair with distinct 'inner' utility levels

There is a strict upgrade g from the population of w_1 to the population of w_2 and there is a strict upgrade h from the population of w_2 to the population of w_1. To construct g we need to split the agents in A into two infinite sets of agents, A_1 and A_2 and we need to split the agents in Z into two infinite sets of agents Z_1 and Z_2. We can depict bijections g and h as follows, where an arrow between sets A and X means that each agent a_i in set A is mapped by bijection g onto an agent x_i in set X, such that all agents in A in w_1 are mapped to an agent in X in w_2 by g and all agents in X in w_2 are mapped to an agent (the same agent) in A in w_1 by g^{-1}:



Define a permutation f_1 of the populations of w_1 and w_2 as follows: $f_1(x) = g(x)$ if x exists at w_1 and $f(x) = g^{-1}(x)$ if x exists at w_2. Define a permutation f_2 of the populations of w_1 and w_2 as follows: $f_2(x) = h(x)$ if x exists at w_2 and $f(x) = h^{-1}(x)$ if x exists at w_1. Let $\langle w_3, w_4 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f_1 and let $\langle w_5, w_6 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f_2. These word pairs are as follows:

	A_1	A_2	B	C	X	Y	Z_1	Z_2
w_1	1	1	4	5	–	–	–	–
w_2	–	–	–	–	1	3	5	5
w_3	–	–	–	–	1	1	4	5
w_4	1	3	5	5	–	–	–	–

	A	B	C	X	Y	Z
w_1	1	4	5	–	–	–
w_2	–	–	–	1	3	5
w_5	–	–	–	1	4	5
w_6	1	3	5	–	–	–

Figure 52: Permutations showing that world pairs with distinct inner utility levels are incomparable

We can see that $\langle w_3, w_4 \rangle$ shows that $w_1 \not\succcurlyeq w_2$ given our auxiliary premises and $\langle w_5, w_6 \rangle$ shows that $w_2 \not\succcurlyeq w_1$ given our auxiliary premises, as expected.

Result 3 is limited in various respects. It does not apply to world pairs in which all agents have the same utility levels, or pairs that don't contain infinitely many agents with utilities equal to the upper and lower utility level, or pairs in which there are agents at utility levels strictly less than or strictly greater than these endpoints.[165] It also does not apply to world

[165] We can offer a modest extension of Result 3 in such cases. If w_1 and w_2 have finitely many agents strictly less than n and strictly greater than m and there still exists a bijection g from w_1 to w_2 that satisfies (a) and (b) then $w_1 \not\succcurlyeq w_2$. This means that if, for every agent at utility level i outside of the interval $[n, m]$ in w_1 there is an agent with utility i in w_2 and vice versa, then we can show that these worlds are incomparable.

pairs with uncountable populations, unbounded utilities, or non-equinumerous populations. However, the goal is not to establish all of the conditions that are sufficient for disjoint worlds to be incomparable by a four world argument, but to show that many disjoint world pairs can be shown to be incomparable by a four world argument. Result 3 shows this.

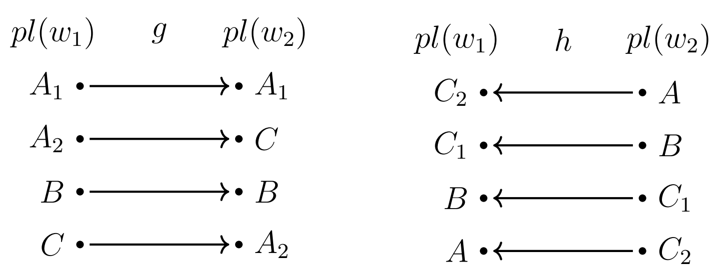
3.2 Incomparability in Identical Population Pairs

So far I have only considered world pairs with disjoint populations. We can also show that many pairs of worlds with identical populations – worlds that contain exactly the same agents – are incomparable by a four world argument. Consider the worlds w_1 and w_2 that contain the same infinite sets of agents with the following utility levels:

	A	B	C
w_1	1	3	7
w_2	7	4	1

Figure 53: An identical population world pair with distinct inner utility levels

We can show that w_1 and w_2 are incomparable by a four world argument. Let us split the agents in set A into infinitely many A_1 agents and infinitely many A_2 agents, and let us split the agents in set C into infinitely many C_1 agents and infinitely many C_2 agents. Consider bijections g and h on $A \cup B \cup C$ as follows, where an arrow between sets X and Y means that g is a bijection from X to Y:



Let $\langle w_3, w_4 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g and let $\langle w_5, w_6 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under h. These world pairs are as follows:

	A_1	A_2	B	C
w_1	1	1	3	7
w_2	7	7	4	1
w_3	1	7	3	1
w_4	7	1	4	7

	A	B	C_1	C_2
w_1	1	3	7	7
w_2	7	4	1	1
w_5	7	7	3	1
w_6	1	1	4	7

Figure 54: Permutations showing that identical population pairs with distinct inner utility levels are incomparable

We can see that $\langle w_3, w_4 \rangle$ shows that $w_1 \not\succcurlyeq w_2$ given our auxiliary premises. For assume $w_1 \succcurlyeq w_2$. Then $w_2 \succ w_3$ by Pareto, $w_3 \succcurlyeq w_4$ by the qualitativeness of $\succcurlyeq$, and $w_4 \succ w_1$ by Pareto, violating transitivity. We can also see that $\langle w_5, w_6 \rangle$ shows that $w_2 \not\succcurlyeq w_1$ by our auxiliary premises. For assume $w_2 \succcurlyeq w_1$. Then $w_1 \succ w_6$ by Pareto, $w_6 \succcurlyeq w_5$ by the qualitativeness of $\succcurlyeq$, and $w_5 \succ w_2$ by Pareto, violating transitivity.

If w_1 and w_2 have identical populations then in order to show that world w_1 is not at least as good as w_2 by a four world argument, it must be the case that these exists a bijection g on the relevant set of agents such that (i) for all agents x, $u_{w_1}(x)$ is less than or equal to $u_{w_2}(g(x))$ for all x in w_1 and all $g(x)$ in w_2 and (ii) for some agent x, $u_{w_1}(x)$ is strictly less than $u_{w_2}(g(x))$. To see why we need (i), suppose that in the case above each agent in A had utility 2 in w_1 instead of utility 1. If this were the case then we would be unable to show that w_1 is not at least as good as w_2 by a four world argument because it would not be possible for w_2 to better than *any* qualitative duplicate of w_1 by Pareto since w_2 contains some agents at utility 1 (the agents in C) while duplicates of w_1 would contain only agents with utility strictly greater than 1. And we need (ii) to guarantee that w_3 can be strictly rather than just weakly better than w_2 by Pareto.

If w_1 and w_2 have identical populations then this is not all we need to show that world w_1 is not at least as good as w_2 by a four world argument, however. To see why, consider the following identical population pair of infinite worlds $\langle w_1, w_2 \rangle$:

	X_1	X_2	Y	Z_1	Z_2
w_1	0	0	1	1	1
w_2	0	0	1	1	1

Figure 55: Identical population world pair with identical utilities

Now consider the following bijection f from $pl(w_1)$ to $pl(w_2)$:

$$\begin{array}{rcl} pl(w_1) & f & pl(w_2) \\ X_1 \bullet & \longrightarrow & \bullet\ X_1 \cup X_2 \\ X_2 \bullet & \longrightarrow & \bullet\ Y \\ Y \bullet & \longrightarrow & \bullet\ Z_1 \\ Z_1 \cup Z_2 \bullet & \longrightarrow & \bullet\ Z_2 \end{array}$$

A bijection that maps an infinite subset of the agents in X in w_1 onto all the agents in X in w_2 and that maps all the agents in Z in w_1 onto an infinite subset of the agents in Z in w_2 is possible because infinite sets can be equinumerous with their proper subsets. For example, the set of all even numbers can be mapped one-to-one onto the set of all natural numbers. Bijection g is a strict upgrade from $pl(w_1)$ to $pl(w_2)$, i.e. for some $x, u_{w_1}(x) < u_{w_2}(g(x))$ – since for all x in X_2 and all y in Y, $u_{w_1}(x) < u_{w_2}(y)$ – and for all $x, u_{w_1}(x) \leq u_{w_2}(g(x))$. But consider the world pair $\langle w_3, w_3 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f:

	X_1	X_2	Y	Z_1	Z_2
w_1	0	0	1	1	1
w_2	0	0	1	1	1
w_3	0	0	0	1	1
w_4	0	0	0	1	1

Figure 56: Permuting the populations of an identical population, identical utility world pair

We can see that $w_2 \succ w_3$ by Pareto. But Pareto does not entail that $w_4 \succ w_1$. In fact, it entails that $w_1 \succ w_4$! World w_4 is not better than w_1 by Pareto because there is some agent x in w_1 such that $u_{w_1}(x) > u_{w_2}(f^{-1}(x))$. Specifically, for all y in Y and for all x in X_2, $u_{w_1}(y)$ is strictly greater than $u_{w_2}(x)$. This means that if w_4 is a qualitative duplicate of w_2 under g then because $u_{w_2}(f^{-1}(x))$ is strictly less than $u_{w_1}(x)$, it must be the case that $u_{w_4}(f(f^{-1}(x))$ is strictly less than $u_{w_1}(x)$. But since $u(f(f^{-1}(x)) = x$, this means $u_{w_4}(x)$ is strictly less than $u_{w_1}(x)$, which entails that w_4 cannot be at least as good as w_1 by Pareto.

So if w_1 and w_2 have identical populations then in order to show that w_1 is not at least as good as w_2 by a four world argument, we must supplement (i) and (ii) with a further condition: for all agents x, the utility that x has in w_1 is less than or equal to the utility that $g^{-1}(x)$ has in w_2. It is easy to see that this condition is satisfied by the bijections g and h in the case above this one. In fact, in that case there was nothing to stop us from choosing both of these function in such a way that they were their *own* inverses.

In this case, however, it is not possible to find a bijection from the population of w_1 to the population of w_2 that satisfies all three of these conditions. Because the agents of $\langle w_1, w_2\rangle$ have the same utility at w_1 and w_2, and bijection such that $u_{w_1}(x) < u_{w_2}(g(x))$ must also be such that $u_{w_1}(g(x)) > u_{w_2}(g^{-1}(g(x)))$. Therefore we cannot show that w_1 is not at least as good as w_2 by a four world argument. This is what we would expect since if we could find such a bijection from the population of w_1 to the population of w_2, we would be able to use such a bijection to show that w_1 is not at least as good as itself by a four world argument.[166]

In light of this, it will be helpful to define a strengthened notion of a strict upgrade. We can define a 'bidirectional upgrade' from $pl(w_1)$ to $pl(w_2)$ as follows:

Bidirectional Upgrade

A bijection g from $pl(w_1)$ to $pl(w_2)$ is a bidirectional upgrade from $pl(w_1)$ to $pl(w_2)$ iff (i) for all agents $x \in pl(w_1)$, $u_{w_1}(x) \leq u_{w_2}(g(x))$, (ii) for all agents $x \in pl(w_2)$, $u_{w_1}(x) \leq u_{w_2}(g^{-1}(x))$, and (iii) for some agent x in w_1, $u_{w_1}(x) < u_{w_2}(g(x))$

I will now show that if w_1 and w_2 have identical populations and there is a bidirectional upgrade from $pl(w_1)$ to $pl(w_2)$ then $w_1 \not\succcurlyeq w_2$ by a four world argument.

Result 4: Identical Pair $\not\succcurlyeq$ by a Four World Argument

If w_1 and w_2 have identical populations and there is a bidirectional upgrade from $pl(w_1)$ to $pl(w_2)$ then $w_1 \not\succcurlyeq w_2$.

To establish result 4 we employ a line of argument that is now familiar. Suppose that w_1 and w_2 have identical populations and there exists a bidirectional upgrade g from the population of w_1 to the population of w_2. Since g is a bijection from the population of w_1 to the population of w_2 and the populations of w_1 and w_2 are identical, g is a permutation the populations of $\langle w_1, w_2 \rangle$. By the Permutation Principle, we can find a pair of worlds $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g, such that $w_2 \succ w_3$ and $w_4 \succcurlyeq w_1$ by Pareto.

To show that world w_2 is strictly better than w_3 by Pareto, note that since g is a bijection from the population of w_1 to the population of w_2, worlds w_2 and w_3 have identical populations. Since $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g, the qualitative properties of

[166] This is exactly the sort of case that shows that, if we accept the Permutation Principle, Pareto is not consistent with the claim that $\succcurlyeq$ is an internal relation. However, $w_1 \sim w_1$ by Pareto and so any qualitative duplicate of the pair $\langle w_1, w_1 \rangle$ will also be equally good by Pareto. So if $\succcurlyeq$ is not internal, it doesn't matter if both members of the qualitative duplicate pair are strictly worse (or better) than w_1 by Pareto.

each agent x in w_1 are the same as the qualitative properties of $g(x)$ in w_3. In particular, the utility of agent x in w_1 is the same as the utility of $g(x)$ at w_3. Since there is some agent x such that $u_{w_1}(x) < u_{w_2}(g(x))$ there will be some agent $g(x)$ in w_3 that contains strictly less utility than the agent $g(x)$ in w_2 does. Since all agents x in w_1 are such that $u_{w_1}(x) \leq u_{w_2}(g(x))$ there will be no agent $g(x)$ in w_3 that contains strictly greater utility than any agent $g(x)$ in w_2 does. Therefore world w_2 is strictly better than w_3 by Pareto.

To show that $w_4 \succcurlyeq w_1$ by Pareto, note that since g is a bijection from the population of w_1 to the population of w_2, worlds w_4 and w_1 have identical populations. Since $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g, the qualitative properties of each agent x in w_2 are the same as the qualitative properties of $g(x)$ in w_4. In particular, the utility of agent x in w_2 is the same as the utility of $g(x)$ at w_4. Since every agent x in w_2 are such that $u_{w_2}(g^{-1}(x)) \geq u_{w_1}(x)$, every agent $g(g^{-1}(x))$ in w_4 also has at least as much utility as agent x in w_1 has. Since $g(g^{-1}(x)) = x$, it follows that w_4 is at least as good as w_1 by Pareto.

We assumed that $w_1 \succcurlyeq w_2$. Since $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g, it follows that $w_3 \succcurlyeq w_4$ by the qualitativeness of $\succcurlyeq$. But $w_2 \succ w_3$ and $w_4 \succcurlyeq w_1$ by Pareto. Given transitivity, this results in the following contradiction:

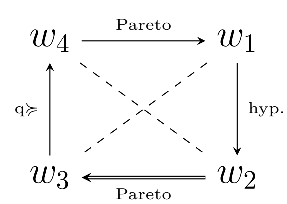


Figure 57: Transitivity violation in identical population world pair with a bidirectional upgrade

It follows that identical population worlds are ethically incomparable – $w_1 \not\gtrless w_2$ – if there exists a bidirectional upgrade from the agents of w_1 to the agents of w_2 and there exists a bidirectional upgrade from the agents of w_2 to the agents of w_1.

Result 5: Identical Pair Incomparability by a Four World Argument

If w_1 and w_2 have identical populations and there is a bidirectional upgrade from $pl(w_1)$ to $pl(w_2)$ and there is a bidirectional upgrade from $pl(w_2)$ to $pl(w_1)$ then w_1 and w_2 are incomparable ($w_1 \not\gtrless w_2$).

If w_1 and w_2 have identical populations and there is a bidirectional upgrade from $pl(w_1)$ to $pl(w_2)$ and there is a bidirectional upgrade from $pl(w_2)$ to $pl(w_1)$ then $w_1 \not\succcurlyeq w_2$ by Result 4 and $w_2 \not\succcurlyeq w_1$ by Result 4. Therefore w_1 and w_2 are incomparable ($w_1 \not\gtrless w_2$).

Notice that we do not require that both $w_2 \succ w_3$ and $w_4 \succ w_1$ by Pareto to show that $w_1 \not\succcurlyeq w_2$ by a four world argument in the claims above – it is sufficient to show that $w_2 \succ w_3$ and $w_4 \succcurlyeq w_1$. For example, consider the following two worlds w_1 and w_2:

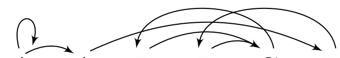
	A	B	C	D
w_1	0	2	0	0
w_2	0	0	2	2

Figure 58: A further example of an identical population world pair

And consider the bijection g from the agents of world w_1 to the agents of world w_2:

$$\begin{array}{rcl} pl(w_1) & g & pl(w_2) \\ A_1 & \longrightarrow & A \\ A_2 & \longrightarrow & D \\ B & \longrightarrow & C \\ C & \longrightarrow & B_1 \\ D & \longrightarrow & B_2 \end{array}$$

Let $\langle w_3, w_4 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g. The utilities of the agents of the pair $\langle w_3, w_4 \rangle$ are as follows:



	A_1	A_2	B_1	B_2	C	D
w_1	0	0	2	2	0	0
w_2	0	0	0	0	2	2
w_3	0	0	0	0	2	0
w_4	0	0	2	2	0	0

Figure 59: A permutation showing that $w_1 \not\succcurlyeq w_2$

In this case $w_1 \succcurlyeq w_2$ by hypothesis, $w_2 \succ w_3$ by Pareto, $w_3 \succcurlyeq w_4$ by the qualitativeness of $\succcurlyeq$ and $w_4 \succcurlyeq w_1$ by Pareto. This violates transitivity and so $w_1 \not\succcurlyeq w_2$.

As with disjoint population worlds, I will not formulate further conditions that are necessary and sufficient for there to be a strict upgrade from w_1 to w_2 and a strict upgrade from w_2 to w_1 in identical population cases. But I will, as an example, show that a class of identical population world pairs are incomparable. Let us say that there is a 'utility isomorphism' between a set of agents in w_1 and a set of agents in w_2 if and only if there exists a bijection g from the set of agents in w_1 to the set of agents in w_2 such that, for all agents x, iff $g(x)$ exists in w_1 and $u_{w_1}(x) = u_{w_2}(g(x))$ and $u_{w_2}(x) = u_{w_1}(g(x))$.[167] Given this, we can show that the following identical population world pairs are incomparable:

Result 6: A Sufficient Condition for Identical Population Pair Incomparability

If w_1 and w_2 have countable identical populations and there is an infinite set of agents x in $\langle w_1, w_2\rangle$ such that $u_{w_1}(x) = n$ and $u_{w_2}(x) = m$ and there is an infinite set of agents y in $\langle w_1, w_2\rangle$ such that $u_{w_2}(y) = n$ and $u_{w_1}(y) = m$, where $n < m$, and there is a utility isomorphism between the remaining agents of w_1 and w_2, then w_1 and w_2 are incomparable.

To establish Result 6, we use the same familiar reasoning employed in the results above to show that if w_1 and w_2 have countable identical populations and there is an infinite set of

[167]If we assume that an existing agent cannot have the same utility as a non-existing agent, this entails that x exists in w_1 iff $g(x)$ exists in w_2 and x exists in w_2. In the next chapter, however, I consider the view that we should treat non-existing agents as we would treat agents with utility zero lives.

agents x in $\langle w_1, w_2 \rangle$ such that $u_{w_1}(x) = n$ and $u_{w_2}(x) = m$ and there is an infinite set of agents y in $\langle w_1, w_2 \rangle$ such that $u_{w_2}(y) = n$ and $u_{w_1}(y) = m$, where $n > m$, and there is a utility isomorphism between the remaining agents of w_1 or w_2, then there exists a bidirectional upgrade g from the population of w_1 to the population of w_2 an there exists a bidirectional upgrade h from the population of w_2 to the population of w_1.

To show that there will exist a bidirectional upgrade g from the population of w_1 to the population of w_2, let X be the set of agents at utility n in w_1 and utility m in w_2 and let Y be the set of agents at utility m in w_1 and utility n in w_2 and let Z be the set of all remaining agents in w_1. We split the Y-agents into two infinite sets – Y_1 and Y_2. Next, for all $x \in X$ and for all $y \in Y_2$, let $g(x) = y$, for all $y \in Y_1$, let $g(y) = y$. For all $y \in Y_2$ and for all $x \in X$, let $g(y) = x$. And for all $z_i \in Z$, let $g(z_i) = z_j$ for some $z_j \in Z$ such that $u_{w_1}(z_i) = u_{w_2}(z_j)$.

Since the agents in X, Y_1, Y_2 and Z comprise all of the agents in $\langle w_1, w_2 \rangle$, g is a bijection from the population of $\langle w_1, w_2 \rangle$ onto itself. Since $u_{w_1}(y \in Y_1) = m < u_{w_2}(y \in Y_1) = n$, bijection g is such that for some agent x, $u_{w_1}(x) < u_{w_2}(g(x))$. Since $u_{w_1}(x \in X) = n \leq u_{w_2}(y \in Y_2) = n$, $u_{w_1}(y \in Y_1) = m \leq u_{w_2}(y \in Y_1) = n$, $u_{w_1}(y \in Y_2) = m \leq u_{w_2}(x \in X) = m$, and $u_{w_1}(z \in Z) = u_{w_2}(z \in Z)$, bijection g is such that $u_{w_2}(x) \leq u_{w_2}(g(x))$ for all agents x. Since for all x, $g^{-1}(x) = g(x)$, $u_{w_1}(x) \leq u_{w_2}(g^{-1}(x))$ for all x. Therefore bijection g is a bidirectional upgrade from $pl(w_1)$ to $pl(w_2)$. This entails that $w_1 \not\succeq w_2$.

Since worlds w_2 and w_1 have the same properties that let us find a bidirectional upgrade g from the population of w_1 to the population of w_2, we can use the same method to find a bidirectional upgrade h from the population of w_2 to the population of w_1. It follows that $w_2 \not\succeq w_1$. Therefore w_1 and w_2 are ethically incomparable.

This establishes Result 6. If w_1 and w_2 have countable identical populations and there is an infinite set of agents x in $\langle w_1, w_2 \rangle$ such that $u_{w_1}(x) = n$ and $u_{w_2}(x) = m$ and there is an infinite set of agents y in $\langle w_1, w_2 \rangle$ such that $u_{w_2}(y) = n$ and $u_{w_1}(y) = m$, where $n < m$, and there are no other agents in w_1 or w_2, then w_1 and w_2 are incomparable.

Result 6 does not pick out all of the identical population world pairs that are incomparable by a four world argument. However, it is sufficient to show that if we accept Pareto, the qualitativeness of $\succcurlyeq$, the Permutation Principle and transitivity, then many infinite world pairs with identical populations are ethically incomparable.

3.3 The General Four World Result

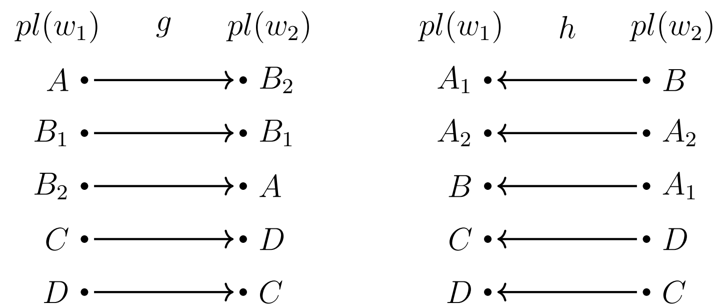
In the previous two sections I offered sufficient conditions for w_1 and w_2 to be incomparable by a four world argument if they have disjoint or identical populations. In this section I will formulate general conditions that entail that w_1 and w_2 can be shown to be incomparable by a four world argument, regardless of whether their populations are identical, disjoint, or overlapping. I will focus primarily on world pairs with overlapping populations (with shared and unshared agents) in this section, since these have yet to be discussed.

Overlapping world pairs contain some shared and some unshared agents. If the population of w_1 is a proper subset of the population of w_2 then w_1 and w_2 have overlapping populations. If w_1 and w_2 both contain some shared and some unshared agents then w_1 and w_2 are also overlapping populations. The following worlds are an overlapping population of the latter sort, where a dash (–) indicates that an agent does not exist:

	A	B	C	D
w_1	4	2	1	–
w_2	2	4	–	1

Figure 60: A world pair with overlapping populations

We can show that $w_1 \not\gtrless w_2$ by a four world argument in the case above because there is a strict upgrade from the shared agents of $\langle w_1, w_2 \rangle$ in w_1 to the shared agents of $\langle w_1, w_2 \rangle$ in w_1, and there is a utility isomorphism between the unshared agents of $\langle w_1, w_2 \rangle$ in w_1 and the unshared agents of $\langle w_1, w_2 \rangle$ in w_1. Consider the bijection g from the agents of world w_1 to the agents of world w_2 and the bijection h from the agents of world w_2 to the agents of world w_1, where we divide the agents in set A into the infinite sets A_1 and A_2 and we divide the agents in set B into the infinite sets B_1 and B_2.



Let $\langle w_3, w_4 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under bijection g and let $\langle w_5, w_6 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under bijection h. The utilities of the agents of the pairs $\langle w_3, w_4 \rangle$ and $\langle w_5, w_6 \rangle$ are as follows:

	A	B_1	B_2	C	D
w_1	4	2	2	1	–
w_2	2	4	4	–	1
w_3	2	2	4	–	1
w_4	4	4	2	1	–

	A_1	A_2	B	C	D
w_1	4	4	2	1	–
w_2	2	2	4	–	1
w_5	2	4	4	–	1
w_6	4	2	2	1	–

Figure 61: Permutations showing that the overlapping population world pair is incomparable

We can see that $\langle w_3, w_4 \rangle$ shows that $w_1 \not\succcurlyeq w_2$ by a four world argument and $\langle w_5, w_6 \rangle$ shows that $w_2 \not\succcurlyeq w_1$ by a four world argument. Therefore we can show that overlapping population pairs of worlds like w_1 and w_2 are incomparable by a four world argument.

We can now show the conditions that are sufficient to show that w_1 and w_2 are incomparable by a four world argument, whether their populations are identical, disjoint, or overlapping.[168]

Result 7: Four World $w_1 \not\succcurlyeq w_2$ (FW$\not\succcurlyeq$)

If there exists a bidirectional upgrade g from from $pl(w_1)$ to $pl(w_2)$ then $w_1 \not\succcurlyeq w_2$.

To establish Result 7, we once again suppose that there exists a bidirectional upgrade g from the population of w_1 to the population of w_2. Since g is such that $x \in pl(w_1)$ iff $g(x) \in pl(w_2)$ and $x \in pl(w_2)$ iff $g(x) \in pl(w_1)$, it follows that the shared populations of w_1 and w_2 are equinumerous and the unshared populations of w_1 and w_2 are equinumerous. Define a permutation f of the populations of w_1 and w_2 as follows: $f(x) = g(x)$ if x exists at w_1 and $f(x) = g^{-1}(x)$ if x exists at w_2 but does not exist at w_1. Since $f(x) = g(x)$ for all shared agents of w_1 and w_2, $f(x) = g(x)$ for all agents that exist in w_1 but not w_2 and $f(x) = g^{-1}(x)$ for all agents that exist in w_2 but not w_1, this results in a permutation the populations of $\langle w_1, w_2\rangle$. By the Permutation Principle, we can find a pair of worlds $\langle w_3, w_4\rangle$ that is a qualitative duplicate of $\langle w_1, w_2\rangle$ under g such that $w_2 \succ w_3$ and $w_4 \succcurlyeq w_1$ by Pareto.

To show that world w_2 is strictly better than w_3 by Pareto, note that since $x \in pl(w_1)$ iff $g(x) \in pl(w_2)$, worlds w_2 and w_3 have identical populations ($g(x)$ exists in w_2 iff $g(x)$ exists in w_3). Since $\langle w_3, w_4\rangle$ is a qualitative duplicate of $\langle w_1, w_2\rangle$ under g, the qualitative properties of each agent x in w_1 are the same as the qualitative properties of $g(x)$ in w_3. In particular, the utility of agent x in w_1 is the same as the utility of $g(x)$ at w_3. Since there is some agent x such that $u_{w_1}(x) < u_{w_2}(g(x))$ there will be some agent $g(x)$ in w_3 that contains strictly less utility than the agent $g(x)$ in w_2 does. Since all agents x in w_1 are such that $u_{w_1}(x) \leq u_{w_2}(g(x))$ there will be no agent $g(x)$ in w_3 that contains strictly greater utility than any agent $g(x)$ in w_2 does. Therefore world w_2 is strictly better than w_3 by Pareto.

[168] If w_1 and w_2 are disjoint then if there exists a g satisfying (d), there exists a g satisfying (e).

To show that $w_4 \succcurlyeq w_1$ by Pareto, note that since $x \in pl(w_2)$ iff $g(x) \in pl(w_1)$, worlds w_4 and w_1 have identical populations ($g(x)$ exists in w_4 iff $g(x)$ exists in w_1). Since $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g, the qualitative properties of each agent x in w_2 are the same as the qualitative properties of $g(x)$ in w_4. In particular, the utility of agent x in w_2 is the same as the utility of $g(x)$ at w_4. Since every agent x in w_2 are such that $u_{w_2}(g^{-1}(x)) \geq u_{w_1}(x)$, every agent $g(g^{-1}(x))$ in w_4 also has at least as much utility as agent x in w_1 has. Since $g(g^{-1}(x)) = x$, it follows that w_4 is at least as good as w_1 by Pareto.

Given transitivity, this results in the following familiar contradiction:

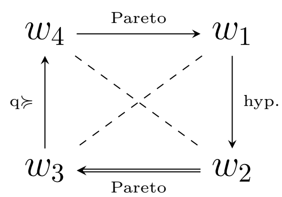


Figure 62: Transitivity violation in a world pair with a bidirectional upgrade

Therefore $w_1 \not\succcurlyeq w_2$, contrary to our assumption.

We can now summarize the general incomparability result as follows:

Result 8: Four World Incomparability (FW$\not\gtreqless$)

If there exists a bidirectional upgrade g from $pl(w_1)$ to $pl(w_2)$ and there exists a bidirectional upgrade g from $pl(w_2)$ to $pl(w_1)$ then $w_1 \not\gtreqless w_2$

If there exists a bidirectional upgrade g from $pl(w_1)$ to $pl(w_2)$ then $w_1 \not\succcurlyeq w_2$ by Result 7. If there exists a bidirectional upgrade g from $pl(w_2)$ to $pl(w_1)$ then $w_2 \not\succcurlyeq w_1$ by Result 7. Therefore if there exists a bidirectional upgrade g from $pl(w_1)$ to $pl(w_2)$ and there exists a bidirectional upgrade g from $pl(w_2)$ to $pl(w_1)$ then w_1 and w_2 are ethically incomparable.

It follows from this that the following overlapping population pairs can be shown to be incomparable by a four world argument:

Result 9: A Sufficient Condition for Overlapping Pair Incomparability

If w_1 and w_2 have countable, non-empty populations and the unshared population of $\langle w_1, w_2 \rangle$ meet the conditions of Result 3 and the shared population of $\langle w_1, w_2 \rangle$ meet the conditions of Result 6 then w_1 and w_2 are incomparable.

To show this, let g be a bijection from the shared agents of $\langle w_1, w_2 \rangle$ that exist in w_1 to the shared agents of $\langle w_1, w_2 \rangle$ that exist in w_2 that has the same properties as the bijection between identical populations in Result 6. Let h be a bijection from the unshared agents of $\langle w_1, w_2 \rangle$ that exist in w_1 to the shared agents of $\langle w_1, w_2 \rangle$ that exist in w_2 that has the same properties as the bijection between identical populations in Result 3 (note that if w_1 and w_2 contain no shared agents then g is just the empty function). And let h^{-1} be a bijection from the unshared agents of $\langle w_1, w_2 \rangle$ that exist in w_2 to the shared agents of $\langle w_1, w_2 \rangle$ that exist in w_2 that is the inverse of h (note that if w_1 and w_2 contain no unshared agents then h and h^{-1} are just the empty function). Let f be the composite bijection $g \circ h \circ h^{-1}$. The bijection f is a permutation of the population of $\langle w_1, w_2 \rangle$. By Result 3 and Result 6, a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f – $\langle w_3, w_4 \rangle$ – will be such that $w_2 \succ w_3$ by Pareto and $w_4 \succ w_1$ by Pareto. Therefore $w_1 \not\succeq w_2$ by transitivity. We can use the same method to find a permutation of $\langle w_1, w_2 \rangle$ to show that $w_2 \not\succeq w_1$. Therefore w_1 and w_2 are incomparable.

We can also extend Result 9 to world pairs that have countable, non-empty populations whose shared population meets the conditions of Result 6 *or* whose unshared population meets the conditions of Result 9 as long as there is a utility isomorphism between the remaining agents of world w_1 and world w_2. If this is the case then if the shared population of $\langle w_1, w_2 \rangle$ meet the conditions of Result 6 we can define g as above and if the unshared population of $\langle w_1, w_2 \rangle$ meet the conditions of Result 3 then we can define h and h^{-1} as above, and all remaining agents x in w_1 are mapped by g_2 to agents with isomorphic utility in w_2 and all remaining agents x in w_2 are mapped by g_2^{-1} to an agent with isomorphic utility in

w_1.[169] This is exactly what we did in the case of w_1 and w_2 at the beginning of this section.

I have now formulated conditions that are sufficient for there to be a four world argument showing that $w_1 \not\gtrless w_2$, regardless of whether w_1 and w_2 have identical, disjoint, or overlapping populations. I have also shown that many identical, disjoint, and overlapping population pairs are incomparable by a four world argument. In the next section, however, I will show that not all worlds that are incomparable by Pareto, transitivity, the qualitativeness of $\succcurlyeq$, and the Permutation Principle are incomparable by a four world argument. I introduce a new argument – the *cyclic argument* – that can be used to show that an even broader class of world pairs whose incomparability is entailed by these four axioms.

3.4 The Cyclic Result

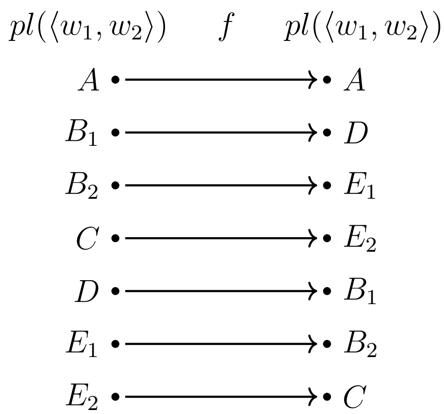
Not all world pairs that are incomparable by Pareto, transitivity, the qualitativeness of $\succcurlyeq$, and the Permutation Principle are incomparable by a four world argument. It is easiest to show this by working through an example. Suppose that the agents in w_1 and w_2 have the following utility levels, where the utility levels of shared agents are in bold:

	A	B	C	D	E
w_1	**1**	1	0	–	–
w_2	**0**	–	–	1	0

Figure 63: Overlapping world pair with distinct utilities at shared population

It is possible to find a bidirectional upgrade from $pl(w_2)$ to $pl(w_1)$. We divide the set B into two infinite sets of agents, B_1 and B_2, and we divide the set E into two infinite sets of agents, E_1 and E_2. We then permute the population of $\langle w_1, w_2\rangle$ by the following bijection:

[169] This means that all overlapping world pairs like the pair $\langle w_1, w_2\rangle$ discussed at the beginning of this section, which had a shared population that met the conditions of Result 2 and an isomorphic unshared population, are ethically incomparable.



Let $\langle w_3, w_4 \rangle$ be a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f.[170] These pairs are as follows:

	A	B_1	B_2	C	D	E_1	E_2
w_1	**1**	1	1	0	-	-	-
w_2	**0**	-	-	-	1	0	0
w_3	**1**	-	-	-	1	1	0
w_4	**0**	1	0	0	-	-	-

Figure 64: Permutation showing that $w_2 \not\succcurlyeq w_2$ in pair with distinct utilities at shared population

We can see that $w_3 \succ w_2$ by Pareto and $w_1 \succ w_4$ by Pareto. Since $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f, $w_2 \succcurlyeq w_1$ iff $w_4 \succcurlyeq w_3$. Therefore $w_2 \not\succcurlyeq w_1$ by transitivity.

It is not possible to find a bidirectional upgrade from $pl(w_1)$ to $pl(w_2)$, however. To see why, let g be a bijection that satisfies conditions (a) and (b) of FW$\not\succcurlyeq$, which state that $x \in pl(w_1)$ iff $g(x) \in pl(w_2)$ and $x \in pl(w_2)$ iff $g(x) \in pl(w_1)$. This means that g must map all agents in set a onto another agent in set a. But since the agents in set a all have strictly greater utility in w_1 than in w_2, we cannot find a qualitative duplicate of $\langle w_1, w_2 \rangle$ under any bijection that satisfies conditions (a) and (b) that is such that $w_2 \succ w_3$ by Pareto and $w_4 \succ w_1$ by Pareto. Since agents in A must be mapped onto agents in A by such a bijection, all agents in A will

[170] For every agent x in w_1, $f(x)$ exists in w_2. Therefore w_2 and w_3 will have identical populations. For every agent x in w_2, $f(x)$ exists in w_1. Therefore w_4 and w_1 will have identical populations.

have utility 1 in w_3 and utility 0 in w_4. Therefore w_2 cannot be strictly better than w_3 by Pareto and w_4 cannot be strictly better than w_1 by Pareto. So we cannot show that these two worlds are incomparable with Pareto using a four-world argument.

This shows that we cannot always find a single permutation of the population of $\langle w_1, w_2 \rangle$ that can be used to show that $w_1 \not\succeq w_2$ by a four world argument. If we use the full strength of the Permutation Principle, however, we can sometimes find a sequence of qualitative duplicate pairs $\langle w_3, w_4 \rangle$ and $\langle w_5, w_6 \rangle$ such that w_2 and w_3 have identical populations and w_4 and w_5 have identical populations, and w_6 and w_1 have identical populations, even though w_1 and w_4 have distinct populations. And if there is some sequence of qualitative duplicate pairs $\langle w_1, w_2 \rangle$, $\langle w_3, w_4 \rangle$, $\langle w_5, w_6 \rangle, ..., \langle w_{n-1}, w_n \rangle$ such that $w_2 \succ w_3$ by Pareto, $w_4 \succ w_5$ by Pareto, $..., w_n \succ w_1$ by Pareto, then it follows that $w_1 \not\succeq w_2$.

We can illustrate how this kind of 'cyclic argument' works using the example above. First, note that $w_2 \not\succeq w_1$ by a four world argument. This follows from Result 9, since there is a bidirectional upgrade from the shared agents of w_2 to the shared agents of w_1 and there is a strict upgrade from the unshared agents of w_2 to the unshared agents of w_1. However, Result 9 does not entail that $w_1 \not\succeq w_2$, since there is no bidirectional upgrade from the shared population of w_1 to the shared population of w_2 in this example.

To construct a cyclic argument showing that $w_1 \not\succeq w_2$, we first divide the agents in C in $\langle w_1, w_2 \rangle$ into three infinite sets C_1, C_2, and C_3 and the agents in D in $\langle w_1, w_2 \rangle$ into three infinite sets D_1, D_2, and D_3. Let J_1, J_2, and J_3 be three countably infinite sets of agents that don't exist in either w_1 or w_2. The Permutation Principle states that for any bijection f from the population of $\langle w_1, w_2 \rangle$ onto any other population, there exists a world pair $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f. We can find a bijection from the population of a world pair that contains the agents of $\langle w_1, w_2 \rangle$ plus countably infinitely many J_1, J_2, and J_3 agents that do not exist in $\langle w_1, w_2 \rangle$. Therefore if f is a permutation of the population of

$\langle w_1, w_2\rangle$ plus infinitely many J_1, J_2, and J_3 agents then, by the Permutation Principle, we can find a qualitative duplicate of $\langle w_1, w_2\rangle$ under f.

As I stated above, a cyclic argument permutes the population of $\langle w_1, w_2\rangle$ in stages. To show that $w_1 \not\succcurlyeq w_2$ we need to find a permutation sequence such that $\langle w_1, w_2\rangle, \langle w_3, w_4\rangle, ..., \langle w_{n-1}, w_n\rangle$ are qualitative duplicates such that $w_2 \succ w_3, w_4 \succ w_5, ..., w_{n-2} \succ w_{n-1}; w_n \succ w_1$ by Pareto. In this case, the following bijections f from the agents of $\langle w_1, w_2\rangle$ plus J_1, J_2, and J_3 agents can be used to show that $w_1 \not\succcurlyeq w_2$ by such a cyclic argument:

$$\begin{array}{ccc} & f & \\ A & \longrightarrow & D_1 \\ B & \longrightarrow & D_3 \\ C_1 & \longrightarrow & A \\ C_2 & \longrightarrow & D_2 \\ C_3 & \longrightarrow & E \\ D_1 & \longrightarrow & C_1 \\ D_2 & \longrightarrow & J_1 \\ D_3 & \longrightarrow & J_3 \\ E & \longrightarrow & J_2 \\ J_1 & \longrightarrow & C_2 \\ J_2 & \longrightarrow & C_3 \\ J_3 & \longrightarrow & B \end{array}$$

As before, the use of set notation here is just a notational convenience. If f maps set X onto set Y this means that for all $x \in X$ and for all $y \in Y$, $f(x_i) = y_i$ and $f^{-1}(y_i) = x_i$. Let f^n be the reapplication of f n-many times, so that $f^3(x) = f(f(f(x)))$ and $f^0(x) = x$. In the case of cyclic arguments we add a further stipulation, namely that if for some n, the the composite bijection f^n maps set X back onto itself, then for all $x \in X$, $f^n(x_i) = x_i$. In other words, f^n maps all of the individual agents in x back onto themselves.

To help keep track of which worlds are qualitative duplicates of which, I will stop using terms like w_3 and w_4 to refer to qualitative duplicates of w_1 and w_2. Instead I will use $\mathbf{f}(w_i)$ to refer to a world that is a qualitative duplicate of w_i but whose population has been permuted by f. I will also use $\mathbf{f}(\langle w_i, w_j \rangle)$ to refer to a world pair that is a qualitative duplicate of $\langle w_i, w_j \rangle$ but whose population has been permuted by f.

Consider permutation f given above. For all agents x, if x exists in $\mathbf{f^n}(w_1)$, then $f(x)$ exists at $\mathbf{f^n}(w_2)$ under this permutation. Therefore for all applications of f, $\mathbf{f^{n+1}}(w_1)$ will have a population that is identical to $\mathbf{f^{n-1}}(w_2)$. The permutation f given above and the utility of the agents at world pairs $\mathbf{f}(\langle w_1, w_2 \rangle)$ and $\mathbf{f}(\mathbf{f}(\langle w_1, w_2 \rangle))$ are as follows:

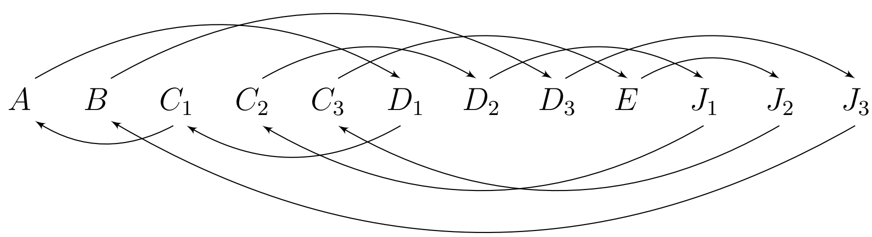


Figure 65: Permutations $\mathbf{f}(\langle w_1, w_2 \rangle)$ and $\mathbf{f}(\mathbf{f}(\langle w_1, w_2 \rangle))$

	A	B	C_1	C_2	C_3	D_1	D_2	D_3	E	J_1	J_2	J_3
w_1	**1**	1	0	0	0	–	–	–	–	–	–	
w_2	**0**	–	–	–	–	1	1	1	0	–	–	–
$\mathbf{f}(w_1)$	0	–	–	–	–	**1**	0	1	0	–	–	–
$\mathbf{f}(w_2)$	–	–	1	–	–	**0**	–	–	–	1	0	1
$\mathbf{f}(\mathbf{f}(w_1))$	–	–	**1**	–	–	0	–	–	–	0	0	1
$\mathbf{f}(\mathbf{f}(w_2))$	1	1	**0**	1	0	–	–	–	–	–	–	–

Figure 66: Agent utilities in $\mathbf{f}(\langle w_1, w_2 \rangle)$ and $\mathbf{f}(\mathbf{f}(\langle w_1, w_2 \rangle))$

Suppose, for reductio, that $w_1 \succcurlyeq w_2$. Since $\succcurlyeq$ is a necessary qualitative relation and $\mathbf{f}(\langle w_1, w_2 \rangle)$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f, it follows that $\mathbf{f}(w_1) \succcurlyeq \mathbf{f}(w_2)$. And since $\mathbf{f}(\mathbf{f}(\langle w_1, w_2 \rangle))$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f^2 (under $f \circ f$) it follows that $\mathbf{f}(\mathbf{f}(w_1)) \succcurlyeq \mathbf{f}(\mathbf{f}(w_2))$. But $w_2 \succ \mathbf{f}(w_1)$ by Pareto and $\mathbf{f}(w_2) \succ \mathbf{f}(\mathbf{f}(w_1))$ by Pareto and $\mathbf{f}(\mathbf{f}(w_2)) \succ w_1$ by Pareto. This results in the following violation of transitivity:

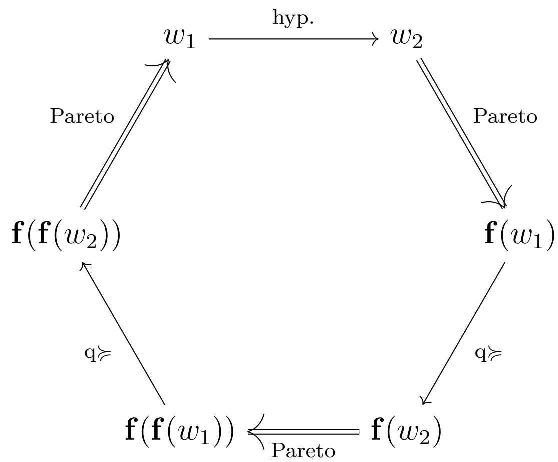


Figure 67: Transitivity violation in a pair with distinct utilities at shared population

Therefore if we accept Pareto, the Permutation Principle, the qualitativeness of $\succcurlyeq$ and transitivity, then $w_1 \not\succcurlyeq w_2$. And since world $w_2 \not\succcurlyeq w_1$ by a four world argument, it follows that world w_1 and world w_2 are ethically incomparable ($w_1 \not\gtrless w_2$).

Let $\mathbb{P}$ denote the set of all possible people. We can now formulate conditions that are sufficient to show $w_1 \not\succcurlyeq w_2$ by a cyclic argument.

Result 10: Cyclic Permutation $w_1 \not\succcurlyeq w_2$ (CP$\not\succcurlyeq$)

If there exists an n-cyclic permutation of $\mathbb{P}$ that is a strict upgrade from the population of w_1 to the population of w_2 and is such that, for all x in w_1, $f^{n+1}(x) = x$, then $w_1 \not\succcurlyeq w_2$

Claim: If there is an n-cyclic permutation of $\mathbb{P}$ that satisfies these conditions then $w_1 \not\succcurlyeq w_2$.

To show this, suppose for reductio that $w_1 \succcurlyeq w_2$ and there is an n-cyclic permutation f of $\mathbb{P}$ that is a strict upgrade from the population of w_1 to the population of w_2 and is such that, for all x in w_1, $f^{n+1}(x) = x$. By the Permutation Principle, there exists a world pair $\mathbf{f}(\langle w_1, w_2\rangle))$ that is a qualitative duplicate of $\langle w_1, w_2\rangle$ under f, and there exists a world pair $\mathbf{f}(\mathbf{f}(\langle w_1, w_2\rangle)))$ that is a qualitative duplicate of $\mathbf{f}(\langle w_1, w_2\rangle))$ under f and so on for any iteration of f. Since $w_1 \succcurlyeq w_2$, it follows from the qualitativeness of $\succcurlyeq$ that $\mathbf{f}(w_1) \succcurlyeq \mathbf{f}(w_2)$, and $\mathbf{f}(\mathbf{f}(w_1)) \succcurlyeq \mathbf{f}(\mathbf{f}(w_2))$, and so on for any iteration of f.

Since f is a strict upgrade from w_1 to w_2 it follows that for all agents x, x exists in w_1 if and only if $f(x)$ exists in w_2, and $f(x)$ exists in $\mathbf{f}(w_1)$ if and only if $f(f(x))$ exists in $\mathbf{f}(w_2)$, and so on. Therefore w_2 and $\mathbf{f}(w_1)$ have identical populations, $\mathbf{f}(w_2)$ and $\mathbf{f}(\mathbf{f}(w_1))$ have identical populations, and so on for each qualitative duplicate of $\langle w_1, w_2 \rangle$ under an iterate of f.

Since f is a strict upgrade from the population of w_1 to the population of w_2 it follows that for all agents $f(x)$ in w_2, $u_{w_2}(f(x)) \geq u_{\mathbf{f}(w_1)}(f(x))$ and for all agents $f(f(x))$ in $\mathbf{f}(w_2)$, $u_{\mathbf{f}(w_2)}(f(f(x))) \geq u_{\mathbf{f}(\mathbf{f}(w_1))}(f(f(x)))$ and so on for each qualitative duplicate of $\langle w_1, w_2 \rangle$ under an iterate of f. In addition, for some agent $f(x)$ in w_2, $u_{w_2}(f(x)) > u_{\mathbf{f}(w_1)}(f(x))$ and for some agent $f(f(x))$ in $\mathbf{f}(w_2)$, $u_{\mathbf{f}(w_2)}(f(f(x))) > u_{\mathbf{f}(\mathbf{f}(w_1))}(f(f(x)))$ and so on for each qualitative duplicate of $\langle w_1, w_2 \rangle$ under an iterate of f. Therefore $w_2 \succ \mathbf{f}(\mathbf{w_1})$ by Pareto, $\mathbf{f}(\mathbf{w_2}) \succ \mathbf{f}(\mathbf{f}(w_1))$ by Pareto, $\mathbf{f}(\mathbf{f}(w_2)) \succ \mathbf{f}(\mathbf{f}(\mathbf{f}(w_1)))$ by Pareto, and so on for each qualitative duplicate of $\langle w_1, w_2 \rangle$ under an iterate of f.

Finally, since there is some n such that $f^{n+1}(x) = x$, and x exists in w_1 iff $f(x)$ exists in w_2 it follows that the population of $\mathbf{f^n}(w_2)$ is identical to the population of w_1. Since for each iterate of f, $f^n(w_2) \succ f^{n+1}(w_1)$ by Pareto, it follows that $\mathbf{f^n}(w_2) \succ w_1$ by Pareto. This results in the following violation of transitivity, where there may exist any finite number of iterations of f between f^2 and f^n such that $\mathbf{f^i}(w_1) \succcurlyeq \mathbf{f^i}(w_2)$ by the qualitativeness of $\succcurlyeq$ and $\mathbf{f^i}(w_2) \succ \mathbf{f^{i+1}}(w_1)$ by Pareto:

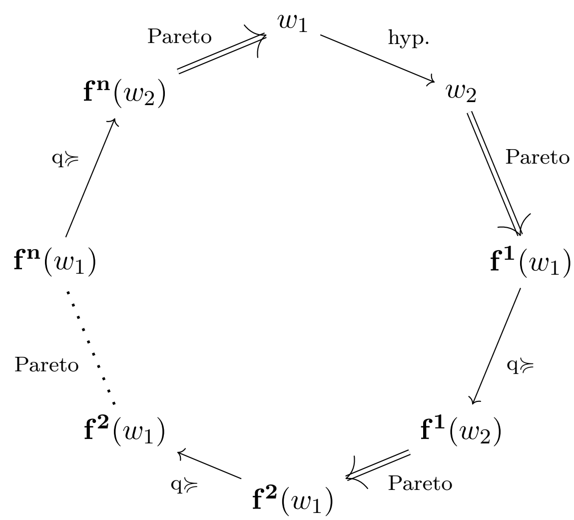


Figure 68: Cyclic transitivity violation

Since this violates transitivity, $w_1 \not\succcurlyeq w_2$, contrary to our assumption. Therefore if there exists an n-cyclic permutation of $\mathbb{P}$ that is a strict upgrade from the population of w_1 to the population of w_2 and is such that, for all x in w_1, $f^{n+1}(x) = x$, then $w_1 \not\succcurlyeq w_2$

For completeness, we can state the cyclic incomparability result.

Result 11: Cyclic Permutation Incomparability (CP$\not\gtrless$)

If (i) there exists an n-cyclic permutation of $\mathbb{P}$ that is a strict upgrade from the population of w_1 to the population of w_2 and is such that, for all x in w_1, $f^{n+1}(x) = x$, then $w_1 \not\succcurlyeq w_2$, and (ii) there exists an n-cyclic permutation of $\mathbb{P}$ that is a strict upgrade from the population of w_2 to the population of w_1 and is such that, for all x in w_2, $f^{n+1}(x) = x$, then $w_2 \not\succcurlyeq w_1$.

If (i) holds then $w_1 \not\succcurlyeq w_2$ by Result 10 and if (ii) holds then $w_2 \not\succcurlyeq w_1$ by Result 10. Therefore w_1 and w_2 are incomparable ($w_1 \not\gtrless w_2$).

Notice that both of the conditions of Result 10 hold in the example of a cyclic argument offered earlier. In that case, f was a strict upgrade from the population of w_1 to the population of w_2, since f mapped all agents in w_1 onto an agent in w_2 with at least as much utility, and f mapped the agents in C_2 to agents in D_2, and all of the agents in C_1

in w_1 have strictly less utility than all of the agents in D_1 in w_2. Moreover, for all x in w_1, $f^4(x) = f(f(f(f(x))) = x$. Therefore $w_1 \not\succcurlyeq w_2$ by a 3-cyclic permutation, as we saw.

Before offering an example of a sufficient condition for cyclic incomparability, we can illustrate how we construct a 3-cycle permutation that satisfies the conditions of Result 10 using a simple example. Consider a pair of worlds $\langle w_1, w_2 \rangle$ that are as follows:

	A	B	C
w_1	**4**	2	–
w_2	**2**	–	4

Figure 69: World pair requiring a 3-cycle permutation

We cannot show that $w_1 \not\succcurlyeq w_2$ or that $w_2 \not\succcurlyeq w_1$ by a four world argument because there is no bidirectional upgrade from the agents of w_1 to the agents of w_2 or from the agents of w_2 to the agents of w_1. But we can show that these worlds are incomparable by a cyclic argument.

To show that $w_1 \not\succcurlyeq w_2$ by a cyclic argument, we have to find a strict upgrade from the population of w_1 to the population of w_2. In this case, we divide the agents in set B into two infinite sets, B_1 and B_2, and we can divide the agents in set C into two infinite sets, C_1 and C_2. We can construct a strict upgrade from the population of w_1 to the population of w_2 by mapping the agents in A (at utility 4 in w_1) onto agents in C_1 (at utility 4 in w_2), all of the agents in B_1 (at utility 2 in w_1) onto agents in A (at utility 2 in w_2), and all of the agents in B_2 (at utility 2 in w_1) onto agents in C_2 (at utility 4 in w_2):

$$\begin{array}{ccc} A & \longrightarrow & C_1 \\ B_1 & \longrightarrow & A \\ B_2 & \longrightarrow & B_1 \end{array}$$

We now need to map C_1 agents and C_2 agents. Since we are looking for a cycle such that for some n, $f^n(x) = x$, we can map C_1 agents onto B_1 agents to complete the cycle above: $A \to C_1 \to B_1 \to A$. We could map C_2 agents onto B_2 agents, but this would result in one even-numbered cycle (the previous one) and one odd-numbered cycle: $C_2 \to B_2 \to C_2$. As a result, the smallest n such that $f^n(x) = x$ is 6, and so we would require six sets of worlds (five iterations of f) to show that $w_1 \not\succcurlyeq w_2$. Instead, we can introduce a set of agents D_1 that don't exist in w_1 or w_2 and cycle the C_2 agents through these agents, resulting in the even-numbered cycle $B_2 \to C_2 \to D_2 \to B_2$. We can therefore complete the permutation:

$$\begin{array}{c} C_1 \longrightarrow B_1 \\ C_2 \longrightarrow D_1 \\ D_1 \longrightarrow B_2 \end{array}$$

Using the permutation above will show that $w_1 \not\succcurlyeq w_2$ by a cyclic argument using only three pairs of worlds. We can depict the permutation above and the world pairs $\mathbf{f}(\langle w_1, w_2 \rangle)$ and $\mathbf{f}(\mathbf{f}(\langle w_1, w_2 \rangle))$ as follows:

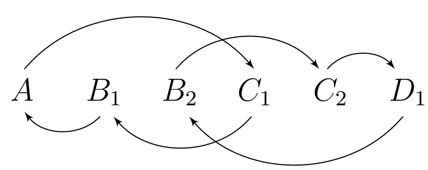


Figure 70: A 3-cycle upgrade permutation

	A	B_1	B_2	C_1	C_2	D_1
w_1	**4**	2	2	–	–	–
w_2	**2**	–	–	4	4	–
$\mathbf{f}(w_1)$	2	–	–	**4**	2	–
$\mathbf{f}(w_2)$	–	4	–	**2**	–	4
$\mathbf{f}(\mathbf{f}(w_1))$	–	**4**	–	2	–	2
$\mathbf{f}(\mathbf{f}(w_2))$	4	**2**	4	–	–	–

Figure 71: Agent utilities under the 3-cycle upgrade permutation

We can see that if $w_1 \succcurlyeq w_2$ then by the qualitativeness of $\succcurlyeq$ and Pareto it follows that $w_1 \succcurlyeq w_2 \succ \mathbf{f}(w_1) \succcurlyeq \mathbf{f}(w_2) \succ \mathbf{f}(\mathbf{f}(w_1)) \succcurlyeq \mathbf{f}(\mathbf{f}(w_2)) \succ w_1$, which violates transitivity.

We can also show that $w_2 \not\succcurlyeq w_1$ by a cyclic argument. One way to do this is to find a 'strict downgrade' from the population of world w_1 to the population of world w_2 – a bijection g from $pl(w_1)$ to $pl(w_2)$ such that for all agents x in w_1, $u_{w_1}(x) \geq u_{w_2}(g(x))$ and for some agent x in w_1, $u_{w_1}(x) > u_{w_2}(g(x))$. In this case we can do this without having to go through any agents that don't exist in w_1 or w_2. First we divide the set of agents A into two infinite sets A_1 and A_2. Given this, the following is a strict downgrade from $pl(w_1)$ to $pl(w_2)$:

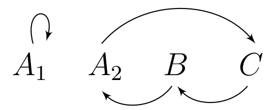


Figure 72: A 3-cycle downgrade permutation

We can depict the world pairs $\mathbf{f}(\langle w_1, w_2 \rangle)$ and $\mathbf{f}(\mathbf{f}(\langle w_1, w_2 \rangle))$ as follows:

	A_1	A_2	B	C
w_1	**4**	**4**	2	–
w_2	**2**	**2**	–	4
$\mathbf{f}(w_1)$	**4**	2	–	**4**
$\mathbf{f}(w_2)$	**2**	–	4	**2**
$\mathbf{f}(\mathbf{f}(w_1))$	**4**	–	**4**	2
$\mathbf{f}(\mathbf{f}(w_2))$	**2**	4	**2**	–

Figure 73: Agent utilities under the 3-cycle downgrade permutation

We can see that if $w_2 \succcurlyeq w_1$ then by the qualiativeness of $\succcurlyeq$ and Pareto it follows that $w_2 \succcurlyeq w_1 \succ \mathbf{f}(\mathbf{f}(w_2)) \succcurlyeq \mathbf{f}(\mathbf{f}(w_1)) \succ \mathbf{f}(w_2) \succcurlyeq \mathbf{f}(w_1) \succ w_2$, which violates transitivity.

In all of the examples of cyclic arguments I have considered thus far, it has been possible to find a single function f such that for some n, $f^n(x) = x$. It is worth noting, however, that it is not necessary that we use the same function in a cyclic argument: we could instead use a sequence of entirely distinct functions f_1, f_2,..., f_n. Since 'is a qualitative duplicate of' is a transitive relation, if $\mathbf{f_1}(\langle w_1, w_2 \rangle)$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under f_1

and $\mathbf{f_2}(\mathbf{f_1}(\langle w_1, w_2\rangle))$ is a qualitative duplicate of $\mathbf{f_1}(\langle w_1, w_2\rangle)$ under f_2, then $\mathbf{f_2}(\mathbf{f_1}(\langle w_1, w_2\rangle))$ is a qualitative duplicate of $\langle w_1, w_2\rangle$ (it is a duplicate under the composite function $f_1 \circ f_2$). It remains an open question whether there are any cases in which a cyclic proof of incomparability requires the use of distinct functions rather than a single iterated function. But if there are a sequence of distinct functions f_1, f_2,..., f_n such that $f_1 \circ f_2 \circ, ..., \circ f_n = n$ and some function f_i in the sequence is a strict upgrade from the population of $\mathbf{f_{i-1}}(w_1)$ to the population of $\mathbf{f_{i-1}}(w_2)$ and all functions f_i in the sequence are a weak upgrade from the population of $\mathbf{f_{i-1}}(w_1)$ to the population of $\mathbf{f_{i-1}}(w_2)$, then $w_1 \not\succcurlyeq w_2$ by a cyclic argument.

Summary

In this chapter I have shown that if we accept Pareto, the qualitativeness of $\succcurlyeq$, the Permutation Principle, and transitivity, then many classes of worlds are ethically incomparable.

These results do not show that all infinite world pairs that are not comparable by Pareto are ethically incomparable. They do not entail that world w_1 containing infinitely many agents at utility level 4 and infinitely many agents at utility level 1 is incomparable with a world containing infinitely many agents at utility 3 and infinitely many agents at utility 2. They also do not extend to world pairs in which there is a finite population of agents in w_1 with utility strictly greater than or strictly less than the utility of any agent in w_2. Nor do they apply to world pairs with uncountable populations or with non-equinumerous populations. In the next chapter, I will offer a modest extension of Pareto and will use this to extend the results of this chapter. I then present problems for more ambitious extensions of Pareto.

Chapter 4

Extending the Incomparability Results

In the previous chapter I showed that many infinite world pairs are incomparable. Although some world pairs can be shown to be comparable or incomparable, there are still 'open question' world pairs: world pairs that are not comparable by the principles that we endorse but are also not demonstrably incomparable. Ideally, we eventually reach a point at which no open question world pairs remain. In this chapter, I explore some possible extensions of the results of this chapter that could reduce the number of open question world pairs. In section 4.1 I offer a modest extension of Pareto. In section 4.2 I show that this extension can be made stronger if we supplement it with addition principles, resulting in the Weak People Criterion and argue that adopting the Weak People Criterion entails that a larger class of worlds can be shown to be incomparable by four world and cyclic arguments than in the previous chapter. Finally, in section 4.3 I consider alternatives to the Weak People Criterion that are intuitively plausible and argue that these have unacceptable consequences and therefore ought to be rejected.

4.1 Weak Catching-Up

In the previous chapters I showed that if we accept Pareto, transitivity, the Permutation Principle, and the qualitativeness of $\succsim$, then we can show that many infinite worlds are ethically incomparable. Pareto is a relatively limited ethical principle, however. The principle can only rank world pairs that have shared populations and they cannot rank any worlds pairs $\langle w_1, w_2 \rangle$ if there is at least one agent with lower utility in w_1 than in w_2 and there is at least one agent with lower utility in w_2 than in w_1.

It seems clear that some world pairs that fail to meet the conditions for Pareto comparability are not ethically incomparable. Consider an identical population world pair $\langle w_1, w_2 \rangle$ in which infinitely many agents are better off in w_1 than in w_2 by an infinite amount of utility and only finitely many agents are better off in w_2 than in w_1 by a finite amount of utility. We can give an example of such a world pair, where # is used to indicate the cardinality of the set of agents and the cardinality of any countably infinite set of agents is $\aleph_0$:

	A	B
#	$\aleph_0$	1
w_1	2	0
w_2	0	1

Figure 74: Identical population world pair with distinct upper utility levels

World w_1 contains infinitely many agents in set A at utility level 2 and a single agent in set B at utility level 0. World w_2 contains infinitely many agents in set A at utility level 0 and a single agents in set B at utility level 1. It seems clear that world w_1 is strictly better than world w_2. The sum of utility at world w_1 is positively infinite and infinitely many agents are better off in w_1 than they are in w_2. But because there is a single agent that has lower utility in w_1 than in w_2, these two worlds are not Pareto comparable.

Pareto also cannot compare identical population world pairs $\langle w_1, w_2 \rangle$ in which the total utility of both worlds is finite if at least one agent has more utility in w_1 than in w_2 and at least one agent has more utility in w_2 than in w_1. Worlds may have finite total utility either because they contain finitely many agents or because they contain infinitely many agents and the series of utility of these agents converges absolutely on some finite value (such that the sequence of agent utilities converges on zero). For example, suppose that w_1 and w_2 have identical populations. World w_1 contains a single agent at utility level $\frac{3}{4}$ and infinitely many agents at utility 0. World w_2 contains infinitely many agents that have utility levels that follow the geometric series $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, ...$ Therefore the total utility of world w_1 is $\frac{3}{4}$ and the total utility of world w_2 is $\sum(\frac{1}{2})^n = 1$. It therefore seems that world w_1 is strictly better than world w_2 but these two worlds are not Pareto comparable. Ideally we would want our preferred principles for ranking infinite worlds to entail our preferred principles for ranking finite worlds we prefer whenever two worlds contain finite utility.

We can formulate a modest extension of the Pareto principle which entails that if $\langle w_1, w_2 \rangle$ is an identical population world pair and the agents of w_1 are better off by a finite or infinite amount of utility than the agents in w_2, then w_1 is strictly better than w_2. In order to do this, it will be helpful to introduce a new concept of 'extended convergence': an extension of the usual concept of absolute convergence introduced by Artnzenius [6, p. 54]. If a sequence extendedly converges then it either converges absolutely on some finite value or it diverges to positive infinity or to negative infinity (but not both) under any ordering of its terms. Given this, we can first extend Pareto by endorsing the Weak Catching-Up principle defended by Lauwers and Vallentyne [148, p. 322], which I formulate as follows:

Weak Catching-Up

($\succcurlyeq$) *If* w_1 *and* w_2 *contain the same agents and for all possible enumerations of the agents of* w_1 *and* w_2*, the lower limit as* J *approaches infinity, of the sum of* $u_{w_1}(x_j) - u_{w_2}(x_j)$ *for agents 1 to* J *(i.e.* $\lim_{J\to\infty} \sum_{j=1,\ldots,J}(u_{w_1}(x_j) - u_{w_2}(x_j)))$ *is extendedly convergent and* ≥ 0 *then* $w_1 \succcurlyeq w_2$

($\succ$) *If* w_1 *and* w_2 *contain the same agents and for all possible enumerations of the agents of* w_1 *and* w_2*, the lower limit as* J *approaches infinity, of the sum of* $u_{w_1}(x_j) - u_{w_2}(x_j)$ *for agents 1 to* J *(i.e.* $\lim_{J\to\infty} \sum_{j=1,\ldots,J}(u_{w_1}(x_j) - u_{w_2}(x_j)))$ *is extendedly convergent and* > 0 *then* $w_1 \succ w_2$

Weak Catching-Up satisfies both of the desiderata outlined above. If w_1 and w_2 contain the same agents and in w_1 infinitely many agents have utility 2 and all remaining agents have utility 0, and in w_2 finitely many agents have utility 1 and all remaining agents have utility 0, then the difference in utilities across the agents of w_1 and w_2 extendedly converges on ∞. Therefore $w_1 \succ w_2$ by Weak Catching-Up. And if w_1 and w_2 contain the same agents and in w_1 there is a single agent with utility $\frac{3}{4}$ and infinitely many agents with utility 0 and in w_2 there are infinitely many agents with $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$ then the difference in utilities across the agents of w_1 and w_2 extendedly converges on $-\frac{1}{4}$. Therefore $w_2 \succ w_1$ by Weak Catching-Up.

If we endorse a principle that is stronger than Pareto then we increase the number of infinite world pairs that we are able to compare. At the same time, endorsing stronger principles often increases the number of world pairs that can be shown to be incomparable by a four world or cyclic argument. If we accept the Weak Catching-Up principle then we can extend the incomparability results of the previous chapter to world pairs that would meet the conditions of Result 8 or Result 11 were it not for the fact that there exists a finite population of agents in w_1 with utility strictly greater than or strictly less than the utility of any agent in w_2 or there exists a finite population of agents in w_2 with utility strictly greater than or strictly less than the utility of any agent in w_1. Consider the following world pair:

	A	B	C	D	E
#	$\aleph_0$	$\aleph_0$	1	1	1
w_1	2	1	4	-1	-1
w_2	1	2	-2	6	6

Figure 75: A pair with finitely many agents with utilities outside the accumulation interval

There is a bidirectional upgrade from the agents in w_1 to the agents in in w_2 and there is a bidirectional upgrade from the agents in w_2 to the agents in w_1 in this case. But we cannot find a bidirectional upgrade from w_1 to w_2 because there is no agent with utility less than or equal to -2 in w_1. And we cannot find a bidirectional upgrade from the population of w_2 to the population of w_1 because there is no agent with utility greater than or equal to 6 in w_1. So we cannot show that $w_1 \not\gtrless w_2$ by a four world argument using Pareto.

If we accept Weak Catching-Up, however, then we can show that w_1 and w_2 are incomparable by a four world argument. To do this, we just need to find a bijection g that is a bidirectional upgrade from w_1 to w_2 such that, for infinitely many agents x, $u_{w_1}(x) < u_{w_2}(x)$ and $\sum_x(u_{w_1}(x) - u_{w_2}(x))$ is negatively infinite, and for all remaining agents only finitely many are such that $u_{w_2}(x) < u_{w_1}(x)$.[171] To find such a bijection from $pl(w_1)$ to $pl(w_2)$ we divide the agents in set B into the infinite sets B_1 and B_2. A bijection from $pl(w_1)$ to $pl(w_2)$ that satisfies the conditions above is as follows:

[171]These conditions are stronger than what we need. All we actually require is that the total utility of the agents x in w_1 such that $u_{w_1}(x) < u_{w_2}(x)$ is less than or equal to the total utility of the agents x in w_2 such that $u_{w_2}(x) < u_{w_1}(x)$. But, as I will show, if there is a bidirectional upgrade from $pl(w_1)$ to $pl(w_2)$ and there is a bidirectional upgrade from $pl(w_2)$ to $pl(w_1)$ then we can find a bijection that satisfies these conditions.

	A	B_1	B_2	C	D	E
#	$\aleph_0$	$\aleph_0$	$\aleph_0$	1	1	1
w_1	2	1	1	4	-1	-1
w_2	1	2	2	-2	6	6
w_3	1	2	1	4	-1	-1
w_4	2	1	2	-2	6	6

Figure 76: A permutation showing $w_1 \not\succcurlyeq w_2$ by Weak Catching-Up

Suppose that $w_1 \succcurlyeq w_2$. Let the world pair $\langle w_3, w_4 \rangle$ be a qualitative duplicates of $\langle w_1, w_2 \rangle$ under the bijection above. Infinitely many agents in $B_2 \cup D \cup E$ have greater utility in w_2 than they do w_3 such that the total difference in utility between the agents in $B_2 \cup D \cup E$ in w_2 and the agents in $B_2 \cup D \cup E$ in w_3 is infinite. Finitely many agents in C have greater utility in w_3 than they do in w_2 and the total difference in utility between the agents in C in w_2 and the agents in C in w_3 is finite (6 utils). All remaining agents in w_2 – the agents in A – have the same utility levels in w_2 and w_3. Therefore $\sum_x (u_{w_2}(x) - u_{w_3}(x))$ extendedly converges to a positive value – since it diverges to ∞ – and so $w_2 \succ w_3$ by Weak Catching-Up. It is easy to see that $\sum_x (u_{w_4}(x) - u_{w_1}(x))$ also diverges to ∞ and so $w_4 \succ w_1$ by Weak Catching-Up. By the qualitativeness of the $\succcurlyeq$ relation, $w_3 \succcurlyeq w_4$. Therefore $w_1 \succcurlyeq w_2 \succ w_3 \succcurlyeq w_4 \succ w_1$. This violates transitivity and so $w_1 \not\succcurlyeq w_2$, contrary to our assumption.

It should be easy to see that we can find a bijection from $pl(w_2)$ to $pl(w_1)$ that also satisfies the conditions above. For completeness, this bijection is as follows:

	A_1	A_2	B	C	D	E
#	$\aleph_0$	$\aleph_0$	$\aleph_0$	1	1	1
w_1	2	2	1	4	-1	-1
w_2	1	1	2	-2	6	6
w_5	2	1	2	4	-1	-1
w_6	1	2	1	-2	6	6

Figure 77: A permutation showing $w_2 \not\succcurlyeq w_1$ by Weak Catching-Up

Suppose that $w_2 \succcurlyeq w_1$. Let the world pair $\langle w_5, w_6 \rangle$ be a qualitative duplicates of $\langle w_1, w_2 \rangle$ under the bijection above. Since $\sum_x (u_{w_5}(x) - u_{w_2}(x))$ diverges to ∞, $w_5 \succ w_2$ by Weak Catching-Up. Since $\sum_x (u_{w_1}(x) - u_{w_6}(x))$ diverges to ∞, $w_6 \succ w_1$ by Weak Catching-Up. By the qualitativeness of the $\succcurlyeq$ relation, $w_5 \succcurlyeq w_6$. Therefore $w_2 \succcurlyeq w_1 \succ w_6 \succcurlyeq w_5 \succ w_2$. This violates transitivity and so $w_2 \not\succcurlyeq w_1$, contrary to our assumption.

Weak Catching-Up lets us extend the comparability results to all world pairs $\langle w_1, w_2 \rangle$ such that infinitely many agents have greater utility in w_1 than in w_2 and infinitely many agents have greater utility in w_2 than in w_1 and there are finitely many agents in w_1 with utility strictly greater than or less than that of any agent in w_2 and/or there are are finitely many agents in w_2 with utility strictly greater than that of any agent in w_1.[172] In order to do this, let us define a 'subpopulation bidirectional upgrade'. There is a subpopulation bidirectional upgrade from w_1 to w_2 if there is bidirectional upgrade from a subset of the agents of w_1 to a subset of the population of w_2.[173]

[172] In this chapter I continue to assume that agents have finite utility. This result does not hold if we believe that individual agents can have lives that contain infinite positive or negative utility (or both).

[173] By this definition, a bidirectional upgrade from the whole population of w_1 to the whole population of w_2 (rather than from or to a proper subset of the population) is also a subpopulation bidirectional upgrade.

Result 12: Cofinite Subset Incomparability

If there us a cofinite subset of the population of w_1 and there is a cofinite subset of the population of w_2 such that these cofinite subsets of w_1 and w_2 satisfy the conditions that the entire populations of w_1 and w_2 satisfy in Result 8 or Result 11 then world w_1 and w_2 are ethically incomparable ($w_1 \not\gtrless w_2$).

To show that if a cofinite subset of the population of $\langle w_1, w_2 \rangle$ satisfies the conditions of Result 8 then $w_1 \not\gtrless w_2$, let X be a cofinite subset of the population of $\langle w_1, w_2 \rangle$ such that there is a subpopulation bidirectional upgrade from the agents in X in w_1 to the agents in X in w_2 and there is an subpopulation bidirectional upgrade from the agents in X in w_2 to the agents in X w_1. Therefore X is a cofinite subset of the population of $\langle w_1, w_2 \rangle$ that satisfies the conditions that the entire population of a world pair satisfies in Result 8.

Let Y_1 be the set of finitely many agents that are not in X in w_1. Let Y_2 by the set of finitely many agents that are not in X in w_2. The populations of Y_1 and Y_2 may be identical, overlapping, or disjoint and so they may have different cardinalities. Let $n = |Y_1| - |Y_2|$ be the cardinality of the agents in Y_1 minus the cardinality of the agents in Y_2, such that n may be any finite (positive or negative) integer. If $n < 0$ then let Z_1 be a subset of the agents in X in w_1 such that $|Z_1| = n$, otherwise let Z_1 be empty. If $n > 0$ then let Z_2 be a subset of the agents in X in w_2 such that $|Z_2| = n$, otherwise let Z_2 be empty.

If there is a subpopulation bidirectional (or strict) upgrade from the agents in X in w_1 to the agents in X in w_2 and there is a subpopulation bidirectional upgrade from the agents in X in w_2 to the agents in X in w_1 then there exists a bijection g from the agents x in $X \setminus Z_1$ in w_1 to the agents x in $X \setminus Z_2$ in w_2 such that $\sum(u_{w_1}(x) - u_{w_2}(g(x)))$ diverges to ∞. We can show this by contradiction. Suppose there is a subpopulation bidirectional (or strict) upgrade from $pl(w_1)$ to $pl(w_2)$ and from $pl(w_2)$ to $pl(w_1)$ that does not satisfy this condition. It follows that finitely many agents in w_1 such that $u_{w_1}(x) < u_{w_2}(g(x))$ and the remaining agents in w_1 are such that $u_{w_1}(x) = u_{w_2}(g(x))$. But if this is the case then there cannot be a subpopulation bidirectional (or strict) upgrade from $pl(w_2)$ to $pl(w_1)$ because there are no agents $g(x)$ in w_2 such that $u_{w_2}(g(x)) < u_{w_1}(g(g(x)))$. This contradicts our assumption that there is a subpopulation bidirectional (or strict) upgrade from $pl(w_1)$ to $pl(w_2)$ and from $pl(w_2)$ to $pl(w_1)$ such that $\sum(u_{w_1}(x) - u_{w_2}(g(x)))$ does not diverge to ∞.

We now construct a bijection g from $pl(w_1)$ to $pl(w_2)$ as follows. First, let g map the agents in $Z_1 \cup Y_1$ in w_1 onto the agents in $Z_2 \cup Y_2$ in w_2. Let g map the agents in $X \setminus Z_1$ in w_1 onto the agents in $X \setminus Z_2$ in w_2 such that $\sum(u_{w_1}(x) - u_{w_2}(g(x)))$ diverges to ∞. Then any qualitative duplicate of $\langle w_1, w_2\rangle$ under g will be such that $\mathbf{g_n}(w_2) \succ \mathbf{g_{n+1}}(w_1)$ by Weak Catching-Up. Therefore if there exists a bijection from the agents in $X \setminus Z_1$ in w_1 to $X \setminus Z_2$ in w_2 that satisfies the conditions of Result 8 or Result 11 and there exists a bijection from the agents $X \setminus Z_2$ in w_2 to the agents in $X \setminus Z_1$ in w_1 that satisfies the conditions of Result 8 or Result 11, then it follows that $w_1 \not\gtrless w_2$ by Weak Catching-Up.

We can offer an example of a condition that is sufficient for a world pair to be incomparable by this result. Let the 'accumulation interval' of world w_1 be the interval between the lowest accumulation point of the utility levels of w_1 and the highest accumulation point of the utility levels of world w_1.[174] So if w_1 has only a single accumulation point n, then its accumulation interval is the degenerate interval $[n, n]$.[175] And if w_1 contains infinitely many agents at utility level n and infinitely many agents at utility level m and there are finitely many agents with utility greater than m and less than n, then the accumulation interval of w_1 is the non-degenerate interval $[n, m]$. If w_1 contains infinitely many agents approaching

[174] The Bolzano–Weierstrass theorem entails that if utilities are bounded then infinite worlds have at least one accumulation point. See [20, p. 78]. If utilities are not bounded then a world may not have an accumulation interval. For example if, for each natural number, there are finitely many agents in w_1 with utility equal to that natural number, then w_1 has no accumulation interval since w_1 has no accumulation points.

[175] If a world contains infinitely many agents and has finite total utility then its accumulation set must be the degenerate set [0,0] since the only possible accumulation point for such a world to have is 0.

utility n from above and infinitely many agents approaching utility m from below and w_1 has no other accumulation points, then the accumulation interval of w_1 is also $[n, m]$. I will focus on world pairs that contain infinitely many agents at each upper and lower accumulation point, however. When I refer to accumulation intervals in this chapter, I assume that there are infinitely many agents at each upper and lower accumulation point.

In this section I introduced the Weak Catching-Up rule and I have shown that if we accept the Weak Catching-Up rule, which is stronger than the Pareto principle,[176] then we can compare more world pairs as well as extend the incomparability results of the previous chapter. In the next section I will introduce a new set of principles that let us further extend these results.

4.2 Addition Principles and the Weak People Criterion

Until this point I have assumed that agents can have lives that contain different finite amounts of utility, but I have not assumed that there is a meaningful zero point of utility. However, many ethicists accept that there is an important point at which a life goes from being 'a life worth living' to 'a life not worth living'.[177] If a life is worth living then the life is good for the agent and contributes value to the world.[178] If a life is not worth living then it is bad for the agent and detracts value from the world. And if a life is neither worth living nor not worth living – it is neither bad for the agent nor good for the agent and it neither contributes nor detracts value from the world – I will say that it is a neutral life.[179] For the

[176]Weak Catching-Up entails Pareto but Pareto does not entail Weak Catching-Up.

[177]See, for example, Parfit [179] and Broome [44].

[178]We may believe that bringing people with lives worth living into existence is not only better for the world but better for the person in question. There is, however, great controversy about whether we can make comparisons between existence and non-existence – whether the $\succcurlyeq$ relation can hold between two relata if one of them exists and the other does not. I will not take a stance on this issue here. For discussion see, for example, Broome [44] and McMahan [162].

[179]The possibility of morally neutral lives or lives with zero welfare is fairly widely accepted. See, for example, Broome [44, Ch. 10].

remainder of this thesis I will set the zero level of lifetime utility at the point at which a life is a neutral. Lives that are worth living have lifetime utility strictly greater than zero and lives that are not worth living have lifetime utility that is strictly less than zero.

A fairly ubiquitous axiom among population ethicists is that if we add agents with lives worth living to a world without reducing the utility of any other agents that exist at that world, then we have not made the world worse and that if we add agents with lives not worth living to a world without increasing the utility of any other agents, then we have not made the more world better. We can also accept stronger versions of these axioms, which state that adding people with positive welfare to a world always makes the world better and, less controversially, that adding agents with negative welfare to a world makes the world worse.[180] A further axiom that we can accept is that adding neutral lives to a world neither makes a world better nor makes a world worse. This axiom is less widely adopted than the first two[181] but has been defended in the literature on infinite ethics. For example, Fishkind, Hamkins, and Montero [85] defend both standard addition principles and neutral addition in infinite worlds. For the sake of the present discussion I will accept both neutral addition and the stronger addition axioms. We can formulate these as the following three principles:

Benevolent Addition

If w_2 contains all of the agents of w_1 with at least at much utility as they have in w_1 plus some positive number of agents with lives worth living, then $w_2 \succ w_1$

[180]The claim that adding people with positive to a welfare necessarily makes a world better is controversial because many people will reject that it is better to bring into existence happy children than to fail to do so. This principle can also be used to generate the repugnant conclusion although, as Arrhenius [**?**] shows, we can generate the repugnant and very repugnant conclusions with a much weaker and more plausible addition axiom, namely the 'Dominance Addition Condition'.

[181]See Broome [44, p. 143-5] for a discussion of neutral addition in finite population ethics. Neutral addition may be controversial because we believe that there is some value to existing even if one has zero welfare. But if this is the case then we can simply incorporate this value into our utilities so that a neutral life is in fact a life with slightly negative welfare (to compensate for the value of existence).

Malevolent Addition

If w_2 contains all of the agents of w_1 with at most at much utility as they have in w_1 plus some positive number of agents with lives not worth living, then $w_1 \succ w_2$

Neutral Addition

If w_1 contains all of the agents of w_2 with at least at much utility as they have in w_2 plus some positive number of agents with neutral lives, then $w_1 \succcurlyeq w_2$

These addition principles allow us to compare various overlapping infinite world pairs that are not comparable by Weak Catching-Up. For example, if w_1 and w_2 both contain infinitely many agents with positive utility but the population of w_1 is a proper subset of the population of w_2 then $w_2 \succ w_1$ by Benevolent Addition. And if w_1 and w_2 both contain infinitely many agents with negative utility but the population of w_1 is a proper subset of the population of w_2 then $w_1 \succ w_2$ by Malevolent Addition. For example, in the cases below $w_1 \succ w_2$ and $w_4 \succ w_3$ even though neither world pair is comparable by Weak Catching-Up:

	A	B
#	$\aleph_0$	n
w_1	2	2
w_2	2	–

	A	B
#	$\aleph_0$	n
w_3	-4	-4
w_4	-4	–

Figure 78: Examples of Benevolent Addition and Malevolent Addition

The most important of these principles for our purposes, however, is the neutral addition principle. For example, consider the following two worlds, which are not incomparable by a four world or cyclic argument by Pareto or Weak Catching-Up:

	A	B	C
#	$\aleph_0$	$\aleph_0$	$\aleph_0$
w_1	3	0	–
w_2	–	–	3

Figure 79: Worlds accumulation intervals [0,3] and [3]

Now consider world w_3, which contains the same agents as world w_2 at the same utility levels plus infinitely many agents in set D at utility zero:

	C	D
#	$\aleph_0$	$\aleph_0$
w_3	3	0

Figure 80: Adding infinitely many utility 0 lives

By Neutral Addition, $w_2 \sim w_3$. So if $w_1 \succcurlyeq w_2$ then $w_1 \succcurlyeq w_3$ and if $w_2 \succcurlyeq w_1$ then $w_3 \succcurlyeq w_1$. But, by Result 2, world w_1 is incomparable with world w_3. Therefore world w_1 and world w_2 must also be ethically incomparable ($w_1 \not\gtrless w_2$).

I will show that the Weak Catching-Up rule and Neutral Addition are jointly equivalent to an objective variant of the principle for comparing infinite world pairs that Arntzenius [6, p. 55] calls the 'Weak People Criterion'. Following Artnzenius, we can suppose that the utility of a person at a world in which they do not exist is zero. This supposition is justified by Neutral Addition since Neutral Addition entails that if the population of $w_{1'}$ contains all of the agents of w_1 at the same utility levels plus any number of agents with neutral lives, then $w_1 \sim w_{1'}$. Therefore if $w_2 \succcurlyeq w_{1'}$ it follows from transitivity that $w_2 \succcurlyeq w_1$. So if there are some agents that exist in w_1 but not w_2 or there are some agents that exist in w_2 but no w_1, we can find a world $w_{1'}$ that contains all of the agents of w_1 plus any agents that are in w_2 but not w_1 at utility zero and we can find a world $w_{2'}$ that contains all of the agents of w_2 plus any agents that are in w_1 but not w_2 at utility zero. Therefore $w_{1'}$ and $w_{2'}$

contain identical populations. And if $w_{1'} \succcurlyeq w_{2'}$ then, by transitivity, $w_1 \succcurlyeq w_2$. Therefore we can suppose that the utility of a person in a world in which they do not exist is zero since any relations that hold between worlds in which these agents exist at utility zero also hold between worlds in which these agents do not exist at all by transitivity.

We can now formulate the Weak People Criterion as follows:

Weak People Criterion

($\succcurlyeq$) *If, summing over all possible agents x, $\sum_x (u_{w_1}(x) - u_{w_2}(x))$ is extendedly convergent and ≥ 0 then $w_1 \succcurlyeq w_2$*

($\succ$) *If, summing over all possible agents x, $\sum_x (u_{w_1}(x) - u_{w_2}(x))$ is extendedly convergent and > 0 then $w_1 \succ w_2$*

Arntzenius calls the principle 'weak' because it fails to rank worlds if the expected utility differences between their populations do not extendedly converge. The Weak People Criterion is equivalent to the Weak Catching-Up rule plus Neutral Addition. This is because the Weak People Criterion is just the Weak Catching-Up rule over world pairs in which agents that do not exist are treated as though they have utility zero. We have already shown that these are equivalent to worlds in which the agents do not exist by Neutral Addition.

Weak Catching-Up is silent about world pairs that have disjoint or overlapping populations, but Neutral Addition lets us extend Weak Catching-Up to the Weak People Criterion. This allows us to compare disjoint and overlapping world pairs $\langle w_1, w_2 \rangle$ by comparing an identical population pair $\langle w_3, w_4 \rangle$ in which all agents that exist only in w_1 in $\langle w_1, w_2 \rangle$ exist in both w_3 and w_4 but have a utility zero life in world w_3 and all agents that exist only in w_2 in $\langle w_1, w_2 \rangle$ exist in both w_3 and w_4 but have a utility zero life in world w_4.

I have shown that accepting addition principles and the Weak People Criterion lets us compare more worlds than we could using the Weak Catching-Up rule alone. To show that accepting these principles also extends the incomparability results, let us define a 'neutral expansion' of a world w_1 as any world that contains the same agents as world w_1 at the same utility levels plus any number of agents with neutral lives. Let us also define a 'neutral contraction' of a world w_1 as any world that contains the same agents as world w_1 at the same utility levels minus any number of agents with neutral lives. (This means that each world is a neutral expansion and contraction of itself, since w_1 plus zero agents with neutral lives is just w_1). We can expand the incomparability results to world pairs that are not directly incomparable by a four world or cyclic argument if there are neutral expansions or contractions of the worlds in the pair that are incomparable.

Result 14: Neutral Expansion/Contraction Incomparability

If there is a neutral expansion [contraction] of w_1 that is is incomparable with a neutral expansion [contraction] of w_2, then w_1 and w_2 are ethically incomparable ($w_1 \not\gtreqless w_2$)

Let $w_{1'}$ be a neutral expansion of w_1 and let $w_{2'}$ be a neutral expansion of w_2 such that $w_{1'} \not\gtreqless w_{2'}$. Since $w_{1'}$ is a neutral expansion of w_1, $w_1 \sim w_{1'}$ by Neutral Addition. Therefore $w_1 \not\gtreqless w_{2'}$. Since $w_{2'}$ is a neutral expansion of w_2, $w_2 \sim w_{2'}$ by Neutral Addition. Since $w_1 \not\gtreqless w_{2'}$ it follows that $w_1 \not\gtreqless w_2$. And if $w_1 \not\gtreqless w_2$ then $w_{1'} \not\gtreqless w_{2'}$ by the same reasoning.

I won't attempt to formulate sufficient conditions for incomparability by neutral expansion here but we can explore some of the consequences of this result. For example, we can show that a weaker condition than the condition of Result 13 is sufficient for incomparability.

Result 15: Disjoint Incomparability by Neutral Expansion

If w_1 and w_2 have disjoint populations and the lower endpoint of the accumulation interval of w_1 is non-negative and the lower endpoint of the accumulation interval of w_2 is non-negative and the the upper endpoint of both intervals are identical and strictly positive then $w_1 \not\gtreqless w_2$

Let the upper endpoint of the accumulation intervals of w_1 and w_2 be n. Since the lower endpoint of the accumulation intervals of w_1 and w_2 are non-negative, there is a neutral expansion of w_1 with the accumulation interval $[0, n]$ and there is a neutral expansion of w_2 with the accumulation interval $[0, n]$. Since n is strictly positive, w_1 and w_2 are incomparable by Result 12.

By the same reasoning we can show that if w_1 and w_2 have disjoint populations and the upper endpoint of the accumulation interval of w_1 is non-positive and the upper endpoint of the accumulation interval of w_2 is non-positive and the lower endpoint of both intervals are identical and strictly negative then world w_1 and world w_2 are ethically incomparable.

We can further extend the incomparability results of this chapter if we accept one further claim – that what matters for the ranking of worlds is the pattern of difference in the utility that agents have in one world rather than the other and not the absolute level of utility they have at each world. We can capture this in the following principle:[182]

Transformation Indifference

If (i) w_1 and w_2 have identical populations, and (ii) $\langle w_{1'}, w_{2'} \rangle$ and $\langle w_1, w_2 \rangle$ have identical populations, and (iii) there exists a positive affine function f such that for all agents x, if x has utility $u(x)$ in $\langle w_1, w_2 \rangle$ then x has utility $f(u(x))$ in $\langle w_{1'}, w_{1'} \rangle$ and if x has utility $u(x)$ in $\langle w_{1'}, w_{1'} \rangle$ then x has utility $f^{-1}(u(x))$ in $\langle w_1, w_2 \rangle$, then $w_1 \succcurlyeq w_2$ if and only if $w_{1'} \succcurlyeq w_{2'}$.

Transformation indifference says that if $w_1 \succcurlyeq w_2$ and f is a positive affine function – one that preserves addition and scalar multiplication but not the zero point – then if $w_1 \succcurlyeq w_2$ and we transform the utilities of the shared agents of $\langle w_1, w_2 \rangle$ by f then $\succcurlyeq$ relation is preserved under this transformation. For example, if w_1 contains infinitely many agents at utility level 2 and w_2 contains infinitely many agents at utility level 3 and $f = 2y + 1$ then the agents of

[182]This should not be confused with the claim that utility functions are unique up to a positive affine transformation. The utility function over world pairs may be unique up to a positive affine transformation, but Transformation Indifference concerns alterations to the utilities of individual agents in these world pairs.

$w_{1'}$ all have utility level 5 and the agents of w_2 all have utility level 7. By Transformation Indifference, $w_{1'} \not\succcurlyeq w_{2'}$.[183] Since we are using positive affine transformations, if, for any agent x in $\langle w_1, w_2 \rangle$, the difference between $u(x)$ in w_1 and $u(x)$ in w_2 is n then the difference between $u(x)$ in $w_{1'}$ and $u(x)$ in $w_{2'}$ is also n. Transformation Indifference is therefore equivalent to the claim that it is the profile of the differences in agent utilities and not the profile of the absolute values of agent utilities that matters for the ethical ranking of worlds.

Transformation Indifference is restricted to identical population world pairs but if w_1 and w_2 contain disjoint or overlapping populations then we can use Neutral Addition to consider positive affine transformations of the utilities of a pair of worlds $\langle w_3, w_4 \rangle$ such that (i) w_3 contains all of the agents that exist in w_1 with the same utility levels they have in w_1 but the agents that exist in w_2 but not w_1 exist with neutral lives in w_3, and (ii) w_4 contains all of the agents that exist in w_2 with the same utility levels they have in w_2 but the agents that exist in w_1 but not w_2 exist with neutral lives in w_4. Transformation Indifference is also more plausible if we try to preserve as many qualitative properties and relations of the original pair as possible while transforming the utilities of its population.

Transformation Indifference is not an uncontroversial principle and not all population ethicists will accept it. The principle is, however, consistent with many theories within population ethics because it is a principle restricted to the ranking of worlds: it does not say anything about how much better one world is than another. Consider prioritarianism, which says that the goodness of a world is a function of the wellbeing of the agents of that world but with greater weight being given to agents with lower utility lives. If the function that the prioritarian uses to weight increases in wellbeing is a monotonically increasing function, which seems plausible, then the prioritarian ranking will satisfy Transformation Indiffer-

[183] Here we can see why Translation Indifference is restricted to positive affine transformations. If it extended to non-positive affine transformations then the scale factor could be non-positive. If the scale factor were zero, the utilities of the agents of $w_{1'}$ and $w_{2'}$ would be identical. If the scale factor were negative, the utilities of the agents of $w_{1'}$ would be strictly greater than the utilities of the agents of $w_{2'}$.

ence.[184] However, any population ethics that attaches significance to absolute utility levels or weighs the utility of agents using a function that is not monotonically increasing will violate Transformation Indifference. For example, if a theory that says that there is some utility level such that increasing the lifetime utility of agents above that level do not make the world better, then that theory will be inconsistent with Transformation Indifference.[185]

Vallentyne and Lauwers [148] endorse a principle called 'Zero Independence' that is similar in spirit to Transformation Indifference in that it entails that it is formulated to capture the idea that it is differences in utility and not absolute utility levels that matter. Let w_1 and w_2 have identical populations. Zero Independence says that $w_1 \succcurlyeq w_2$ if and only if w_3 – a world with the same population of $\langle w_1, w_2 \rangle$ but where each agent has utility equal to their utility in w_1 minus their utility in w_2 – is at least as good as w_4 – a world a world with the same population of $\langle w_1, w_2 \rangle$ but where each agent has utility zero. For example, if w_1 contains agents all with utility 3 lives and w_2 contains the same agents all with utility 2 lives then w_3 contains agents all with utility 1 lives and w_4 contains the same agents all with utility 0 lives. By Zero Independence, $w_1 \succcurlyeq w_2$ if and only if $w_3 \succcurlyeq w_4$. In support of Zero Independence, Vallentyne and Lauwers point out that in finite worlds the total value of a world is greater if and only if the sum of the differences in value is greater than zero. This means that finitely additive theories will entail the finite equivalent of both Zero Indifference and Transformation Indifference in these worlds.

A more controversial principle than Transformation Indifference is what I will call Subpopula-

[184] Otsuka states that 'The priority-weighted utility of any given outcome is the sum of everyone's utility in that outcome, after each person's utility has been assigned the relevant positive but decreasing marginal moral importance in prioritarian fashion.' [178, p. 3]. The claim that each person's utility is assigned positive but decreasing marginal importance supports the claim that the prioritarian function is monotonically increasing.

[185] Transformation Indifference is consistent with views that don't treat all increases in wellbeing as equally valuable. If we favor egalitarian principles, for example, then we may believe that a world pair in which there is greater variance in lifetime utility is worse than a world pair in which there is lower lifetime utility variance, all else being equal. But we might nonetheless accept that the pattern of differences in utility between worlds determines whether one world is at least as good as another, even if it doesn't determine whether one pair of worlds is better than another pair of worlds.

tion Transformation Indifference. According to Subpopulation Transformation Indifference, if we transform the utilities of a subpopulation of a world pair by some positive affine function then if $\succcurlyeq$ holds between the original world pair, it also holds between the new pair:

Subpopulation Transformation Indifference

If (i) w_1 and w_2 have identical populations, and (ii) $\langle w_{1'}, w_{2'} \rangle$ and $\langle w_1, w_2 \rangle$ have identical populations, and (iii) for all agents x, there exists a positive affine function f such that if x has utility $u(x)$ in $\langle w_1, w_2 \rangle$ then x has utility $f(u(x))$ in $\langle w_{1'}, w_{1'} \rangle$ and if x has utility $u(x)$ in $\langle w_{1'}, w_{1'} \rangle$ then x has utility $f^{-1}(u(x))$ in $\langle w_1, w_2 \rangle$, then $w_1 \succcurlyeq w_2$ if and only if $w_{1'} \succcurlyeq w_{2'}$.

Subpopulation Transformation Indifference says that if $w_1 \succcurlyeq w_2$ and we transform some agent utilities by a positive affine function f and we transform other agent utilities by a different positive affine function g, resulting in $\langle w_{1'}, w_{1'} \rangle$, then $w_{1'} \succcurlyeq w_{2'}$. For example, suppose w_1 and w_2 contain infinitely many agents that are at utility 3 in w_1 and utility 1 in w_2. Let X denote this set of agents. Suppose w_1 and w_2 also contain infinitely many agents that are at utility 2 in w_1 and utility 4 in w_2. Let Y denote this set of agents. Suppose w_1 and w_2 contain no other agents. We can transform the utilities of the agents in X by $f = y - 2$ so that they have utility 1 in $w_{1'}$ and utility -1 in $w_{2'}$. We can transform the utilities of the agents in Y by $g = y - 3$ so that they have utility -1 in $w_{1'}$ and utility 1 in $w_{2'}$. By Subpopulation Transformation Indifference, $w_1 \succcurlyeq w_2$ if and only if $w_{1'} \succcurlyeq w_{2'}$. Such a transformation is not possible if we must use the same function, however.

Subpopulation Transformation Indifference is much more controversial than Transformation Indifference. It entails that ethical rankings depend on the differences in utility that agents experience in w_1 and w_2 but are not sensitive to whether the differences in utility are experienced by high or low utility agents. It is therefore not likely to be endorsed by egalitarians or prioritarians since, according to these theories, whether and to what extent improving the lives of agents improves the world is in part a function of the absolute level of utility that those agents would have had otherwise. For example, suppose that w_1 and w_2 both are both finite worlds containing two agents – x and y – and that x has utility 0 in w_1 and utility 1 in w_2 while y has utility 100 in w_1 and utility 99 in w_2. Suppose that $\langle w_{1'}, w_{2'} \rangle$ have the same population as $\langle w_1, w_2 \rangle$ but x has utility 50 in w_1 and utility 51 in w_2 and y has utility 51 in w_1 and utility 50 in w_2. By Subpopulation Transformation Indifference, $w_1 \succcurlyeq w_2$ if and only if $w_{1'} \succcurlyeq w_{2'}$. But egalitarians may argue that $w_1 \not\succcurlyeq w_2$ and $w_{1'} \succcurlyeq w_{2'}$ because w_2 has a more equal distribution of utility than w_1 and $w_{1'}$ and $w_{2'}$ have equal utility distributions.

I acknowledge that Subpopulation Transformation Indifference will therefore not be attractive to those who believe that the distribution of utilities between agents matters. Let us assume, however, that we are attempting to find extensions of aggregative ethical principles to infinite worlds. For those that are indifferent to the distribution of utilities between agents in finite worlds, Subpopulation Transformation Indifference will seem like a much more reasonable principle. If we accept Subpopulation Transformation Indifference then we are able to show that many disjoint worlds with distinct accumulation intervals are ethically incomparable. For example, consider the following pair of worlds:

	A	B	C	D
#	$\aleph_0$	$\aleph_0$	$\aleph_0$	$\aleph_0$
w_1	-1	0	2	–
w_2	-3	-2	–	4

Figure 81: Worlds with accumulation intervals [-1, 2] and [-3, 4]

Worlds w_1 and w_2 are not comparable by the Weak People Criterion. But we have not shown that world pairs like this – pairs that have distinct accumulation intervals[186] – are ethically incomparable. If we accept Neutral Addition then we can infer what relation holds between w_1 and w_2 from what relation holds between the neutral expansions $w_{1'}$ and $w_{2'}$:

[186] The accumulation interval of w_1 is $[-1, 2]$ and the accumulation interval of w_2 is $[-3, 4]$.

	A	B	C	D
#	$\aleph_0$	$\aleph_0$	$\aleph_0$	$\aleph_0$
$w_{1'}$	-1	0	2	0
$w_{2'}$	-3	-2	0	4

Figure 82: The neutral expansion of this world pair

If we accept Subpopulation Transformation Indifference, we can infer what relation holds between $w_{1'}$ and $w_{2'}$ from what relation holds between utility transformations $w_{1''}$ and $w_{2''}$:

	A	B	C	D
#	$\aleph_0$	$\aleph_0$	$\aleph_0$	$\aleph_0$
$w_{1''}$	4	4	2	0
$w_{2''}$	2	2	0	4

Figure 83: A subpopulation transformation of the neutral expansion pair

By neutral contraction, we can infer what relation holds between $w_{1'}$ and $w_{2'}$ from what relation holds between $w_{1'''}$ and $w_{2'''}$, which are neutral contractions of $w_{1''}$ and $w_{2''}$:[187]

	A	B	C	D
#	$\aleph_0$	$\aleph_0$	$\aleph_0$	$\aleph_0$
$w_{1'''}$	4	4	2	-
$w_{2'''}$	2	2	-	4

Figure 84: A neutral contraction of the subpopulation transformation

In Chapter 3 we showed that world pairs in which agents have the utilities of $w_{1'''}$ and $w_{2'''}$ are incomparable by a cyclic argument. Therefore w_1 and w_2 are ethically incomparable.

One final set of principles that we may wish to consider are 'transfer principles'. In finite worlds, the total utility of a world remains the same regardless of how we distribute utility across agents, since transferring utility from one agent to another can never change the total.

[187] Alternatively, we could simply show that $w_{1''}$ and $w_{2''}$ are incomparable directly.

In infinite worlds, transfers are more complicated. We might think that if one world can be obtained from another by transferring utility between the agents of the world, then those worlds must be equally good. But in infinite worlds we cannot endorse such a principle without violating the principles outlined above. To see why, suppose that we give each agents of world w_1 a name corresponding with a natural number – a_1, a_2, a_3, and so on – and that every agent in w_1 has a utility 0 life, So if we arrange the agents by their names then their utilities are $(0, 0, 0, 0, 0, 0, 0, 0, ...)$. Now suppose that the odd-numbered agents all donate 1 util to the agent whose name is 1 greater than their own, so the agent utilities are $(-1, 1, -1, 1, -1, 1, -1, ...)$. But if we accept Subpopulation Transformation Indifference then these two worlds are incomparable, since they are equivalent to the following:

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	...
$w_{1'}$	1	0	1	0	1	0	1	0	...
$w_{2'}$	0	1	0	1	0	1	0	1	...

Figure 85: A counterexample to unrestricted transfer principles

Since these worlds are incomparable by a four world argument, we cannot endorse unrestricted transfer principles in infinite worlds without violating one of these principles. Vallentyne and Lauwers argue that when comparing infinite worlds we should, however, accept a weaker transfer principle. They define a 'restricted transfer' as '(1) a transfer of a positive amount of value from a location with positive value to a location with negative value such that (2) after the transfer, the donor location still has non-negative value and the recipient location still has non-positive value.' [148, p. 323]. The example above did not involve restricted transfers since the transfers fail to satisfy either of these conditions.

Restricted transfers can only be used to bring both the donor agent and the recipient agent closer to zero. Vallentyne and Lauwers believe that the following principle is plausible:

Restricted Transfers

If w_1 and w_2 contain identical populations and $w_1 \succ w_2$ and we can obtain w_3 by performing finitely or infinitely many restricted transfers between the agents of w_1, then $w_3 \succ w_2$.

It is important that we only let each agent in a world transfer utility to at most one other agent in the world or else Restricted Transfers will conflict with the irreflexivity of $\succ$.[188]

Vallentyne and Lauwers argue that Restricted Transfers and Zero Independence plus certain further axioms entail that Weak Catching-Up is necessary and sufficient for ranking infinite world pairs with identical populations. We can use Neutral Addition to show by a similar method that if w_1 and w_2 are not comparable by the Weak People Criterion then w_1 and w_2 are incomparable if we accept the following additional principle:

No Transfer Dominance

If w_1 and w_2 contain identical populations and we can obtain w_2 by performing finitely or infinitely many restricted transfers between the agents of w_1, then $w_1 \nsucc w_2$.

We can show that if we accept Neutral Addition, Transformation Indifference and these transfer principles then w_1 and w_2 are comparable if and only if they are comparable by the Weak People Criterion.

Result 16: The Weak People Criterion is Necessary and Sufficient for $\succcurlyeq$

If we accept Neutral Addition, Transformation Indifference, Restricted Transfers, No Transfer Dominance, and transitivity then if w_1 and w_2 are not comparable by the Weak People Criterion then w_1 and w_2 are incomparable.

To show this, suppose for contradiction that w_1 and w_2 are not comparable by the Weak

[188]Suppose we don't adopt this restriction. Consider ordered streams of agent utilities: a_1's utility always comes first, a_2's always second, and so on. we can transform the utilities of the ordered stream (2,-1,2, -1,2,-1,...) into the ordered stream (0,0,0,0,0,0,...). Considered the stream (1,-1,1,-1,1,-1,...) that is worse than the first by Pareto. We can also transform this into the ordered stream (0,0,0,0,0,0,...).

People Criterion and that w_1 and w_2 are comparable: either $w_1 \succcurlyeq w_2$ or $w_2 \succcurlyeq w_1$.

Suppose that $w_1 \succcurlyeq w_2$. Let w_3 be a neutral expansion of world w_1 such that for all agents x that exist in w_2 but not w_1, x exists in w_3 with a utility zero life. Let w_4 be a neutral expansion of world w_1 such that for all agents x that exist in w_1 but not w_2, x exists in w_4 with a utility zero life. Therefore w_3 and w_4 have identical populations. Since $w_1 \succcurlyeq w_2$ by hypothesis and $\langle w_3, w_4 \rangle$ is a neutral expansion of $\langle w_1, w_2 \rangle$, $w_3 \succcurlyeq w_4$ by Neutral Addition.

Since w_1 and w_2 are not comparable by the Weak People Criterion it must be the case that there are infinitely many agents such that $u_{w_3}(x) < u_{w_4}(x)$ and infinitely many agents such that such that $u_{w_4}(x) < u_{w_3}(x)$ and there may be any number of agents such that $u_{w_3}(x) = u_{w_4}(x)$. Let y be an agent that has strictly greater utility in w_4 than in w_3 and let X be the set of all other agents $\langle w_3, w_4 \rangle$. By performing infinitely many restricted transfers we can, without altering the utility of agent y, transfer utility between the positive utility agents in X in w_3 to the negative utility agents in w_4 and between the positive utility agents in X in w_4 to the negative utility agents in w_3 until all for all $x \in X$, $u_{w_3}(x) - u_{w_4}(x)$=0.

Let w_5 be a world we obtain by performing these restricted transfers between the agents of w_3 and let world w_6 be a world we obtain by performing these restricted transfers between the agents of w_4. World w_5 and world w_6 have identical populations. For all $x \in X$, $u_{w_5}(x) = u_{w_6}(x) = 0$ and for the agent y, $u_{w_5}(y) < u_{w_6}(y)$. Therefore $w_6 \succ w_5$ by Pareto.

Since we obtained w_6 be performing infinitely many restricted transfers between the agents of w_4, $w_4 \succ w_5$ by Restricted Transfers. Since $w_3 \succcurlyeq w_4$ and $w_4 \succ w_5$ it follows that $w_3 \succ w_5$ by transitivity. But we derived w_5 by performing infinitely many restricted transfers between the agents of w_3. Therefore $w_3 \not\succ w_5$ by No Transfer Dominance. We have derived a contradiction from our supposition that $w_1 \succcurlyeq w_2$. Therefore $w_1 \not\succcurlyeq w_2$.

We can use the same method to show that $w_2 \not\succcurlyeq w_1$ by letting y be an agent that has strictly greater utility in w_3 than in w_4. Therefore $w_1 \not\succcurlyeq w_2$ and $w_2 \not\succcurlyeq w_1$. This contradicts our assumption that w_1 and w_2 are comparable.

The key problem with the restricted transfers principles is that they lack independent motivation. Many transfer principles in infinite ethics produce results that are in tension with Pareto. Moreover, restricted transfers treat the zero point as ethically relevant for transfers, despite the fact that principles like Transformation Indifference and Zero Independence are in tension with the claim that there is a privileged zero point of utility. For these reasons I am inclined to think that we should be suspicious of any result that is based on restricted transfer principles, though I include this one here for completeness.

I have now shown that if we accept the Weak Catching-Up and Neutral Addition, which are jointly equivalent to the Weak People Criterion, then we can expand the set of infinite worlds that are comparable beyond those that are merely Pareto comparable and, in doing so, we expand the set of world pairs that are incomparable by four world and cyclic arguments. If we accept Subpopulation Transformation Indifference then we can further expand the set of world pairs that are incomparable by four world and cyclic arguments. Finally, if we accept the less well-motivated Restricted Transfers and No Transfer Dominance principles, we can show that infinite worlds are comparable only if they are comparable by the Weak People Criterion. If we accept the Weak People Criterion and these axioms, we must conclude that infinite worlds are comparable if and only if they are comparable by the Weak People Criterion. This will entail a great deal of incomparability between infinite worlds.

I have argued that Restricted Transfers and No Transfer Dominance are fairly objectionable principles, and so many will reject the claim that infinite worlds are comparable if and only if they are comparable by the Weak People Criterion. Weak Catching-Up, Neutral Addition, and Subpopulation Transformation Invariance are also not entirely uncontroversial principles, however, and how much we can expand the results of the previous chapter depends on which of these principles we accept (if any). The goal of this section was not to definitively show that all infinite world pairs can be shown to be either comparable or incomparable, however. The goal was merely to show that if we accept certain principles that at least some will find plausible, then more infinite worlds can be shown to be comparable than are comparable by Pareto and more infinite worlds can be shown to be incomparable by four world and cyclic arguments. I have shown that this is true of each of the principles outlined in this section.

We might be tempted to offer alternative principles for comparing infinite worlds than those outlined in this section. In particular, we might be tempted to use facts about the accumulation intervals of infinite worlds to try to rank them. In the next section I argue that we ought to reject these stronger principles.

4.3 Against Accumulation Principles

Before moving on I want to briefly consider a class of principles that we might be inclined use to rank infinite world pairs that are not comparable by the Weak People Criterion (assuming that we reject at least one axiom appealed to in Result 16). When w_1 and w_2 are not comparable by the Weak People Criterion, we might think that we can ethically rank these worlds using their accumulation intervals. This has some intuitive appeal to it. After all, if world w_1 contains infinitely many agents at utility 100 and infinitely many agents at utility 4, while world w_2 contains infinitely many agents at utility 5, it is tempting to conclude that w_1 is better than w_2. Consider the following accumulation principles for disjoint population world pairs:

Accumulation $\ngtrless$ *If w_1 and w_2 have disjoint populations and identical, non-degenerate accumulation intervals then w_1 and w_2 are ethically incomparable ($w_1 \ngtrless w_2$)*

Accumulation $\sim$ *If* w_1 *and* w_2 *have disjoint populations and identical, degenerate accumulation intervals then* w_1 *and* w_2 *are equally good* ($w_1 \sim w_2$)

Accumulation $\succ$ *If* w_1 *and* w_2 *have disjoint populations and non-identical accumulation intervals and the midpoint of the accumulation interval of* w_1 *is strictly greater than the midpoint of the accumulation interval of* w_2 *then* w_1 *is strictly better than* w_2 ($w_1 \succ w_2$)

Accumulation $\not\gtreqless$ follows from Result 12, since if w_1 and w_2 have disjoint populations and identical, non-degenerate accumulation intervals then there is a strict upgrade from a subset of w_1 to w_2 and there is a strict upgrade from a subset of w_2 to w_1. This subset satisfies the conditions of Result 1 and therefore of Result 8. If we accept the Accumulation $\not\gtreqless$ principle then we won't run into conflicts with Pareto if we accept transitivity, the Permutation Principle, and that $\succcurlyeq$ is a qualitative relation.

The Accumulation $\sim$ principle says that if two worlds with disjoint populations have degenerate accumulation intervals – both worlds contain infinitely many agents at some utility level n and there is no other utility level such that infinitely many agents have that utility level in w_1 or w_2 – then w_1 and w_2 are equally good. Accumulation $\sim$ is consistent with Lauwers' 'Infinite Sensitivity' principle from Chapter 1 but is inconsistent with Pareto using four world or cyclic arguments in cases where no agents are worse off in w_1 than in w_2 and finitely many agents are strictly better off in w_1 than in w_2. Suppose w_1 contains a thousand agents at utility 100 and infinitely many agents at utility 0, while world w_2 contains infinitely many agents at utility 0. The accumulation interval of w_1 and w_2 are the same: the degenerate interval [0,0] and yet w_1 is clearly better than w_2. I believe this is sufficient reason to reject this principle and so I will focus on the Accumulation $\succ$ principle below.

The Accumulation $\succ$ principle says that one infinite world is strictly better than another if the midpoint of its accumulation interval is higher. For example, if w_1 comprises infinitely many agents at utility 4 and infinitely many agents at utility 1 and w_2 comprises infinitely many agents at utility 2, then w_1 is strictly better than w_2 by the Accumulation $\succ$ principle since the midpoint of the accumulation interval of w_1 is 3 and the midpoint of the accumulation interval of w_2 is 2. We can think of this as a kind of 'average utility' principle in infinite worlds. This may be appealing to those who favor average views in finite ethics, since it is difficult to transfer these principles to infinite worlds.[189]

The Accumulation $\succ$ principle also entails plausible weaker principles like 'if the lower endpoint of the accumulation interval of w_1 is strictly greater than the upper endpoint of the accumulation interval of w_2 then $w_1 \succ w_2$'. The Weak People Criterion entails that if the accumulation interval of w_1 is non-negative and the accumulation interval of w_2 is non-positive then $w_1 \sim w_2$ if the accumulation interval of both worlds is $[0, 0]$ and $w_1 \succ w_2$ for any other values of the accumulation intervals of w_1 and w_2. But, unlike the Accumulation $\succ$ principle, it does not entail that that if the lower endpoint of the accumulation interval of w_1 is strictly greater than the upper endpoint of the accumulation interval of w_2 then $w_1 \succ w_2$.

I have restricted Accumulation principles to disjoint population world pairs. This is because we cannot straightforwardly use accumulation intervals to rank identical or overlapping world pairs as we do above. To see why, suppose that w_1 and w_2 have identical populations and the accumulation interval of w_1 and w_2 is $[2, 4]$. This is consistent with w_1 and w_2 being incomparable (if infinitely many agents that have utility 2 in w_1 have utility 4 in w_2 and infinitely many agents that have utility 2 in w_2 have utility 4 in w_1) or with w_1 being strictly better than w_2 by Pareto (if all of the agents that have utility 4 in w_2 have utility 4 in w_1 and some of the agents that have utility 2 in w_2 have utility 4 in w_1) or with w_1 and w_2 being equally good by Pareto (if all of the agents that have utility 2 or 4 in w_1 have the same

[189] Average utility is not well-defined in infinite worlds if agents lack a privileged order. If we reject the midpoint view then it is likely that the only alternative will be to identify average utility of a world with its entire accumulation interval, since its average utility as defined by something like a natural density function can, under different orderings, be made to converge on any value in the accumulation interval.

utility in w_2). It is possible to extend Accumulation principles to identical or overlapping worlds, but I will focus on the disjoint population principles here.[190]

I believe that we ought to reject the Accumulation $\succ$ principle because it is inconsistent with Benevolent Addition, Malevolent addition and the Weak People Criterion. We can see this in the following two cases:

	A	B	C	D
#	∞	∞	∞	∞
w_1	3	-	-	-
w_2	-	3	2	-
w_3	3	-	-	1

	A	B	C	
#	∞	∞	∞	
w_4	-3	-	-	-
w_5	-	-3	-2	-
w_6	-3	-	-	-1

Figure 86: Two problematic cases for the Accumulation $\succ$ principle

By the Accumulation $\succ$ principle, $w_1 \succ w_2 \succ w_3$. But w_3 is strictly better than w_1 by both Benevolent Addition and by the Weak People Criterion. Therefore the Accumulation $\succ$ principle violates both of these principles. And by the Accumulation $\succ$ principle, $w_6 \succ w_5 \succ w_4$. But w_6 is strictly worse than w_4 by both Malevolent Addition and by the Weak People Criterion. Therefore the Accumulation $\succ$ principle violates both of these principles.

The following weaker Accumulation $\succ$ principle does not violate the addition principles: 'if the lower endpoint of the accumulation interval of w_1 is strictly greater than the upper endpoint of the accumulation interval of w_2 then $w_1 \succ w_2$'. However, this principle is

[190]To extend Accumulation principles to the overlapping sets of agents in worlds, we need to appeal not to the accumulation interval but to what I will call the 'difference interval' of the world pair. A difference interval is an accumulation interval in which accumulation points have been replaced with differences in utility across the two worlds (meaning that difference interval of each world with itself is $[0, 0]$). So if infinitely many agents have utility 2 in w_1 and infinitely many agents have utility -2 in w_1 and the same agents all have utility 1 in w_2 then the difference interval of (w_1, w_2) is $[-3, 1]$ and the difference interval of (w_2, w_1) is $[-1, 3]$. Accumulation theorists can then say, that if the difference interval of w_1 and w_2 is identical and non-degenerate then w_1 and w_2 are incomparable, and if the the difference intervals are non-identical and the midpoint of the difference interval of (w_1, w_2) is strictly greater than the midpoint of the difference interval of (w_2, w_1) then $w_1 \succ w_2$, and if the midpoint is identical then $w_1 \sim w_2$. We can combine this with the disjoint Accumulation principles to rank overlapping worlds, but I will not do so here.

produces counterintuitive results of its own. Consider the following case:

	A	B	C
#	∞	∞	∞
w_1	3	3	-
w_2	-	-	$3+\epsilon$
w_3	$(3+\epsilon)+\epsilon$	-	-

Figure 87: A problematic case for the weakened Accumulation $\succ$ principle

According to the weak Accumulation $\succ$ principle, $w_3 \succ w_2 \succ w_1$. This means that it can be good to prevent infinitely many agents in set B from existing at utility level 3 in order to increase the utility of infinitely many agents in set A by an arbitrarily small amount of utility. It is not clear why we should consider this an improvement.

For these reasons I believe that we should not attempt to rank infinite world pairs using Accumulation principles. Despite their prima facie plausibility, they often conflict with Benevolent and Malevolent Addition or produce counterintuitive rankings.

Summary

At the beginning of this chapter I showed that a modest extension of Pareto – the Weak Catching-Up principle – allows us to compare more worlds than Pareto and to extend the incomparability results of the previous chapter. I then demonstrated that we can increase the number of demonstrably comparable worlds and demonstrably incomparable worlds if we accept further principles like Neutral Addition and Subpopulation Transformation Indifference. I also showed that if we accept these principles plus the Restricted Transfer and No Transfer Dominance principles then we can reduce the number of worlds that are neither demonstrably comparable nor demonstrably incomparable to zero. Finally, I argued that Accumulation principles for comparing infinite worlds are not a plausible extension of or alternative to the Pareto principle. I do not take a strong stand on which of the principles I have discussed in this chapter we should accept, but I believe that I have demonstrated the potential to extend the incomparability results of the previous chapter.

Chapter 5

The Implications for Ethics

In this chapter I consider the ethical consequences of the incomparability results of Chapters 2 and 3. In section 5.1 I characterize the incomparability results as an impossibility result: we cannot accept the Transitivity of $\succcurlyeq$, the Permutation Principle, the Qualitativeness of $\succcurlyeq$ and Pareto without concluding that most infinite worlds are incomparable. In section 5.2 I argue that the first three axioms are highly plausible and show that rejecting any of them has unacceptable consequences. In section 5.3 I defend the Pareto principle. If we accept the Transitivity of $\succcurlyeq$, the Permutation Principle, the Qualitativeness of $\succcurlyeq$ and Pareto, however, then we must conclude that there is pervasive incomparability between infinite worlds.

In section 5.4 I formulate new problems for objective and subjective permissibility that arise if there is pervasive incomparability between infinite worlds. I show that if we accept incomparability then we are forced to reject at least one analogue of the four axioms above for objective and subjective permissibility. This problem applies not only to consequentialist theories of permissibility but to all theories of objective and subjective permissibility, including deontological theories and virtue ethics, which are not premised on there being a close connection between the value of worlds and the permissibility of actions. This shows that the pervasive incomparability problem creates serious problems for ethics and not merely for consequentialists. I conclude that the incomparability results of Chapters 2 and 3 leave ethics radically altered, regardless of which axioms we ultimately reject.

5.1 The Incomparability Result as an Impossibility Result

In the previous chapter I considered several possible extensions of the incomparability results of Chapters 2 and 3. In this chapter I am going to return to the original incomparability results that did not assume any ethical principles stronger than Pareto. As I argued in Chapters 2 and 3, if we accept Pareto, the Qualitativeness of $\succcurlyeq$, the Permutation Principle, and the Transitivity of $\succcurlyeq$, then we can show that many infinite worlds are incomparable: w_1 is not at least as good as w_2 and w_2 is not at least as good as w_1. If we accept these four axioms we must therefore conclude that the completeness axiom is false: not all worlds are ethically comparable. This is not merely incomparability between a small subset of infinite world pairs, however. If we accept these axioms then incomparability between infinite world pairs is ubiquitous. Whenever there is a bidirectional upgrade from w_1 to w_2 and vice versa, w_1 and w_2 are incomparable (Result 8). And for any randomly selected pair of infinite worlds $\langle w_1, w_2 \rangle$, it is likely that we will be able to find a bidirectional upgrade from w_1 to w_2 and from w_2 to w_1.[191] Therefore incomparability between infinite worlds will be ubiquitous.

If one infinite world least as good as the other by Pareto or some extension of Pareto such as the Weak People Criterion, then those worlds be shown to be incomparable by a four world or cyclic argument. This makes up an extremely small fraction of infinite world pairs. however. To give an intuitive story about why this is uncommon, suppose that two agents were to play a game in which they each flip a fair coin independently of one another. An agent wins a round every time their coin comes up heads and the other person's comes up

[191]There is not a bidirectional upgrade from w_1 to w_2 and from w_2 to w_1 only if there is some agent in w_1 with utility strictly greater than or strictly less than that of any agent in w_2 (and, if we accept the Weak People Criterion, only if there are infinitely many such agents). Unless there is something preventing any of the agents in w_2 from having utility levels above or below some threshold, it would be extremely surprising to find that this condition is satisfied. Suppose we rolled two infinite-sided dice (representing possible utility levels) infinitely many times (representing the number of agents in each world). The probability that the maximum [minimum] throw of the first die is strictly greater than [less than] the maximum [minimum] throw of the second die is close to zero. The case is trickier if agents can have any real-valued utility level.

tails. If the two coins land on the same side, the round is tied. Now suppose that the two agents played infinitely many rounds of this game. We would be *extremely* surprised if we were to find out that one agent had won the game only finitely many times over the course of infinitely many rounds. But if two identical population worlds are strictly ranked by the Weak People Criterion, it must be the case that only finitely many locations are worse off in one world than they are in the other. This is akin to finding out that one agent winning only finitely many games over infinitely many rounds.[192] Finding out that one world is better than another by Pareto is akin to finding out that one of the agents never lost a game!

Arntzenius [6, p. 51-52] points out that if worlds are infinite then almost all 'random walk' world pairs $\langle w_1, w_2 \rangle$ will such that infinitely many agents are better off in w_1 than in w_2 and infinitely many agents are better off in w_2 than in w_1, and will therefore be incomparable by either the Pareto principle or extensions of it like the Weak People Criterion.

The incomparability results of Chapters 2 and 3 can therefore be characterized as an impossibility result showing that we cannot jointly accept the following five axioms:

(1) Transitivity of $\succcurlyeq$ *If $w_1 \succcurlyeq w_2$ and $w_2 \succcurlyeq w_3$ then $w_1 \succcurlyeq w_3$*

(2) Permutation Principle *For any world pair $\langle w_1, w_2 \rangle$ and any bijection g from the population of $\langle w_1, w_2 \rangle$ onto any population, there exists a world pair $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under bijection g.*

(3) Qualitativeness of $\succcurlyeq$ *If the pair $\langle w_3, w_4 \rangle$ is a qualitative duplicate of the pair $\langle w_1, w_2 \rangle$, then $w_3 \succcurlyeq w_4$ if and only if $w_1 \succcurlyeq w_2$*

(4) Pareto Principle *If w_1 and w_2 have identical populations and for all agents x in w_1 and w_2, $u_{w_1}(x) \geq u_{w_2}(x)$, then $w_1 \succcurlyeq w_2$. If there is also some agent x in w_1 and w_2 such that $u_{w_1}(x) > u_{w_2}(x)$, then $w_1 \succ w_2$.*

[192] See Zame [247, p. 51-2].

(5) Minimal Completeness of $\succcurlyeq$ *Comparability between infinite worlds ($w_1 \succcurlyeq w_2$ or $w_2 \succcurlyeq w_1$) is not incredibly rare*

Most of us would want to retain all five of the axioms above, but the arguments of Chapters 2 and 3 show that they cannot be jointly retained. We must therefore determine which of the five axioms is the least plausible. If we want to retain the Pareto principle and Minimal Completeness, for example, then we must reject the Transitivity of $\succcurlyeq$, the Permutation Principle, or the Qualitativeness of $\succcurlyeq$ (we may, of course, reject more than one axiom). In the next section I will consider the consequences of rejecting the first three axioms.

5.2 Transitivity, the Permutation Principle, and Qualitativeness

The four world and cyclic arguments of Chapters 2 and 3 relied on the Transitivity of $\succcurlyeq$, the Permutation Principle, and the Qualitativeness of $\succcurlyeq$. We can therefore block the incomparability results of these chapters if we deny at least one of these three axioms. In this section I will consider the consequences of denying each of these axioms in turn.

5.2.1 The Transitivity of $\succcurlyeq$

The four world and cyclic arguments appeal to the Transitivity of $\succcurlyeq$ when they claim that if $w_2 \succ w_3$ by Pareto and $w_3 \succcurlyeq w_4$ by the Qualitativeness of $\succcurlyeq$ and $w_4 \succ w_1$ by Pareto, then it follows that $w_1 \not\succcurlyeq w_2$. If we reject Transitivity then we would not be entitled to conclude this, since we could instead conclude that $w_2 \succ w_3 \succcurlyeq w_4 \succ w_1 \succcurlyeq w_2$.

While Transitivity was once considered entirely uncontroversial, some objections to this principle have emerged in recent years. The objection that is most relevant to the four world and cyclic arguments is Temkin's 'different criteria' objection.[193] Temkin argues that if multiple criteria are being used to generate a ranking, then those criteria can result in intransitive rankings if criteria are given different weights across comparisons.[194] We might think that this is at play in the four world argument: the verdict that $w_2 \succ w_3$ is driven by the Pareto criterion but the verdict that $w_1 \succcurlyeq w_2$ is driven by some other criterion, such as spatiotemporal configuration. Perhaps spatiotemporal configuration should be given less weight in the comparison between worlds that are Pareto comparable than it is in the comparison between Pareto-incomparable worlds, resulting in intransitivity.[195]

I do not believe that we have good reasons for thinking that this is what is going on in comparisons between infinite worlds. If criterion A is given greater weight when we compare x and y, such that $x \succ y$, and criterion B is given greater weight when we compare y and z, such that $y \succ z$, then surely we should at least be somewhat compelled to maintain that $y \succ z$ even if this conflicts with Transitivity.[196] When we compare infinite worlds, however, the intransitivity results seem to *undermine* our belief that the non-Paretian criterion should be given any weight. Consider the case of Clement and Stormy from Chapter 2:

[193]Temkin's other main objections to Transitivity are based on 'spectrum arguments', which are less relevant to four world and cyclic arguments. For responses, see Norcross [175], Voorhoeve [240], and Nebel [168]. Further arguments against Transitivity can also be found in Rachels [187] and Temkin [224].

[194]This argument can be found in [226, p. 219-221]. Note that if there are multiple criteria but their weight does not vary across comparisons, then the $\succcurlyeq$ relation would remain transitive.

[195]This would make the four world argument a lot like the infinite puzzle discussed by Kagan [119, p. 471-8].

[196]For example, suppose we believe that when comparing one finance job over another, we should give a lot of weight to which has the greater salary and less weight to how pleasant we would find the work. We also believe that when comparing a finance job with a job in the arts, we should give less weight to which has the greater salary and more weight to how pleasant we would find the work. We could end up concluding that having a high paying but unpleasant finance job is better than having a low paying but very pleasant finance job (since it has a higher salary) and having a low paying but very pleasant finance job is better than having a typical job in the arts (since it affords greater pleasure) but that having a typical job in the arts is better than having a high paying but unpleasant finance job (since it affords greater pleasure). If we genuinely believe that these criteria are given different weight across comparisons, then we should continue to find each of these pairwise rankings plausible even after we discover that they are jointly intransitive.

	A	*B*	*C*	*D*	*E*	*G*
Clement	☻	☹	☹	—	—	—
Stormy	—	—	—	☻	☻	☹
Duplicate Clement	—	—	—	☻	☹	☹
Duplicate Stormy	☻	☻	☹	—	—	—

Figure 88: Clement, Stormy, and duplicates again

When we see that Stormy is better than Duplicate Clement by Pareto and that Duplicate Stormy is better than Clement by Pareto, this undermines our initial intuition that Clement is better than Stormy and that Duplicate Clement is better than Duplicate Stormy. Facts about the spatiotemporal configuration of agents suddenly seem to lose their relevance for comparing infinite worlds: they don't strike us as different criteria that should be given more weight when the Pareto criterion can be applied and no weight when it can be applied. If this is correct then we may be using locations-based criteria as some kind of proxy for improving the lives of agents when worlds are not Pareto comparable. Once we realize that locations-based criteria are not proxies for improving lives, we cease to find them compelling.

Temkin also acknowledges that if the goodness of a world is an intrinsic property – the goodness of a world is fixed solely by its 'internal features' [226, p. 229] – then the 'better than' relation would indeed be transitive because the goodness of each world would be fixed by these features: it would not vary depending on what we are comparing it with and can therefore be represented on the same linear scale. Therefore Temkin would presumably agree that the relation 'is at least as good as' is transitive if how good a world is does not vary depending on which world we are comparing it with.

Broome [46, p. 124] argues that the Transitivity of comparative properties like goodness is an undeniable principle of logic. In defense of Transitivity, Broome [46] argues that we can derive the Transitivity of betterness from the fact that goodness is a property of worlds, regardless of whether it is an intrinsic property of worlds. He states:

> Now, this derivation of the transitivity of betterness does not actually require the premise that goodness is an intrinsic property. If A is at least as good as B and B is at least as good as C, then A's goodness is at least as great as B's, and B's is at least as great as C's. Consequently A's goodness is at least as great as C's. So A must be at least as good as C. The basis of this argument is simply that goodness is a property, and that betterness is the comparative of goodness. There is no need for goodness to be an intrinsic property. The comparative of *any* property is necessarily transitive. [46, p. 233]

Broome argues that if different criteria are relevant across comparisons, then we are not using the comparative of a single property: namely the property of goodness. He concludes that if our judgments about worlds are intransitive even upon reflection – we do not drop our initial judgment that Clement is at least as good as Stormy – then those judgments cannot be about the $\succcurlyeq$ relation, since this is the comparative of goodness. [46, p. 234].

Broome's argument is consistent with the claim that $\succcurlyeq$ is an incomplete relation. Broome notes that Transitivity is consistent with what he calls 'hard indeterminacy'. He states: 'Transitivity requires that if x is Fer than y and y is Fer than z, then x is Fer than z. If it turns out that neither x is Fer than y nor y Fer than x, then in this case Transitivity is vacuously satisfied.' [46, p. 126] Broome's argument is also consistent with the claim that $\succcurlyeq$ is a qualitative but not a qualitative internal relation. If goodness were an intrinsic property of worlds, it would follow that if $\succcurlyeq$ is a qualitative relation it is a qualitative internal relation. But, as Broome points out, goodness need not be an intrinsic property.

I will not offer an extensive defense of the Transitivity $\succcurlyeq$ here. Most of us consider this principle so plausible that we will be compelled to reject our judgment that Clement is at least as good as Stormy or to reject our judgment that Stormy is at least as good as Duplicate Clement long before we will doubt our judgment that the 'at least as good as' relation is transitive.[197] The Transitivity of $\succcurlyeq$ seems like one of the axioms we are least likely to deny out of the five axioms listed above.

5.2.2 The Permutation Principle

The Permutation Principle acts as a kind of existence axiom in the four world and cyclic arguments. It states that for any world pair and any bijection from the population of that world pair onto any population of possible people, there exists another world pair that is a qualitative duplicate of the original world pair under that bijection. If we reject the Permutation Principle then we will not always be able to find a qualitative duplicate of $\langle w_1, w_2 \rangle$ that results in a violation of Transitivity. So we will not be able to show that w_1 and w_2 are incomparable even if we accept Transitivity, the Qualitativeness of $\succcurlyeq$ and Pareto.

The Permutation Principle could be interpreted in two ways, depending on whether we take the relevant notion of possibility in the principle to metaphysical possibility or something broader than this, such as logical possibility or some intermediate concept of possibility that is broader than metaphysical possibility but narrower than logical possibility. Two notable views are inconsistent with the stronger, metaphysical interpretation of the Permutation Principle. The first is essentialism. Essentialists believe that agents cannot play any qualitative role whatsoever in a world: for ancestor essentialists they cannot play the role of someone with different ancestry, for example. But facts about ancestry can be qualitative. You have many ancestors who existed before you. Therefore, according to ancestor essentialism, you cannot play the qualitative role of an agent that lacks any ancestors. Any permutation that fails to preserve these essential qualitative properties and relations does not correspond to a metaphysically possible world pair. The second view that is in tension with this claim is

[197] If we reject the first judgment then we must reject either the Permutation Principle, the Qualitativeness of $\succcurlyeq$, or Minimal Completeness. If we reject the second judgment then we must reject the Pareto principle.

anti-haecceitism. Anti-haecceitists believe that a difference in identity entails, as a matter of metaphysical necessity, a difference in some qualitative property or relation at the relevant world pair. This means that no world pair that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under a non-trivial permutation is a metaphysically possible world pair.

Suppose we believe that facts about whether one world is ethically at least as good as another (facts like '$w_1 \succcurlyeq w_2$') are metaphysically necessary but not logically necessary. When combined with the view that there are essential characteristics, this would prevent us from constructing many four world and cyclic arguments. If there are differences between the essential characteristics of different possible people then permuting agents in ways that change their essential characteristics will result in metaphysically impossible world pairs. And if facts of the form 'w_1 is at least as good as w_2' need not extend to metaphysically impossible worlds and the world pair $\langle w_3, w_4 \rangle$ is metaphysically impossible qualitative duplicate of $\langle w_1, w_2 \rangle$, then we can maintain that $w_1 \succcurlyeq w_2$ and deny that $w_3 \succcurlyeq w_4$ without violating the claim that $\succcurlyeq$ is a necessary qualitative relation. If $\succcurlyeq$ need not hold between metaphysically impossible worlds, then the claim that $\succcurlyeq$ is qualitative only entails that if $\langle w_3, w_4 \rangle$ is a metaphysically possible qualitative duplicate of $\langle w_1, w_2 \rangle$, then $w_1 \succcurlyeq w_2$ iff $w_3 \succcurlyeq w_4$.

If we hold this view, the reasons that we had for accepting the Permutation Principle are only reasons for accepting the a version of that principle that is restricted to the qualitative duplicates of a world pair that are metaphysically possible:

Restricted Permutation Principle

For any world pair $\langle w_1, w_2 \rangle$ and any bijection g from the population of $\langle w_1, w_2 \rangle$ onto any population that preserves essential properties and relations, there exists a metaphysically possible world pair $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under bijection g.

The Restricted Permutation Principle can be used to show that infinite worlds are incomparable whenever there is a metaphysically possible permutations of the agents of a world pair: a possible permutation of the agents of the world pair that preserves the essential characteristics of the populations. Note that unless we have an extremely broad account of essential characteristics – including features like spatial location and birth order – we will often be able to find such a metaphysically possible world pair. Therefore rejecting the Permutation Principle on these grounds will not entirely recover completeness. It will, however, reduce the amount of incompleteness that can be established. The more properties and relations that are metaphysically essential, the fewer world pairs that can be shown to be incomparable by a four world or cyclic argument using the Restricted Permutation Principle. Just what qualitative properties are essential is an open question, but there are few who would claim that an agent can play any qualitative role whatsoever.[198]

The term 'fragility' is used to describe the extent to which our identity depends on the qualitative properties and roles that we have. One extreme is that identity is not fragile at all: our identity does not depend on any of our qualitative properties whatsoever. The other extreme is that identity is perfectly fragile: any change in qualitative properties will result in a change in identity. This view is clearly inconsistent with the Permutation Principle.[199]

If identity is 'perfectly fragile' across all qualitative dimensions (psychology, genetics, etc.) then any change in qualitative properties will result in a change in identity. On this view I am just the sum of my qualitative parts. I could not have been happier than I currently am because this would mean that I have different qualitative properties than those I actually

[198]Suppose world w_1 contains infinitely many happy humans and infinitely many sad humans and w_2 contains infinitely many happy hamsters and infinitely many sad hamsters. In order to show that $w_1 \not\succeq w_2$ by a four world argument, we would have to suppose that each of the humans in w_1 can play the qualitative role of a hamster in w_4. Many of us will be inclined to think that this is not possible. We need not restrict ourselves to objects capable of happiness and sadness when making this point if we extend the Permutation Principle to permutations of all objects.

[199]If we hold this view of personal identity then we must believe that world w_3 can only be a qualitative duplicate of world w_2 that contains the agents of w_1 if world w_1 is a qualitative duplicate of w_2.

have, and this is not possible.[200] This view is not only inconsistent with the Permutation Principle: it also entails that no world can be strictly better or worse than another world by Pareto because it is not possible for identical populations to have distinct utility levels.[201]

The primary response to the this argument against the Permutation Principle is simply that we have no reason for thinking that identity is perfectly fragile across all qualitative dimensions. How robust or fragile our identity is remains an open question, but the claim that identity is perfectly fragile is in tension with our intuitions about cases and our ordinary use of the term. Surely you could have been slightly happier when you woke up this morning. Surely you could have had slightly shorter hair than you currently do.[202] For this reason many, including Hare [102, Ch. 7], simply assume that identity (or essence) is not perfectly fragile across all qualitative dimensions. I will also assume that this is not the case.[203]

The claim that qualitative duplicates under a permutation are metaphysically possible is also in tension with anti-haecceitism. Consider Lewis's characterization of anti-haecceitism:

[200]The claim that identity is perfectly fragile is consistent with the claim that there can be identity across time. If agents have temporal parts as Lewis [153] argues then the view that identity is perfectly fragile entails that people could not have had temporal parts that are different from those that they actually have.

[201]Those who believe that identity is perfectly fragile could in principle accept if the Pareto principle is metaphysically necessary then it is vacuously satisfied: it just entails that two qualitatively identical worlds are equally good. But we cannot use Pareto to generate four world or cyclic arguments because these arguments involve qualitatively distinct worlds that contain identical populations, which is impossible.

[202]Those who believe that identity is perfectly fragile could adopt an error theory about ordinary identity talk, but the fact that our concept of identity is not perfectly fragile is a pro tanto reason against the view. If we accept that identity is perfectly fragile then we are still faced the difficult task of producing a suitably plausible and complete ranking of infinite worlds.

[203]If essence is not perfectly fragile then we Hare's [103] [102, Ch. 11] 'morphing' argument can be raised as a challenge for essentialist and anti-haecceitists. For any world pair $\langle w_1, w_2 \rangle$ we can find a sequence of (non-duplicate) world pairs $\langle w_1', w_2' \rangle$, $\langle w_1'', w_2'' \rangle$, and so on such that w_1' has the same population as w_2 but the agents of w_1' play qualitative roles that are slightly closer to those played by agents in w_1, and so on. Suppose that at some point in this sequence we reach a world pair $\langle w_1^{'',...,'}, w_2^{'',...,'} \rangle$ such that $w_1^{'',...,'}$ contains the same agents as w_2 but playing the qualitative roles played by agents in w_1 and $w_2^{'',...,'}$ contains the same agents as w_1 but playing the qualitative roles played by agents in w_2. But if we accept that each 'almost duplicate' and its predecessor contain identical populations, we can use such sequences to generate violations of Transitivity across metaphysically possible world pairs.

> 'If two worlds differ in what they represent *de re* concerning some individual, but do not differ qualitatively in any way, I shall call that a haecceitistic difference. Haecceitism, as I propose to use the word, is the doctrine that there are at least some cases of haecceitistic difference between worlds. Anti-haecceitism is the doctrine that there are none.' [152, p. 221]

Haecceitism, as Lewis understands it, is the view that two possible worlds can be qualitatively identical in all respects and yet different individuals can be represented in those worlds. The claim that it is metaphysically possible to permute the population of a disjoint world pair so that a world w_3 contains the same agents as w_1 even though w_3 is qualitatively identical to w_2 would seem to commit us to haecceitism as Lewis understands it.

Anti-haecceitism does not commit us to the claim that the same agent could not have had different qualitative properties and relations.[204] However, anti-haecceitists will reject the claim that for some non-trivial bijection from the population of a world pair onto another population, there exists a metaphysically possible qualitative duplicate of that world pair under that bijection, since this requires some haecceitistic difference between the two world pairs. Lewisian anti-haecceitists will therefore reject that the claim that permuting the population of a world pair results in a metaphysically possible qualitative duplicate.

In summary: essentialists will reject the claim that all qualitative duplicate world pairs under a non-trivial permutation are metaphysically possible because they believe that certain qualitative properties are necessary for identity and these will not always be preserved by the permutation in question. Anti-haecceitists will reject the claim that any qualitative duplicate world pairs under a non-trivial permutation are metaphysically possible because

[204]Lewis's [152, §4.3] view is that the same individual cannot exist in different possible worlds but that the same individual can be be represented *de re* in different worlds even if their qualitative properties are different. Representations of the same individual in different possible worlds are the individual's counterparts. Therefore to say that I could have been happier yesterday is to say that I have a counterpart in a possible world that has this property. We can need not adopt this counterpart-theoretic framework, however. Section 1.3 of Hare [103] discusses similar issues without appeal to counterpart theory.

the believe that qualitative differences are necessary for differences in identity. They will therefore deny that we can produce a qualitative duplicate world pair under a permutation. So it seems that all but the most extreme haecceitists will agree that many permutations of a population will result in metaphysically impossible qualitative duplicates.

In order for this to constitute an objection to the Permutation Principle, however, we must also hold that the domain of the $\succcurlyeq$ relation does not include metaphysically impossible worlds. I believe that we should reject this crucial claim. The claim that facts about the $\succcurlyeq$ relation holds necessarily between some metaphysically impossible worlds seems to be consistent with what we generally take to be entailed by claims involving the $\succcurlyeq$ relation.

Suppose that the following is a fact: if w_1 and w_2 contain only a single agent and that agent is happy in w_1 and miserable in w_2 then $w_1 \succ w_2$. Suppose that John is happy in one world and miserable in another and no other agents exist in these worlds. It seems we must conclude that the first world is better than the second. But now suppose that you found out that John was born to different parents in the second world than in the first. Would we now feel in no way compelled to say that the first world is better than the second? Surely not. If we believe that if w_1 and w_2 contain only a single agent and that agent is happy in w_1 and miserable in w_2 then $w_1 \succ w_2$, then it seems to follow that the world in which John is happy is better than a world in which John is sad and was born to different parents. The second world may not be metaphysically possible but the world pair still has a property that we believe is sufficient for the better than relation to hold between them.

Claiming that the domain $\succcurlyeq$ includes metaphysically impossible worlds does not commit us to the claim that the relation is logically necessary since its domain need not include all logically possible worlds. It is even possible that the domain of the $\succcurlyeq$ relation includes some metaphysically impossible worlds but that it does not include all metaphysically possible worlds.[205] All that we require is that the $\succcurlyeq$ relation holds necessarily between any qualitative duplicate of a world pair $\langle w_1, w_2 \rangle$, even if these worlds are metaphysically impossible.

It seems plausible that for any world pair $\langle w_1, w_2 \rangle$ and any bijection g from the population of $\langle w_1, w_2 \rangle$ onto any population, there is a logically possible world pair $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under g even if the metaphysical possibility of $\langle w_3, w_4 \rangle$ is not guaranteed. Rejecting the Permutation Principle therefore rests on the claim that facts of the form 'necessarily, if w_1 and w_2 stand in relation R to one another then $w_1 \succcurlyeq w_2$' can be true even if two logically possible but metaphysically impossible worlds w_1 and w_2 stand in relation R to one another and $w_1 \not\succcurlyeq w_2$. If we believe that the $\succcurlyeq$ relation is a necessary relation that holds between these worlds even if they are metaphysically impossible, then it is difficult to see any grounds on which we could challenge the Permutation Principle.[206]

If we reject the claim that the domain of $\succcurlyeq$ includes metaphysically impossible world pairs, we can still show that worlds w_1 and w_2 are incomparable if there is a bidirectional upgrade from w_1 to w_2 and from w_2 to w_1 under a restricted permutation. This constrains the class of worlds that can be shown to be incomparable. Before concluding this section, I wish to point out that many infinite wolrds can be shown to be incomparable using restricted permutations if we adopt a plausible strengthening of the claim that $\succcurlyeq$ is a qualitative relation.

Suppose that $w_1 \succcurlyeq w_2$. The Permutation Principle guarantees that we can find a pair

[205] A detailed discussion of the relationship between metaphysical and normative necessity is beyond the scope of this document. For more on this topic, see [81], [130, §5.3.2], and [150].

[206] Those who insist that the $\succcurlyeq$ relation is at most metaphysically necessary face the difficult task of showing how facts about haecceities can be used to generate an adequately complete ranking of infinite worlds that is not entirely ad hoc or inequitable. Haecceitistic rankings are not difficult to generate. We could, for example, rank infinite worlds based entirely on Obama's happiness levels (and be indifferent between worlds in which Obama does not exist and worlds at which he has zero utility) but this would hardly be plausible. Those who advocate using haecceitistic facts to rank infinite worlds must also offer an account of how haecceitistic relations hold between existent and non-existent objects in order to rank non-identical populations. Whether a haecceitistic relation holds between an existent and non-existent object may be subject to vagueness, which would mean that we could at best yield only vague rankings of infinite worlds from such haecceitistic rankings. There are clearly great difficulties facing anyone who chooses to go down this path.

of worlds $\langle w_3, w_4 \rangle$ under a permutation that is a duplicate of $\langle w_1, w_2 \rangle$ in absolutely every qualitative respect. This is important because the claim that $\succcurlyeq$ is a necessary qualitative relation entails that $w_3 \succcurlyeq w_4$ only if this pair is a perfect qualitative duplicate of $\langle w_1, w_2 \rangle$. We cannot conclude that $w_3 \succcurlyeq w_4$ if the pair differs in any qualitative respect, even if the difference between the pairs seems entirely ethically irrelevant. For example, suppose that $\langle w_3, w_4 \rangle$ is qualitatively identical to $\langle w_1, w_2 \rangle$ in all respects except the following: in w_1 there is a blade of grass that is wilting, while in w_3 the same blade of grass is not wilting. Since $\langle w_3, w_4 \rangle$ is not a perfect qualitative duplicate of $\langle w_1, w_2 \rangle$ we can no longer conclude from the qualitativeness of $\succcurlyeq$ that $w_3 \succcurlyeq w_4$. Yet surely if $w_1 \succcurlyeq w_2$ and $\langle w_3, w_4 \rangle$ differs from $\langle w_1, w_2 \rangle$ in this ethically irrelevant respect only, we must conclude that $w_3 \succcurlyeq w_4$.

It seems uncontroversial to claim that if $\langle w_3, w_4 \rangle$ differs from $\langle w_1, w_2 \rangle$ only in respects that are ethically irrelevant, such as the dry blade of grass, then if $w_1 \succcurlyeq w_2$ it must also be the case that $w_3 \succcurlyeq w_4$. If this is correct then we are in a position to endorse a claim is stronger than the claim that $\succcurlyeq$ is a necessary qualitative relation, namely the following:

Ethically Relevant Qualitativeness of $\succcurlyeq$ *If the pair $\langle w_3, w_4 \rangle$ is a qualitative duplicate of the pair $\langle w_1, w_2 \rangle$ in all ethically relevant respects, then $w_3 \succcurlyeq w_4$ if and only if $w_1 \succcurlyeq w_2$*

We need not take strong stand on which qualitative properties and relations are ethically relevant or irrelevant. Many qualitative properties, such as agent utilities, relationships between agents, and so on, are plausibly ethically relevant. But the ethical irrelevance of some qualitative properties and relations, such as the color of an agent's hair or the name of a street or the most popular flavor of jam, seems to be uncontroversial. The principle above commits us to the view that if one world pair differs from another in only these ethically irrelevant respects, whatever these turn out to be, then $\succcurlyeq$ relation holds between the first pair of worlds if and only if it holds between the second pair of worlds.

If we accept the Ethically Relevant Qualitativeness of $\succcurlyeq$ and there is a bidirectional upgrade from w_1 to w_2 and from w_2 to w_1 then we can show that w_1 and w_2 are incomparable if there exists a metaphysically possible world pair $\langle w_3, w_4 \rangle$ that is an ethically relevant qualitative duplicate of $\langle w_1, w_2 \rangle$ under some bijection g. This greatly reduces the extent to which anti-haecceitism represents a challenge to the incomparability results. Suppose that $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ in all ethically relevant respects and w_3 contains the same population as w_2 and w_4 contains the same population as w_1. Anti-haecceitists objected to this possibility in pairs that are perfect qualitative duplicates because it required that there can be a difference in the identity of agents between w_1 and w_3 (and w_2 and w_4) without there being a qualitative difference between the two worlds. This is not true, however, if the world pair is an ethically relevant qualitative duplicate rather than a perfect qualitative duplicate. Assume that w_1 contains men with red hair and w_2 contains women with blond hair. If gender and hair color are not ethically relevant then we can suppose that the identical populations of w_2 and w_3 are composed of women that have blond hair and that the identical populations of w_4 and w_1 are composed of men that have red hair. Since we are assuming that gender and hair color are not ethically relevant, this is consistent with the claim that w_1 and w_3 (and w_2 and w_4) are qualitatively identical in all ethically relevant respects.

The only world pairs that meet the conditions of the results of Chapter 2 but that cannot be shown to be incomparable using the Restricted Permutation Principle and the Ethically Relevant Qualitativeness of $\succcurlyeq$ are those in which there are ethically relevant essential properties or relations that some agent in w_1 has and that no agent in w_2 has. Suppose that 'having ancestors' is an essential property that the agents in world w_1 have and the agents in world w_2 lack. If 'having ancestors' is not ethically relevant then we can find a metaphysically possible world pair $\langle w_3, w_4 \rangle$ that is a qualitative duplicate of $\langle w_1, w_2 \rangle$ in all ethically relevant respects in which w_2 and w_3 have identical populations of agents with no ancestors and w_4

and w_1 have identical populations of agents with ancestors. But suppose that it is good to honor your ancestors if you have ancestors and that the agents of w_1 all honor their ancestors while the agents of w_2 do not (since they do not have any ancestors). If w_3 is a qualitative duplicate of w_1 in all ethically relevant respects then this would require that the agents of w_3 all honor their ancestors, which is not possible since they do not have any. Therefore if having ancestors is an ethically relevant essential property, the pair $\langle w_3, w_4 \rangle$ cannot be a qualitative duplicate of $\langle w_1, w_2 \rangle$ in all ethically relevant respects.

Whether there are any ethically relevant essential properties and how common these properties are depends on both what kind of essentialism or anti-haecceisitism is true, if any, and what qualitative properties are ethically relevant. Answering these questions are not always trivial. For example, suppose that species essentialism is true. This might seem like a plausible candidate for an ethically relevant essential property because, on many moral theories, species is treated as an ethically relevant property. Humans may have rights that birds lack and the wellbeing of humans may be given more moral weight than the wellbeing of birds. But the property of 'being a human' may not actually be ethically relevant according to these theories. Instead, other properties that humans have and birds lack may be what truly matters. For example, if there existed a bird with the same capacity to reason, the same capacity to foresee the future, and the same capacity to suffer as a human, then many moral theories that generally distinguish between birds and humans would not distinguish between this bird and humans.[207] If this is the case then it is not 'being human' that is ethically relevant, but one or more of these underlying properties. And species essentialism may be true while essentialism about these underlying properties is not.

I cannot offer a compelling answer to the question of whether and to what extent there are

[207] Whether there are any ethically relevant qualitative properties that can ground distinct moral treatment of human and non-human animals is the subject of much debate. Singer [209] argues that none of the qualitative properties that none of the commonly cited properties of humans is able to play this role. And the property of 'being human' is not, by itself, considered sufficient grounds for distinct moral treatment.

ethically relevant essential properties. But I believe we can say with some confidence that world pairs in which there are ethically relevant essential properties or relations that some agent in w_1 has and that no agent in w_2 has are likely to be far less common than world pairs in which there are any essential properties or relations that some agent in w_1 has and that no agent in w_2 has if we consider the most commonly held views about which qualitative properties are essential and which qualitative properties are ethically relevant. If this is correct then endorsing the Ethically Relevant Qualitativeness of $\succcurlyeq$ seems to greatly weaken the objection to the Permutation Principle on essentialist and anti-haecceitist grounds.

5.2.3 The Qualitativeness of $\succcurlyeq$

In the four world and cyclic arguments it is assumed that $\succcurlyeq$ is a necessary qualitative relation. If $\succcurlyeq$ is a non-qualitative relation then even if $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ under some permutation of the population of $\langle w_1, w_2 \rangle$, it is possible that $w_1 \succcurlyeq w_2$ but $w_3 \not\succcurlyeq w_4$. Therefore if we deny that $\succcurlyeq$ is a qualitative relation then we can block the four world and cyclic results even if we accept Transitivity and the Permutation Principle.

If $\succcurlyeq$ is not a qualitative relation then then whether it holds between two worlds depends not only on facts about whether the same people or different people exist at a given world pair, but on facts about the identities of the agents that exist in each world. For example, it's possible that $\langle w_3, w_4 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ but $w_1 \succcurlyeq w_2$ and $w_3 \not\succcurlyeq w_4$ simply because one of the agents in $\langle w_1, w_2 \rangle$ is Obama and the agent playing the qualitative role of Obama in $\langle w_3, w_4 \rangle$ is merely a qualitative duplicate of Obama and not Obama himself.

There are two key arguments against the claim that $\succcurlyeq$ is a non-qualitative relation. The first is that, as I argued in Chapter 2, the Qualitativeness of $\succcurlyeq$ seems to capture a very basic notion of equity. Those who accept Pareto acknowledge that whether one world is better than another is sensitive to facts about whether the two worlds contain the same agents or not. But if ethics is to be equitable then it seems that, at a minimum, it should be insensitive to facts about which *particulars* exist at those worlds. Ethics should be sensitive to the fact that the same agent exists in w_1 and w_2 and is better off in w_1 than in w_2, but it should not be sensitive to the fact that the agent in question is Obama. This notion of equity is captured by the claim that $\succcurlyeq$ is a necessary qualitative relation. The claim that $\succcurlyeq$ is a non-qualitative relation is in tension with this basic notion of equity.

A further argument against the claim that $\succcurlyeq$ is a non-qualitative relation is simply that it is not clear how we could possibly use haecceitistic facts to produce a ranking of infinite worlds that avoids the incomparability results that is not entirely ad hoc, even if we did accept that haecceitistic facts can affect the ethical ranking of worlds.[208] Given this, I do not believe that rejecting the Qualitativeness of $\succcurlyeq$ is a viable option for those wishing to avoid the incomparability results of the preceding chapters.

5.3 Rejecting Pareto

A final option for avoiding ubiquitous incomparability is to reject the Pareto principle. If we reject Pareto then we can deny that Stormy is better than Duplicate Clement and that Duplicate Stormy is better than Clement in the case from Chapter 2, thus blocking the four world result in this case. This generalizes to all of the four world and cyclic results of Chapter 3. If we reject Pareto then we can then attempt to identify qualitative features

[208] Haecceitistic rankings are not difficult to generate. We could, for example, rank infinite worlds based entirely on Obama's happiness levels and be indifferent between worlds in which Obama does not exist and worlds at which he has zero utility, but this would hardly be plausible or equitable. Those who advocate using haecceitistic facts to rank infinite worlds must also offer an account of how haecceitistic relations hold between existent and non-existent objects in order to rank non-identical populations. Whether a haecceitistic relation holds between an existent and non-existent object may be subject to vagueness, which would mean that we could at best yield only vague rankings of infinite worlds from such haecceitistic rankings.

of world pairs, such as agents' spatiotemporal locations and utility levels, that can be used produce something close to a complete and transitive ranking of infinite worlds. For example, in the case of Stormy and Clement, Expansionism entails that the four worlds can be ranked as follows: Clement $\sim$ Duplicate Clement $\succ$ Stormy $\sim$ Duplicate Stormy.[209]

I have argued that Pareto is a fundamental principle of ethics. Much of the discussion of this principle has occurred in previous chapters. In Chapter 2, I considered an objection to Pareto offered by Hamkins and Montero. They argue that we should reject Pareto because it is inconsistent with the claim that a world is as good as a copy of that world which preserves 'the topological structure of locations and and the amount of local goodness at those locations' of a world [100, p. 235]. I showed that this argument assumes that $\succcurlyeq$ is a qualitative internal relation, which we have no reason to believe.

The structure of this section is as follows. First, I will ask whether there are independent reasons to reject the Pareto principle. I consider what I believe constitute the most noteworthy independent reasons to reject Pareto: objections to the moral importance of persons rather than experiences in the principle, and objections to the principle based on its tension with distributive principles. Second, I will explore what options are available to us if we do reject Pareto on independent grounds and how attractive these non-Paretian views are. I do so because, although we should want to look for independent reasons to give up Pareto, how compelled we will be to give up the axiom also depends on how attractive the alternatives to Pareto are. I will argue that although non-Paretian methods for ranking infinite worlds produce intuitively acceptable rankings that are either total orderings or at least generate less incomparability than Pareto does, they are not plausible on reflection.

[209] Some qualitative properties, such as the midpoints of the accumulation intervals of worlds, may be used to produce a total order over all infinite world pairs. (See Armstrong [1] for another proposed method for producing a ordering using accumulation intervals.) Other qualitative properties may be used reduce but not entirely eliminate incomparability between infinite worlds.

The core assumption of the Pareto principle is the claim that people are the 'basic locations of value' in our sensitivity principles (principles that prevent us from being indifferent about increases in utility at these locations). If people are not the sort of thing that our ethical theories can or should be sensitive to then we ought to reject the Pareto principle.

One prominent view that is in tension with the Pareto principle is the Reductionist view of personal identity: the view that facts about personal identity can be reduced to further facts, typically about physical or psychological continuity, that can be described wholly impersonally [179, p. 210-211] and that 'the existence of a person to involve nothing more than the occurrence of interrelated mental and physical events' [179, 341]. Parfit characterizes the ramifications of Reductionist beliefs about personal identity as follows:

> 'These [Reductionist] beliefs support certain moral claims. It becomes more plausible, when thinking morally, to focus less upon the person, the subject of experiences, and instead to focus more upon the experiences themselves. It becomes more plausible to claim that, just as we are right to ignore whether people come from the same or different nations, we are right to ignore whether experiences come within the same or different lives.' [179, p. 341]

> 'If we cease to believe that persons are separately existing entities, and come to believe that the unity of a life involves no more than the various relations between the experiences in this life, it becomes more plausible to be more concerned about the quality of experiences, and less concerned about whose experiences they are.' [179, p. 346]

Parfit [179, p. 339] argues that if we hold this view of personal identity then those do not think that the equal distribution of benefits at different times is of moral importance independent of its effects should also not believe that the equal distribution of benefits for different people is of moral importance independent of its effects. He appeals to an analogy between persons and nation states. If we could yield a greater benefit to the people of one nation over another then many of us would not see the fact that the first nation happened to have a history less full of of suffering than the second nation as a reason to relieve the suffering of fewer people in the second nation today. He argues that, by analogy, 'We may believe that, when we are trying to relieve suffering, neither persons nor lives are the morally significant unit. We may again decide to aim for the least possible suffering, whatever its distribution.' [179, p. 341].

Parfit does not deny the existence of persons or claim that identity is perfectly fragile. Instead, he argues that whether an instance of suffering occurs in the same person or in a different person is no more morally relevant than whether an instance of suffering occurs at one time or at a different time. In worlds that contain finite utility, there Parfit's position is consistent with the Pareto principle. For example, the utilitarian criterion 'world w_1 is better than world w_2 iff the total utility of world w_1 is greater than the total utility of world w_2' is clearly consistent with Parfit's claims about the role of personal identity in ethics. And if world w_1 and world w_2 contain finite utility then whenever w_1 is better than w_2 by Pareto, w_1 is better than w_2 by the utilitarian criterion.

Tension arises between Parfit's claim that persons are not morally significant units and the Pareto principle when we compare worlds that contain infinite utility. Suppose that world w_1 and world w_2 both contain infinitely many days of subjective experience that are at utility level 1 and infinitely many days of subjective experience that are at utility level -1. If we believe that persons and lives are not morally relevant units, then whether these days of subjective experience occur in one agent or another agent should make no difference to our evaluation of w_1 and w_2. We may conclude that w_1 and w_2 are equally good (perhaps because both have infinitely many utility 1 and infinitely many utility -1 days of subjective experience) or that w_1 and w_2 are incomparable (perhaps because the total utility of each world is not well defined) or that w_1 is better than w_2 or vice versa (perhaps because we accept Expansionism or some other principle). The key point is that our evaluation is not sensitive to which agents these subjective experiences occur in.

Suppose that in w_1 each agent lives for three days: two days at utility -1 and two days at utility 1. In w_2 the same agents live for three days: two days at utility 1 and two days at utility -1. World w_2 is better than world w_1 by Pareto. If this were to coincide with our person-indifferent evaluation – e.g. w_2 is better than w_1 by Expansionism – then we could construct a symmetrical case that is inconsistent with our person-indifferent evaluation. Therefore the claim that people or lives are not morally relevant units is in tension with the Pareto principle when we are comparing worlds that contain infinite utility.[210]

One response to this Reductionist objection to Pareto is to reject Parfit's claim that the Reductionist view entails that people or lives are not morally relevant units. Parfit concludes from fission cases – cases in which each of an agent's hemispheres are transplanted into two bodies such that each half is psychologically continuous with the original whole brain – that the continued existence of someone psychologically continuous with ourselves ('survival') matters but identity does not. Lewis [153] offers an alternative four-dimensionalist account of personal identity.[211] Lewis argues that there are two temporally extended agents in each of the hemispheres that coincide prior to fission and that cease to coincide after fission. This view is consistent with the Reductionist view of personal identity and, on this view, the identity relation holds between the whole brain and the two transplanted hemispheres. Those who adopt a view like Lewis's can maintain that people and lives are morally relevant units without rejecting the Reductionist view of personal identity.[212]

[210] Note that even when we are comparing infinite worlds, Pareto is consistent with Parfit's view that how much an agent has suffered in the past is not morally significant.

[211] Not all four-dimensionalists will endorse Lewis's account of personal identity. For an in depth discussion of the distinction between four-dimensionalism and three-dimensionalism, see Sider [207, Ch. 3].

[212] An interesting question that arises is how we should count the utility experienced by coinciding agents. Suppose that world w_1 contains infinitely many brains that exist for one day at utility 1 and then experience fission and on day two each hemisphere experiences utility 1. Is this world composed of infinitely many agents with utility 2 lives, or infinitely many agents with utility 1.5 lives, or something else? See Briggs & Nolan [43] for a discussion of these issues.

If we do decide to reject Pareto because we believe that identity is perfectly fragile (as discussed in section 5.2.2) or because we believe that people are not morally relevant units, an attractive alternative to the view that people are 'basic locations of value' may be Parfit's [179, p. 344] view that 'people's experiences at each particular time' are the basic location of value. On this view, whether these subjective experiences occur 'in the same agent' is ethically irrelevant.[213] If a world contains agents with lives that are composed of four subjective periods of time s_1, s_2, s_3, and s_4, then all that matters is the amount of utility being experienced at s_1, s_2, s_3, and s_4 regardless of the order these periods come in or who is experiencing them.[214]

As we have seen, this view is in tension with Pareto because in infinite worlds it is possible to increase the utility experienced by some agents without making anyone worse off without increasing the cardinality of positive utility subjective experiences in the world. Suppose that world w_1 and world w_2 are both composed of infinitely many days in which there is one agent experiencing a utility 1 day. In w_1, each agent in the world lives for a single utility 1 day before dying. In w_2, the same agent lives for 100,000 utility 1 days before dying. Since w_1 and w_2 contain identical populations and each agent experiences 99,999 more utility 1 days in w_2 than in w_2, world w_2 is strictly better than world w_1 by Pareto. But since w_1 and w_2 are both composed of infinite streams of utility 1 subjective experiences, these two worlds will not be differentiated by the subjective experience view.

Those who endorse the subjective experiences view must either endorse or reject an analogue of Pareto for subjective experiences: if w_1 and w_2 contain the same subjective experiences and every subjective experience in w_1 is at least as good as it is in w_2 then w_1 is at least

[213]Since this view ascribes no importance to the distinction between one person and the next. It therefore represents a total rejection of the 'separateness of persons' – the claim that ethics should take seriously the distinction between persons. Rawls [189, p. 26-7] famously argues that Utilitarianism does not respect the separateness of persons'. See Norcross [?] for a critical response to this objection to Utilitarianism.

[214]The view may also be indifferent to the length of each period, as long as they are all of equal length.

as good as world w_2, and if some subjective experiences in w_1 are strictly better than they are in w_2, then w_1 is strictly better than w_2. Let us call this 'Experiences Pareto'. It is not entirely clear what it means for worlds to contain the same subjective experiences, but suppose that defenders of the subjective experiences view offer some plausible account of the modal profile of subjective experiences.[215]

Suppose that those who adopt the subjective experience view endorse Experiences Pareto. It is worth noting that even fewer worlds will be comparable by this principle than they are by Pareto. Experiences Pareto requires not only that a world contains the same agents but that it contains the same experiences. It requires not only that no agent has a worse life in w_1 than in w_2, but that no experience is worse in w_1 than in w_2. These conditions will rarely be satisfied. Moreover, unless they reject Transitivity, the Permutation Principle, or the Qualitativeness of $\succcurlyeq$, it may be possible to use Experiences Pareto to generate even greater incomparability between infinite world pairs using four world and cyclic arguments, since two worlds will be incomparable if there is a biconditional upgrade from the subjective experiences of w_1 to the subjective experiences of w_2. Many infinite world pairs will have this property, including many world pairs that are comparable by Pareto.[216]

Suppose that those who endorse the subjective experience view reject Experiences Pareto. The problem is that we are still left with no obvious way to rank worlds that contain infinitely many experiences at some upper level of utility and infinitely many experiences at some lower level of utility. Therefore those who accept the subjective experiences view and reject

[215] Perhaps we could claim that the subjective experiences of agent x on day n of their life in w_1 are the same as the subjective experiences of agent x on day n in their life in w_2 or that the same subjective experiences are just experiences with the same phenomenological properties. I take no stand on this here.

[216] Suppose that w_1 contains infinitely many agents with utility 2 lives composed of three subjective experiences of utility 1, utility 0, and utility 1 (in that order), while w_2 contains the same agents with utility 1 lives composed of three subjective experiences of utility 0, utility 1, and utility 0 (in that order). Suppose also that the subjective experiences of agent x on day n of their life in w_1 are the same as the subjective experiences of agent x on day n in their life in w_2. These worlds are comparable by Pareto, but can be shown to be incomparable by a four world argument if we accept Experiences Pareto plus our auxiliary axioms.

Experiences Pareto must supplement their view with some method for ranking infinite worlds if want to establish some minimal level of comparability between infinite worlds.

A very different independent objection to Pareto principle that deserves consideration is that the Pareto principle is in tension with certain distributive principles in ethics. Those who endorse distributive principles believe that when we are ranking worlds the distribution of utility across agents in those worlds matters. They may reject the claim that making some agents better off and no agents worse off necessarily makes a world better (rejecting the $\succ$ component of Pareto) or that if we make no agents worse off then we do not make the world worse (rejecting the $\succcurlyeq$ component of Pareto) because in doing so we can negatively affect the distribution of utility. Objections to Pareto based on distributive principles are consistent with the claim that people are 'basic locations of value'.

Consider the following distributive claims, the 'strong' claim that is inconsistent with the $\succ$ condition of Pareto and the 'weak' claim that is inconsistent with the $\succcurlyeq$ condition of Pareto:

Strong Distributive Claim[217]

It is possible that w_1 and w_2 contain the same agents and some agents are better off in w_2 than in w_1 and no agents are better off in w_1 than in w_2 but $w_2 \not\succ w_1$ because the distribution of utility in w_2 than is less equal than the distribution of utility in w_1

Weak Distributive Claim

It is possible that w_1 and w_2 contain the same agents and no agents are better off in w_2 than in w_1 but $w_1 \not\succcurlyeq w_2$ because the distribution of utility in w_1 is less equal than the distribution of utility in w_2

[217] The Strong Distributive Claim assumes that it is possible for us to improve the welfare of some agents while reducing the equality of the utility distribution across all agents, but if we vastly increase the utility of the agent that already has the highest utility level in a world, we will increase the inequality the distribution of utility in that world according to most standard measures of inequality. See De Maio [65] for a survey of standard measures of income inequality in economics.

The Strong Distributive Claim conflicts with Pareto because it entails that if $w_1 \succ w_2$ by Pareto but the distribution of utility in w_1 is less equal than the distribution of utility in w_2, then w_1 may not be strictly better than w_2. The Weak Distributive Claim conflicts with Pareto because it entails that if $w_1 \succcurlyeq w_2$ by Pareto but the distribution of utility in w_1 is less equal than the distribution of utility in w_2, then w_1 may not be at least as good as w_2.

Many distributive principles do not entail either of these distributive claims. For example, the 'average utilitarian' claim that w_1 is better than w_2 if and only if the average utility of the population of w_1 is greater than the average utility of w_2 does not entail either claim. If w_1 is strictly or weakly better than w_2 by Pareto then the average utility at w_1 must be at least as great as the average utility at w_2.[218] There are, however, distributive principles that entail one or both of these claims and that therefore conflict with Pareto.

One example of a distributive principle that entails the Strong Distributive Claim (and is therefore inconsistent with Pareto) is the following version of Maximin: world w_1 is strictly better than world w_2 only if the life with the minimum utility level in w_1 is strictly greater than the life with the minimum utility level in w_2.[219] This Maximin principle entails the Strong Distributive Claim because if we improve the lives of those who are not the worst off then this does not result in a better world according to this Maximin principle. This is an atypically strong formulation of Maximin, however. Often Maximin is formulated in such a way that it does not adjudicate between worlds if the worst off agents are not affected. In such worlds Leximin is often proposed as an alternative.[220] And the Leximin principle does not entail the Strong Distributive Claim or conflict with Pareto.

Other principles that entail the Strong Distributive Claim include egalitarian principles that

[218]See Tungodden and Vallentyne [229] for a discussion of Paretian egalitarian theories.

[219]The original Maximin criterion is a criterion for decision making under uncertainty. It identifies the value of a gamble with its lowest possible payoff.

[220]See Tungodden [228, p.415] [227] for a discussion of these principles.

are committed to 'Leveling Down': the claim that if the welfare of those with utility levels at the higher end of the distribution is reduced, this is either better than or at least not worse than the original distribution.[221] The fact that certain views, such as prioritarianism, entail Leveling Down has generally been considered an objection to those views, though Temkin [222] argues that the Leveling Down objection is far from fatal.

If the Weak Distributive Claim is true then one world can fail to be at least as good as another even though all of the agents have at least as much utility in the first world as they do in the second. This also seems most plausible if we accept an egalitarian or prioritarian theory that embraces Leveling Down, since the principle implies that we can make a world better without improving the lives of any of its agents. I will therefore assume that the sort of distributive principles we have in mind are principles that tolerate a certain degree of Leveling Down in order to increase the equality of the distribution of utility.[222]

This gives us some reasons to be wary of Pareto principles in finite worlds. Showing that there is a conflict between Pareto and the Strong Distributive Claim or the Weak Distributive Claim in infinite worlds is not so easy, however. Suppose that we have two distributions of utility and we are trying to determine which of them is more equal. In finite worlds, a distribution is typically considered less equal the more that individual's utility levels tend vary from the utility levels of other agent world. For example, Rabinowicz [186, p. 61] proposes that 'the degree of inequality in a social state X is then just the average pairwise inequality in X, i.e., the average welfare distance between the individuals in X.' It is not clear how we should extend such definitions to worlds that contain infinitely many agents. Consider two possible distributions of infinite utility: $(2, 2, 2, ..., 1, 1, 1, ..., 0, 0, 0, ...)$ and $(2, 2, 2, ..., 0, 0, 0, ...)$. We might be inclined to think that the first distribution is more

[221] See Rabinowicz [185, p.81] on the leveling down objection.

[222] Some theories might state that we can make a world worse by improving the lives of immoral agents. This principle concerns the justice of adding utility to these agents' lives rather than the equality of the distribution of utility, but such principles of justice are also inconsistent with Pareto.

equal than the second because in the second distribution there are two extreme utility levels, while in the first distribution there are some agents with utilities between these two extremes.[223]

This conclusion does not follow from Rabinowicz's measure in infinite utility distributions like the ones above, however, because the average pairwise inequality is not well-defined if agents come in no natural order. This reasoning also holds for any measure of inequality that appeals to average utility since this is also not well-defined in most infinite worlds.[224] Even if we attempt to identify inequality by looking at how far each agent's utility is from the midpoint of the accumulation set of a world (intuitively, the world with infinitely many 1's has more agents near the midpoint) we can arrange the agents of either distribution above so that the limit of the average distance from the midpoint is anything in the [0,2] interval.[225]

Many principles of distributive justice are not based on utility distribution alone, however. It is certainly possible for many of these principles to come into conflict with Pareto in infinite cases. Suppose the agents in w_1 have qualitative properties that make them less deserving of utility than the agents in w_2: for example, the agents of w_1 have some important character failing such as being greedy while the agents of w_2 do not. Suppose that $\langle w_2, w_3 \rangle$ is a qualitative duplicate of $\langle w_1, w_2 \rangle$ such that $w_2 \succ w_3$ by Pareto and $w_3 \succ w_4$ by Pareto. We might endorse a desert-based principles which says that even if some agents are better off in w_2 than in w_3 and no agents are better off in w_3 than in w_2 then w_2 is not better than w_3 if the agents that are better off in w_2 do not deserve to be better off than the agents in w_3. Since the agents in w_1 are less deserving of utility than the agents in w_2 because they

[223]Notice that the finite distribution $(2, 2, 2, 0, 0, 0)$ is less equal than the distribution $(2, 2, 1, 1, 0, 0)$ by Rabinowicz's proposed measure of inequality because the average distance between individuals in the first distribution is 1 while the average distance between individuals in the second distribution is $\frac{8}{9}$.

[224]If agents came in a natural order, we could define the average utility as the Cesáro sum of the series. However, these utility levels can be arranged so that its Cesáro sum is anything in the $[0, 2]$ interval, and so this cannot be used to define average utility if agents come in no natural order.

[225]We might be able to formulate distributive principles for worlds with distinct accumulation intervals but – as was noted above – most infinite world pairs will not have distinct accumulation intervals.

are greedy, it follows that when the the agents of w_2 play the qualitative roles of the agents in w_1 they too are less deserving of utility because they are now playing the roles of greedy agents. Therefore w_2 is not better than w_3 by our imaginary desert-based principle.

There are at least two problems for this view. The first problem is that for any qualitative property that we use in a distributive principle, we will be able to find a permutation of the population of $\langle w_1, w_2 \rangle$ such that w_3 is strictly worse than w_2 by both Pareto and our distributive principle. For example, since there are infinitely many agents in w_1 and w_2, it is likely that we can find a permutation of this pair such that each agent in w_2 has at least as much utility in w_3 and at least as much of any other property we use to ground desert. The second problem is this: for most infinite world pairs $\langle w_1, w_2 \rangle$ that contain agents at some upper and lower level of utility, we can find a qualitative duplicate pair $\langle w_3, w_4 \rangle$ such that *infinitely* many agents are better off in w_2 than they are in w_3. Even if we accept principles of distributive justice, we may be disinclined to endorse the claim that world w_3 can be at least as good as world w_2 because of the way in which utility is distributed in w_3 if infinitely many agents are worse off in world w_3 than they are in world w_2.

If we reject Pareto either because we deny that people are 'basic locations of value' or because it is in tension with distributive principles, many theories for ranking infinite worlds become available to us. Rejecting Pareto lets us accept theories that rely on some natural ordering of utility levels in infinite worlds – often an ordering based on the spatiotemporal location of the bearers of these utilities – without rejecting the Transitivity of $\succcurlyeq$, the Permutation Principle, or the Qualitativeness of $\succcurlyeq$. Most order-dependent theories also have the advantage of producing a much more complete ranking of infinite worlds than order-independent theories do. At the beginning of this section, I noted that how attractive we will find rejections of Pareto depends on how attractive we find non-Paretian theories for ranking infinite worlds. Let us therefore turn to consider some of the most plausible alternatives to Pareto.

Many order-dependent theories in infinite ethics, some of which we encountered in Chapters 1 and 2, are in tension with Pareto if we accept the Transitivity of $\succcurlyeq$, the Permutation Principle, and the Qualitativeness of $\succcurlyeq$. Discounting principles rank infinite worlds by applying a spatiotemporal discount rate to utility from some point of origin.[226] UDASSA, which stands for 'Universal Distribution plus the absolute self-selection assumption' is a theory developed by Paul Christiano that gives ethical weight to an agent's experience in accordance with the length of the shortest specification of that experience, including its spatiotemporal location.[227] Ordered Catching-Up, Ordered Overtaking, utility density principles, the hyperreal approach and the surreal approach using limits all depend on there being a natural ordering of agents or subjective experiences. Expansionism ranks infinite worlds based on their distribution of utility across spacetime. All of these theories conflict with Pareto if we accept Transitivity, the Permutation Principle, and the Qualitativeness of $\succcurlyeq$.[228]

We can use the term 'locationism' to describe this collection of theories, since each of them uses the spatiotemporal locations of agents to generate a privileged order of utility that can then be used to generate a ranking of worlds. One of the most promising locationist theories in infinite ethics is the Expansionist view that was discussed in Chapter 2. I will argue that the appeal to the spatiotemporal locations of agents to generate a ranking of worlds in Expansionism is unjustified. This objection to using the spatiotemporal locations of agents

[226] The most natural point of origin is the point of action, which makes discounting views more plausible as accounts of right action than as methods for determining whether one world is at least as good as another.

[227] Paul Christiano [57] argues that the goal of UDASSA is to give a probability distribution over all possible experiences or observer-moments, though we can imagine a version of UDASSA that gives a probability distribution over all possible agents insofar as it is possible to describe whole agents rather than observer-moments. If we believe that persons are not basic locations of value and that the identity of observer-moments is perfectly fragile, then UDASSA may appear to be a plausible alternative to Expansionism in infinite worlds. UDASSA has some disadvantages relative to Expansionism, however. For example, the simplest description of observer-moments relies on some 'naturally specified time zero'. This means that two worlds with identical distributions of subjective experiences across spacetime can be ranked differently if the 'naturally specified time zero' is different in each world (see the case below for an example). This is not an objectionable feature of the view, however, if there exists a naturally specified time zero and it is of ethical relevance.

[228] I have give several examples of permutations that conflict with Expansionism in Chapter 2, but it is not difficult to generate similar examples for the other theories mentioned here.

to generate a ranking of worlds generalizes to the other locationist views mentioned above.

As a reminder, Arntznius [6, p.53] defines an 'allowable expansion' as one that 'at at each time expands at the same rate in each direction in space, and at each location in space expands at the same rate in each direction of time'. Expansionism states that if w_1 and w_2 share the same spacetime with the same metric, and for all allowable expansions in this metric there exists an integer n such that for every $k > n$, the total utility in r_k at w_1 is greater than or equal to the total utility in r_k at w_2 then w_1 is at least as good as w_2. And if there exists an integer n such that for every $k > n$, the total utility in r_k at w_1 is strictly greater than the total utility in r_k at w_2 then w_1 is strictly better than w_2.

Expansionism does not entirely recover completeness, even among worlds have the same spacetime with the same metric. For example, we saw that Expansionism cannot compare the Sphere of Suffering world and the Sphere of Happiness world from Chapter 1 since, as Artnzenius [6, p. 41] notes, 'In this case whether one gets dominance depends on how fast one expands in the temporal and spatial directions. It follows that for any pair of such worlds it is not true that the utility of the one world is greater or smaller than the utility of the other.' This is a result that Arntzenius takes to be intuitively satisfactory, however, and so he does not take this result to be a problem for the Expansionist view.

Even if Expansionism does not entirely recover completeness, it can be used to rank most infinite worlds that are physically plausible. It therefore avoids the sort of ubiquitous incompleteness that characterizes theories that are consistent with Pareto. Expansionism also ranks infinite worlds in a way that we generally find plausible when considering the spatiotemporal distribution of utility. For example, suppose w_1 and w_2 both contain agents that have utility 1 and utility 0 lives. The agents are distributed across spacetime as follows:

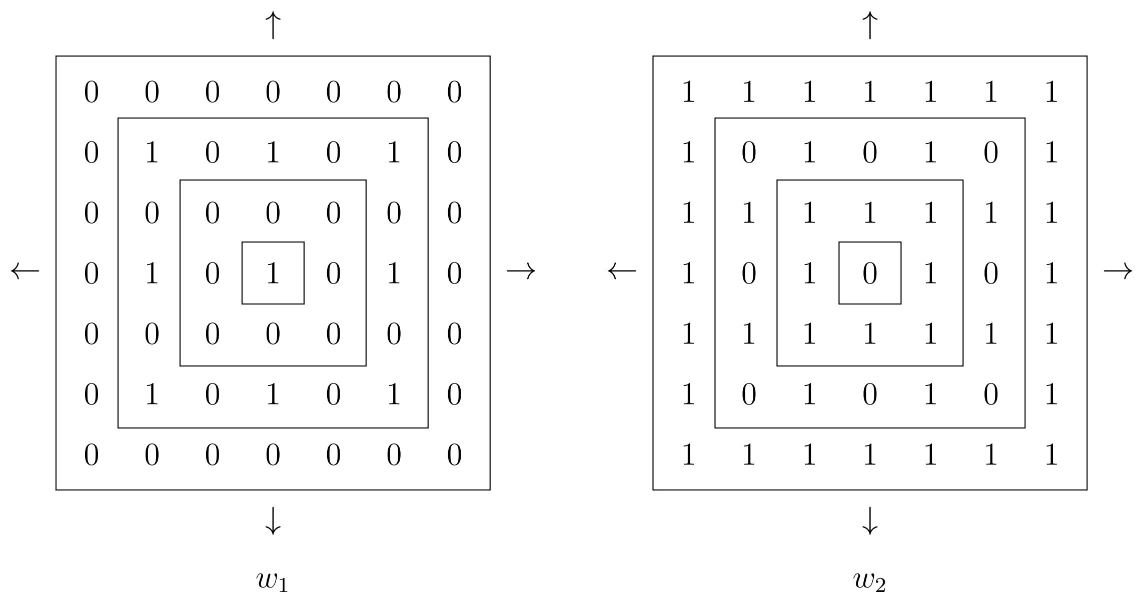


Figure 89: A world pair strictly ranked by Expansionism

The total utility of each region of this allowable expansion in w_1 is $1, 1, 9, 9, 25, 25, ...$ while the total utility of each region of this allowable expansion in w_2 is $0, 9, 16, 40, 56, 96, ...$ There is therefore a region (in this allowable expansion, the second region) such that, for every region after this one, the total utility of the regions of w_2 are strictly greater than the total utility of the regions of w_1. There will be such a region regardless of which point of origin and rate of expansion we select. Expansionism therefore entails that world w_2 is strictly better than world w_1. It seems intuitively plausible that we should prefer worlds in which agents or subjective experiences with positive utility are more densely distributed across spacetime.

My main concern for Expansionism is that the intuition that we should prefer a world with a more dense distribution of utility across spacetime may arise from a misapplication of intuitions about finite worlds to infinite worlds. One reason that we might be inclined to say that world w_2 is better than world w_1, for example, is that if we had to gamble on being an agent in w_1 or an agent in w_2, we would prefer to be an agent in w_2. This likely arises out of an intuition that we have a greater chance of being an agent with a utility 1 life in w_2 than in w_1. But we are only justified in believing that we have a greater chance of being an agent with a utility 1 life in w_2 than in w_1 if we are justified in using the spatiotemporal ordering of agents to assign probabilities to the claim that we will have a utility 1 life in w_1 or in w_2 from behind the veil of ignorance.[229] If we are justified in using the spatiotemporal ordering of agents for this purpose, we can assign probabilities to being in a given infinite subset of the population of w_1 and w_2 in accordance with their densities under this ordering.

I believe we should not be so quick to think that we are justified in using the spatiotemporal ordering of agents to assign a greater probability to being a utility 1 agent in w_2 than in w_1. To show this, suppose, first, that w_1 and w_2 contain identical populations and that you know from behind the veil of ignorance that you are in the population of w_1 and w_2 but you do not know whether you will have a utility 1 or a utility 0 life in these worlds. Suppose you are told that all of the agents that exist at utility 1 in world w_1 are in set X_1 that all the agents that exist at utility 0 in w_1 are in two infinite sets X_2 and X_3. And the agents in X_1 and in X_2 both have utility 1 lives in w_1 and the remaining agents in X_3 have utility 0 lives in w_1. This means that if you are an agent in X_1 or X_3 then your utility will be the same in both w_1 and w_2. But if you are an agent in X_2 then you will have a utility 1 life in w_1 and a utility 0 life in w_2. If you use the spatiotemporal ordering of agents to assign probabilities to being a utility 1 agent in w_1 and w_2, however, then you will assign a greater probability to being a utility 1 agent in w_1 than in w_2 even though, without knowing more about the identities of agents in w_1 and w_2, you have no clear reason to think that the distribution of utilities described above is less likely than a distribution that would cause you to prefer w_2.

What if you know that the populations of w_1 and w_2 are not identical?[230] When you consider

[229] We might think that spatiotemporal ordering is a natural ordering of agents but deny that it has this epistemological significance.

[230] If you do not know whether the population is identical or overlapping or disjoint from behind the veil of ignorance, then the reasoning I give in both the identical population case and the variable population case should cause you to doubt your intuition that you have a greater chance of a utility 1 life in w_1 than in w_2.

whether you prefer one world or another world from behind the veil of ignorance and the populations of the two worlds are non-identical, it is important that you do not assume that our existence is guaranteed in either world.[231] If you were to assume that our existence is guaranteed then you would prefer a world that contains a single agent with a utility 2 life over either w_1 or w_2 as depicted above. Yet such a world is clearly worse than w_1 and w_2.[232]

Suppose that you do not know whether you will exist in w_1 or w_2 but you do know that these worlds have non-identical populations. You are told that the agents that exist in world w_2 are the same as the agents that exist in world w_1 at utility level 1. Let X be this set of shared agents between w_1 and w_2 and let Y be the remaining agents that exist at utility 0 in w_1 and do not exist at all in w_2. It seems that, insofar as you believe that you could be any possible person, you should think that you have a greater chance of a utility 1 life in w_1 than in w_2. After all, if you are an agent not in Y then you won't exist in w_1 or w_2. If you are an agent in Y then you will have a utility 0 life in w_1 and you won't exist in w_2. If you are an agent in X then you are guaranteed to have a utility 1 life in w_1 but you may have a utility 1 or utility 0 life in w_2. If you use the spatiotemporal ordering of agents to assign probabilities to being a utility 1 agent in w_1 and w_2, however, then you will assign a greater probability to being a utility 1 agent in w_1 than in w_2 in this case. This seems wrong.

I believe that these scenarios give us reasons to doubt our intuition that we have a greater chance of being an agent with a utility 1 life in w_2 than in w_1. Whether we have a greater chance of being a utility 1 agent in w_1 than in w_2 depends on the identities of the utility 1 and utility 0 agents in w_1 and w_2. And we cannot infer the likelihood that the identities of

[231] The veil of ignorance was formalized by Harsanyi [105] [106] to analyze social policy from the perspective of one who does not know which position in society one will occupy. If we are assessing the value of different possible populations of agents, however, it seems important that one also does not know whether one will exist in the population in question for the reasons that I give below.

[232] If we assume that we will exist in the populations in question, then it is likely that we will favor something like average utilitarianism. We will care about whether a world produces greater expected utility for those who exist and will be indifferent to the number of people that exist in a given population.

the agents in w_1 and w_2 will be such that we have a greater chance of being a utility 1 agent in w_1 or in w_2 from the spatiotemporal distribution of those agents.[233]

The 'greater chance of a better life' reason for favoring Expansionism and my response to it may not be compelling to those who believe that subjective experiences rather than whole lives are what matter in ethics, since this reasoning seems to be premised on the idea that improving lives is of ethical importance. However, Expansionism produces highly counterintuitive results even if we adopt the subjective experiences view. According to Expansionism, a world with more densely packed agents with low utility levels can be better than a world with less densely packed agents at higher utility levels. We can also make worlds worse by having each positive subjective experience be moved a little further apart in spacetime. It is difficult to see why we should think that rearranging the subjective experiences of the same utility level can make a world better or worse. This problem arises regardless of whether we take subjective experiences or agents to be of importance.

We might think that we can get around this kind of problem by replacing 'total utility' with 'average utility' in our formulation of Expansionism. If we adopt this average utility variant of Expansionism, however, we can still improve a world by moving agents with higher utility closer together. For example, imagine an infinite world composed of infinitely many uniform squares spread out like a checkerboard. On each square there lives two agents: one with a utility 2 life and one with a utility 1 life. Now suppose we move each of the utility 1 agents so that they are further apart so that there is one utility 1 agent for every four squares, while keeping the utility 2 agents in their original locations. For all allowable expansions there will

[233] This claim seems to be in tension with the 'Limiting Proportion' principle formulated by Dorr and Arntzenius [71, p. 413]. To give a rough summary, this principle says that if H entails that every finite region of the world contains finitely-many finitely-lived agents and agents can move only a finite distance then, if the limit of the proportion of agents that have a certain property G is n across all allowable expansions, you are justified in having credence n that you have property G. The tension between my argument and this claim is somewhat indirect, since Limiting Proportion is a principle about what credence we should have conditional on hypothesis H rather than being on conditional on finding ourselves in a given world. If my argument is correct, however, then we may be able to construct analogous worries for this principle.

exist a region such that the average utility at every subsequent region is greater in the world in which we have moved the utility 1 agents than in the original world. Therefore the second world is better than the first according to the average utility variant of Expansionism. Again, it seems implausible that we can make a world better simply by rearranging the subjective experiences of the same utility level across spacetime.

Expansionists might respond by arguing that spatiotemporal regions are 'basic locations of value', rather than agents or subjective experiences. If this is correct then we should care about how much utility there is at each spatiotemporal region rather than at each agent or at each subjective experience or in each agent's life. The main consideration against this claim is that in cases where increasing utility at each spatiotemporal regions decreases the amount of utility per agent or per subjective experience, the claim that the spatiotemporal distribution of utility matters does not seem compelling.

Arntzenius [6, p. 55] states that a reason for dealing with locations rather than agents is that 'the structure of locations can provide us with a natural ordering which allow us to apply dominance reasoning to infinite worlds'. The key concern I want to express for Expansionism or any other theory in infinite ethics that appeals to the spatiotemporal locations of subjective experiences is that there doesn't appear to be any reason to think that where positive and negative subjective experiences are located in spacetime is of any ethical relevance. In finite worlds, ethical theories that are sensitive to increases in utility but reject spatiotemporal discounting tend to be indifferent to when and where positive and negative utility experiences occur in spacetime. It is difficult to see why this is a feature of that we should start to consider morally relevant if worlds happen to be spatiotemporally infinite.[234]

[234]Here is a possible response to this argument that I do not find convincing. Suppose that in finite worlds it is better for there to exist a series of utility 2 subjective experiences every three seconds for one minute rather than a series of utility 1 subjective experiences every 1 second for 20 seconds. The former sequence has fewer utils per second but more total utils (40 vs. 10). If such sequences lasted an equally long time, however, then it will always better to have a sequence with more utils per second: one minute of the second sequence is better than one minute of the first sequence. If we apply this same reasoning to worlds that are

Expansionism is just one example of an intuitively plausible theory that generates a far more complete ordering of worlds than is possible if we endorse Pareto. Most theories of this sort rely either on a natural ordering of agents or of subjective experiences – generally one that is based on their location in spacetime – or on the claim that 'basic locations of value' come in a natural order – perhaps because spatiotemporal regions play this role. The principal objection to both kinds of theory is that there is no clear reason to think that spatiotemporal orderings are of ethical importance, even if they happen to make it easier to rank infinite world pairs. There is no use in producing a complete ordering of infinite world pairs unless that ordering is based on something that is of ethical importance.

In Chapters 1 and 2 I responded to objections to Pareto that were premised on the claim that $\succcurlyeq$ is a qualitative internal relation. In this section I have considered two alternative classes of objection to Pareto. I have argued that if we reject the primacy of people in infinite ethics then we face even greater incomparability unless we can find a privileged ordering of 'basic locations of value'. I then argued that spatiotemporal locations – the most plausible candidate to ground a privileged ordering of agents or subjective experiences – do not seem to have the ethical significance required to provide such a privileged ordering. It may be tempting to appeal to such orderings because they can be used to produce rankings of infinite worlds that conform to our prima facie intuitions about cases. If spatiotemporal orderings are not ethically relevant, however, then I believe that these intuitions may fade upon greater reflection. Finally, I have noted that although some principles of distributive justice are inconsistent with Pareto in finite worlds, it is difficult to formulate similar objections when comparing worlds with infinitely many agents at some upper and lower level of utility.

---

temporally infinite, surely we can conclude it is better to have a utility 1 experience every 1 second for an infinitely long period of time than it is to have a utility 2 experience every three seconds for an infinitely long period of time. But this depends entirely on what we think is grounding the claim that having more utility per second is better if the lengths of time are equal. If we believe that this is true in the finite case because such sequences produce more total utility, then we are not justified in concluding that if two temporal sequences are infinite then we ought to prefer the sequence with more utility per second.

When faced with the impossibility result I have formulated above, rejecting Pareto may ultimately be considered the least bad option out of a set of bad options. My goal has been to explore this option and show that rejecting this axiom is far from cost-free. If we are not willing to accept these costs then we must either reject one of the first three axioms or we must accept that most infinite worlds are incomparable. In the next section I will consider the consequences of embracing ubiquitous incomparability.

5.4 Embracing Incomparability

If we accept the first four axioms listed at the beginning of this chapter – Transitivity, the Permutation Principle, the Qualitativeness of $\succcurlyeq$, and the Pareto principle – then we must conclude that all infinite world pairs that can be shown to be incomparable by a four world or cyclic argument are, in fact, incomparable. Incomparability is not the same as mere ignorance about whether one of the worlds is better than another: if the two worlds are genuinely incomparable then neither world is better than or as good as the other. Even if we knew all true facts, we would still conclude that these two worlds are incomparable. If we accept the first four axioms we must therefore reject the Minimal Completeness axiom, which says that comparability between infinite worlds is not incredibly rare.

Rejecting completeness might not seem like such a cost if it were only a small fraction of world pairs that were expected to be incomparable. Perhaps ethics can function adequately if most worlds are incomparable but a small subset of worlds are incomparable. If we accept the four axioms above, however, then ethical incomparability between infinite worlds is not uncommon: it is ubiquitous. If worlds contain infinitely many agents and those agents can have any finite level of utility, then almost all infinite worlds will contain infinitely many agents at each possible utility level. This means that the fraction of world pairs such that there is *not* a bidirectional upgrade from the population of w_1 to the population of w_2 and from the population of w_2 to the population of w_1 will be very small. Therefore, by Result 8 alone, very many infinite world pairs will be incomparable.

Since it has been shown that the five axioms listed in section 5.1 are not jointly satisfiable, one compelling reason to accept ubiquitous incomparability between infinite worlds is simply that this is the least bad option available. This is not the only reason to accept the incomparability results, however. The claim that almost all infinite world pairs are incomparable may, on reflection, begin to seem somewhat intuitive as we reflect more on the cases in question. Suppose that a single world contains infinitely many agents at utility 1 and infinitely many agents at utility 0. Would switching the utilities of the agents of this world result in a world that is better than, equal to, or worse than the original? We may initially be inclined to think that the resulting world must be just as good as the first world, perhaps because the cardinality of utility 1 and utility 0 agents remains the same after they are switched. However, once we notice that the cardinality of utility 1 and utility 0 agents would remain the same even if we were to increase the utility of infinitely many agents at utility 0 while also leaving infinitely many agents at utility 0 (resulting in a better world by Pareto) or if we were to decrease the utility of infinitely many agents at utility 1 while also leaving infinitely many agents at utility 1 (resulting in a worse world by Pareto), we may divorce the concept 'is as good as' from the concept 'has the same cardinality of agents at each utility level'.[235]

In the case above we could appeal to some other qualitative properties of to rank the original world and the world in which its agents' utilities have been inverted. We could, for example, use facts about the spatiotemporal locations of utility 1 and utility 0 agents at each world to rank them. But, as I have argued, it seems implausible that the ethical ranking of worlds

[235]Suppose that it is not the cardinality of utility 1 and utility 0 agents that generates our equality intuition but the measure or ratio of these agents. This will not vindicate the intuition since the measure of utility 1 and utility 0 agents is not well-defined, and the ratio of utility 1 to utility 0 agents is an indeterminate form.

depends on qualitative properties that have nothing to do with improvements in agents' lives, such as the order in which the agents are born at each world or their spatiotemporal locations. If the ethical better than relation does not supervene on such qualitative properties or on the cardinality of agents' utility levels in infinite worlds, it does not seem so implausible conclude that in this case the original world is not better than, worse than, or equal to the world in which the utilities of its agents have been inverted.

A principal concern for views that accept ubiquitous incomparability between worlds is the affect that this will have on what people ought to do. So far I have focused almost entirely on the objective rankings of worlds: on whether world w_1 is in fact better than world w_2. But the primary concern of ethics is arguably not whether a given world is better than or worse than another, but whether a given action is permissible, obligatory, or forbidden.

One might think that incomparability between infinite worlds will only generate problems for permissibility and obligation only for consequentialist moral theories. This is because, for consequentialists, there is a strong connection between the permissibility of an action and the and the world that results from that action. This is not true of non-consequentialist moral theories, however. Most deontological theories are premised on much weaker connections between the permissibility of an action and the world that results from that action. Deontologists argue that we can be permitted or obligated to perform actions even if these actions do not produce the best consequences.[236] For example, doctors have a duty to avoid killing their patients even if, by doing so, they could save the lives of many other patients.[237] Many deontologists do take into account the consequences of action to some degree, but the

[236] Scheffler [199] identifies two different kinds of agent-centered non-consequentialist moral theories. The first are 'hybrid' theories, which 'depart from consequentialism to the extent of incorporating an agent-centred prerogative, and which thus hold that one is always permitted but not always required to do what would have the best available outcome overall' [199, p. 116]. The other are 'fully agent-centred' theories, which 'incorporate both an agent-centred prerogative and agent-centred restrictions, and which thus hold that one is neither required nor permitted always to do what would have the best outcome' [199, p. 116].

[237] See Foot [88] for a discussion of this well-known case.

consequences of an action do not fully determine its permissibility.[238]

We might think that any problems of permissibility that arise out of the incomparability of worlds cannot affect these deontological moral theories the link between worlds and the permissibility of actions is much weaker on these theories. In the next two sections I will show that this is not the case. I show that we can use the incomparability of infinite worlds to generate new puzzles for permissibility that highly general. These puzzles apply to consequentialist theories and to non-consequentialist theories – such as deontological and virtue ethical moral theories – that are premised on a much weaker connection between the outcomes of actions and their permissibility. The puzzles that I present in these sections do not depend on the assumption of a consequentialist ethical framework.

In ethics it is typical to distinguish between what agents objectively ought to do and what they subjectively ought to do. Objective oughts are not sensitive to the information that the agent has, while subjective oughts are sensitive to the information available to the agent.[239] What an agent ought to do in a given scenario therefore generally coincides with what she subjectively ought to do in that scenario if she had perfect information.[240] In this section I am going to offer a separate treatment of objective and subjective permissibility. I will show that the incomparability results generate a puzzle for theories of objective permissibility, before

[238]Kagan [118, p. 79] characterizes 'moderate deontologists' as those who 'believe that the constraint [against harming] has a threshold: up to a certain point - the threshold point - it is forbidden to kill or harm an innocent person, even if a greater good could be achieved by doing it; but if enough good is at stake - if the threshold has been reached or passed - then the constraint is no longer in force, and it is permissible to harm the person.' In *The Right and the Good*, Ross [192] formulates an influential form of moderate deontology.

[239]Here I assume that the subjective ought is sensitive to an agent's empirical uncertainty only, but it is possible that the subjective ought should also be sensitive to an agent's moral uncertainty (see Bykvist [47]) or to the agent's moral imperfections and the options available to her (see Hedden [107]).

[240]As Yetter Chappell [246] notes, we might be worried that identifying the objective ought with the subjective ought under conditions of perfect information will involve committing Schope's [206] 'conditional fallacy'. For example, it will never be the case that an agent objectively ought to gather more information because if she had perfect information then information gathering would never be required. I am sympathetic to the view that information gathering is only ever something that an agent subjectively ought to do, but we need not take a stand on this issue here since cases like this will not arise in the discussion below.

showing that we can generate an analogous puzzle for theories of subjective permissibility.

The case most often used to demonstrate how objective and subjective oughts can diverge is the 'miners puzzle'.[241] To give a variant of this puzzle: suppose that ten miners are trapped in either shaft A or shaft B and you do not know which. You can press a button to secure shaft A or you can press a button to secure shaft B but not both. If you secure the correct shaft then all of the miners will live but if you secure the wrong shaft then all of the miners will die. Alternatively, you can press a button that will partially secure both shafts. If you press this button then it is certain that one and only one miner will die. If you had full information about this case then you would know where the miners are located and would press the button corresponding to their location. You would certainly not press the button that partially secures both shafts. Given your imperfect information about the miners' location, however, it would seem reckless of you to gamble and press one of the first two buttons rather than pressing the button that will partially secure both shafts. So what you objectively ought to do and what you subjectively ought to do seem to diverge.

In the next section I will show that rejecting Minimal Completeness generates problems for theories of objective permissibility. In the final section I will show that rejecting Minimal Completeness generates similar problems for theories of subjective permissibility.

5.4.1 Incomparability and Objective Permissibility

In this section I will address the question of what agents are *objectively* permitted to do if some of the actions available to them will result in infinite worlds that are incomparable.[242] It will be helpful to proceed in one of two ways. First, we could begin by assuming that all actions are objectively permissible if some of the actions available to the agent will result in

[241]This case can be found in Parfit [180] who attributes the case to Regan [190].

[242]The outcome of an action is the entire world that will result if the action is performed, including both its immediate and long-term consequences. Therefore the outcomes of actions can be infinite and incomparable.

infinite worlds that are incomparable. We could then search for principles that restrict the set of permissible actions. Alternatively, we could begin by assuming that all actions are objectively impermissible if some of the actions available to the agent will result in infinite worlds that are incomparable. We could then search for principles that broaden the set of permissible actions. I will begin by assuming that all actions objectively permissible and we that must search for search for principles that restrict the set of permissible actions. After this, I will briefly address what happens if we adopt the opposite view.

In this section I will talk about an agent's 'choice set'. This is the set of all acts available to the agent in a given scenario. Following Savage [198, p. 13-15] I will assume that actions are functions from states of the world to outcomes. The outcome of an action is the entire world that will result if the action is performed, including both its immediate and long-term consequences.[243] We can therefore treat outcomes as worlds that result from an act being performed in a particular state of the world.

According to utilitarians, the objective permissibility of actions depends entirely on the total utility that the world will have if the action is performed. In finite worlds, an act a is objectively permissible according to utilitarianism if and only if there is no other available act b such that the total utility that the world will have if b is performed is strictly greater than the total utility that the world will have if a is performed. In this section I will formulate a general puzzle for permissibility that does not assume consequentialism. Let us start by exploring the problems that incomparable infinite worlds generate for consequentialist theories, however, while noting that these theories are controversial and will be rejected by many ethicist. In the interest of exploring the difficulties that consequentialist theories of objective permissibility faces if some of the outcomes of our actions are incomparable, let us

[243]If the world is infinite and someone saves the life of a stranger and that stranger goes on to have living descendants, for example, the outcome of this action could be a future in which infinitely many different people exist than those thats would have existed otherwise.

assume for the time being that the objective permissibility depends entirely on the utility levels of the people that will exist if we undertake a given action.

Those who adopt a consequentialist approach to objective permissibility cannot simply appeal to the utilitarian criterion of objective permissibility in infinite worlds because we do not yet have a method for assigning real-valued utilities to worlds that contain infinite utility. If one world contains infinitely many agents at utility level 2 and another world contains the same agents at utility level 1, the total utility of both worlds is positively infinite. If we were to infer from this that both worlds have the same total utility then the utilitarian criterion would entail that if act a results in the first world and act b results in the second world, then act a and b are both permissible. Since the outcome of a is better than the outcome of b by Pareto, this means that if we extended the utilitarian criterion to infinite worlds in such a way that it treated 'positively infinite' worlds as the same, it would conflict with Pareto. I assume that most utilitarians would not wish to extend their theory in this way.

In order to identify what actions are objectively permissible or impermissible for outcomes-based moral theories like consequentialism, we can first identify weak restrictions on what an agent objectively permitted to do if the outcomes of some of her actions are infinite before assessing whether more robust restrictions are possible. I believe that most consequentialists will be inclined to accept the following restriction on what an agent is objectively permitted to do, based on the Weak People Criterion outlined in the previous chapter:

Weak People Criterion for Objective Permissibility (WPCO)

It is objectively permissible for an agent to perform action a only if there is no action b available to the agent such that the outcome of action b is strictly better than the outcome of action a by the Weak People Criterion.

In finite worlds, WPCO entails the maximizing utilitarian principle 'an act a is objectively permissible if and only if there is no action b available to the agent that produces more utility than action a does'. This should at least make it somewhat attractive to those who are inclined to accept utilitarian principles in finite cases. In infinite worlds, however, WPCO places fewer restrictions on what an agent is objectively permitted to do.

Suppose that an agent can choose whether to cure a small population of people (action a) or to leave them sick (action b). If the agent cures the people then they will be happier: they will each experience utility 1 instead of utility 0. Some of the cured people will also conceive children other than those they would have conceived if they had been left sick, and this will completely change who exists in the future. We can represent the utilities of the future people in this world as follows, where S is the actual state of the world and X is the small population of people that the agent must decide whether to cure:

	S				
	X 50	Y_1 ∞	Y_2 ∞	Z_1 ∞	Z_2 ∞
a	1	1	0	–	–
b	0	–	–	1	0

Figure 90: Curing a small population or leaving them sick

In this case we are inclined to say that it is objectively permissible (or even obligatory) to cure the small population and objectively impermissible to fail to do so. But the WPCO principle does not entail that a or b is objectively impermissible.

The WPCO does place some restrictions on what an agent is objectively permitted to do. Suppose, for example, that the agent in the case above has a third option (action c) that involves curing the small population but passing on the cost of this to their first 50 happy descendants (50 agents in Y_1) reducing their utility from 1 to 0:

	S					
	X	Y_{1a}	Y_{1b}	Y_2	Z_1	Z_2
	50	50	∞	∞	∞	∞
a	1	1	1	0	–	–
b	0	–	–	–	1	0
c	1	0	1	0	–	–

Figure 91: Curing the small population but passing on the cost

Let's use w_a to represent the world that results if act a is performed. By WPCO the outcomes of a, b, and c in the case above form the following partial order:

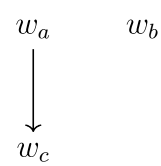


Figure 92: A partial order of outcomes under WPCO

If we accept WPCO then act a and act b are both objectively permissible but act c is not. More generally, an act will be objectively permissible according to WPCO if and only if there is nothing higher than its outcome in the partial order of outcomes generated by WPCO.

The main problem with WPCO, as noted above, is that is that if the world is infinite then the WPCO principle will rarely entail that an action is impermissible. The only actions that are impermissible will be those with outcomes that are strictly worse than the outcomes of another action by the WPCO principle. But if the world is infinite then we can expect this to make up a small fraction of the actions available to an agent, since we have already argued that only a small fraction of infinite worlds are comparable by the Weak People Criterion.

It might be objected that WPCO would place adequate restrictions on which actions are objectively permissible if the causal ramifications of our actions are finite since, if this is the case, the outcomes of all of our actions can be totally ordered by the Weak People Criterion.

An act a is impermissible by WPCO only if the sum of the difference of utility between the outcome of a and the outcome of an alternative action b converges absolutely and is positive, or diverges to positive infinity. But if the causal ramifications of our actions are finite then the sum of the difference of utility between any two actions will always converge absolutely. It will consist of a finite sequence of positive and negative differences in utility while at least one of the actions have causal ramifications. Since the part of the world that is beyond our causal influence will be the same regardless of how we act, this finite sequence will be followed by an infinite sequence of zeros after the point at which neither action has any further causal ramifications. Therefore, if the causal ramifications of our actions are finite, the outcomes of all actions can be totally ordered using the Weak People Criterion. However, there is no guarantee that the causal ramifications of our actions are finite. Arntzenius [6, p. 52] points out that 'in Newtonian worlds, and in relativistic worlds with an infinite future, on the most natural understanding of what our causal sphere of influence is, it is infinite'. If the world is in fact finite or the causal ramifications of our actions are in fact finite then WPCO can be used to identify a small set of objectively permissible actions. If, however, the causal ramifications our actions are infinite, then WPCO will generally entail that only a very small fraction of the actions available to an agent are objectively impermissible.

The inverse problem arises if we assume that all actions are impermissible by default and then use an analogue of the WPCO principle to broaden the set of actions that are objectively permissible. The most natural analogue to use would be the following: it is objectively permissible for an agent to perform an action a if, for every action b available to the agent, the outcome of a is at least as good as the outcome of b by the Weak People Criterion plus Transitivity. But if the world is infinite then in most situations all no action available to an agent will be objectively permissible by this principle. A single pair of incomparable act outcomes would be sufficient to prevent any act from being permissible by this principle.

If consequentialists cannot appeal to something stronger than WPCO-like principles to deem an action objectively impermissible or permissible, they will face serious difficulties in infinite worlds. I have shown that if we assume that actions are objectively permissible by default, WPCO fails to adequately restrict the set of permissible actions. And if we assume that actions are objectively impermissible by default, its analogue fails to adequately broaden the set of actions that are permissible. Consequentialists must therefore look for principles of objective permissibility and impermissibility that are stronger than WPCO.

One option available to the consequentialist is to try to appeal to more restrictive principles than WPCO. For example, the consequentialist could claim that if act a produces more spatiotemporally discounted utility than act b, then act b is impermissible. This rule depends not only on the distribution of utility but on qualitative properties of the outcomes of actions, but it is still in the spirit of consequentialism. Consequentialists could also try to find rules that depend solely on the distribution of utility but that are more restrictive than WPCO.[244] The key point is that principles as weak as WPCO do not produce plausible accounts of objective permissibility if we accept that many infinite worlds are incomparable.

Thus far I have been discussing consequentialist principles of objective permissibility. At this point I want to turn to principles are not premised on the idea that the objective permissibility of actions depends solely on the utility levels of the people that will exist if we undertake a those actions. Many theories within ethics, such as deontological theories and virtue-based theories, do not assume that there is such a direct connection between the objective permissibility of actions and the utility levels of the people that will exist if we undertake a given action. These theories can base the permissibility of curing the small population in the case above on features of the action itself rather than its outcome: for

[244] For example, they could claim that if act a produces an outcome with a higher utility midpoint than act b, then act b is impermissible. One problem is that these purely utility-based rules will often not entail that it is impermissible to fail to cure the small population in the case above. According to the midpoint rule, neither curing the population nor failing to cure the population are objectively impermissible.

example, they can state that one has a duty to alleviate the suffering of nearby agents, or that acting with the intention of alleviating the suffering of nearby agents is virtuous.

I believe that many non-consequentialist principles can be used to vindicate this judgment. Kant [129, 4: 398], for example, argues that we have a duty to be beneficent when we can be. Curing the small population of agents clearly seems to constitute a beneficent act, since it is done in order to help those who are sick. Aristotle [111, 1120a33-b3] argues that those who are virtuous will use their property to help others.[245] Insofar as this indicates that donating one's money to cure the small population would be virtuous, it surely indicates that curing the small population if one can do so at no cost to oneself would also be virtuous.

Suppose we are not convinced that non-consequentialist ethicists will claim that it is impermissible to fail to cure the small population of sick agents in the case above. We can simply replace this act pair with any other act pair whose outcomes result in the same utility levels as the act above. For example, suppose that instead of choosing between curing the small population of agents and failing to do so, the agent is choosing between allowing a small population of agents to live happy lives (replacing act a) or depriving them of their liberty and torturing them so that they live miserable lives (replacing act b). The outcomes of these actions in terms of their utility may be identical to those in the case above, but many non-consequentialist theories will state that it is not permissible to deprive the small population of their liberty and torture them. If the reader is not convinced that non-consequentialist moral principles entail that it is impermissible to fail to cure a sick population of agents at no cost to oneself, they may substitute each instance of this act in the discussion that follows with the act of torturing an otherwise happy population of agents for no reason.

Suppose we endorse a non-consequentialist principle like 'if you can easily alleviate the suffer-

[245] Aristotle believes that it worse to fail to help those with whom we have special relationships than it is to fail to help strangers [111, 1160a4-1166] but this is not relevant to the case considered here.

ing of nearby agents at little or no cost to yourself, you are permitted to do so and forbidden from failing to do so'. This principle would entail that failing to cure the sickness in the case above is objectively impermissible. Perhaps we can use principles like this – principles that appeal to properties of actions and not just properties of outcomes – to determine whether actions are permissible or impermissible when their outcomes are incomparable. Rather than assessing such principles one by one, I will show that we can construct a new puzzle for *any* principle that entails curing the small population is objectively permissible and that failing to do so is objectively impermissible.[246] In order to construct this general puzzle, I will first assume that even non-consequentialist ethical theories should endorse the following very weak link between objective permissibility and the outcomes of actions:

Infinite Act Pareto

If the outcome of act a and the outcome of act b contain the same people and infinitely many people are worse off by at least some finite amount if act b is performed than if act a is performed and no people are better off if act b is performed than if act a is performed, then (i) act b is objectively impermissible and (ii) if a and b are the only actions available to the agent, then act a is objectively permissible

Infinite Act Pareto simply says that an action cannot be permissible if it results in a world that is worse for infinitely many people and better for no people. It also says that if our only options are to perform an action that results in a world that is worse for infinitely many people or to perform an action that does not, it is permissible to perform the action that does not. I believe that a violation of Infinite Act Pareto would be considered unacceptable by most ethical theorists. Even if we believe that we have a general duty to avoid lying, we are surely not permitted to avoid lying if, in doing so, infinitely many people will be worse off

[246] This includes both consequentialist principles that go beyond WPCO and non-consequentialist principles that appeal to properties like intentions, rights, duties, and so on. It also applies to any principle that says it is impermissible to torture an otherwise happy population of agents for no reason in the case given above.

than they would have been and no one will be better off than they would have been. Infinite Act Pareto says that the same is true not only of lying, but of any property of actions that could ground facts about objective permissibility and impermissibility.

This link between the permissibility of actions and the outcomes of actions is extremely weak and will hopefully appear plausible to those who reject stronger principles linking the objective permissibility of actions to the outcomes of those actions. It is a principle that will surely be endorsed by moderate deontologists who believe that we are permitted to violate deontological constraints when the stakes are high enough.[247] This will not be true of all deontologists, however. Kant believed that many duties are absolute. A stark example of this is given in *On a Supposed Right to Lie because of Philanthropic Concerns* [124] in which Kant argues that if a man bent on murder comes to our door asking for the location of a potential victim, it is our duty to tell the truth to them.[248] Williams [210, p. 98-100] indicates that it may be better to let twenty people die at the hands of others than to kill one person oneself. Since some non-consequentialists are willing to commit to the view that certain actions are wrong even if, by performing them, we can prevent large finite amounts of harm from occurring, they may also be willing to commit to the view that those actions are to wrong even if, by performing them, we can prevent an infinite amount of harm from occurring. Non-consequentialists that adopt this view will reject Infinite Act Pareto.

I do not claim that all non-consequentialists will be committed to Infinite Act Pareto. I do claim that if non-consequentialists deny Infinite Act Pareto then this will, in the eyes of many, constitute a powerful objection to those non-consequentialist theories. Denying similar finite principles strikes us as implausible enough. Even if I have a duty not to lie

[247] Nussbaum [83, p. 101] discusses similar views under the banner of 'sensible deontologies'. For discussion of the possibility moderate deontology see Smilansky [212] and Cook [61].

[248] Korsgaard argues that lying to the murderer can be permissible within a Kantian framework. [128, p. 339] She argues that, on her modification of Kant's views, aways telling the truth is an ideal to live up to but it is not always feasible to live up to this ideal when doing so would involve great evil. [128, p. 349]

or cheat or kill, there is surely some point at which, if the harm of my carrying out these duties is great enough, I am no longer permitted to do so. Suppose that I am forced to choose between the following by some evil mastermind: either I kill a single person or the mastermind will torture every person on earth for a decade before killing them (this includes the person I am being asked to kill). To avoid killing is one of the strongest duties we have, and yet in such a case as this it seems highly immoral to privilege satisfying this duty over protecting the lives of those who will be tortured and killed. And yet this case only involves finite harms. How much stronger the case against the absolutist seems to be when there were infinitely many people who would be tortured and killed in this scenario.

Given this, I believe that most non-consequentialists will accept Infinite Act Pareto and that a failure to do so constitutes a powerful objection to a moral theory. I will argue that we should also accept three further principles about objective permissibility:

Qualitativeness of Objective Permissibility

If act pair $\langle c, d\rangle$ is a qualitative duplicate of act pair $\langle a, b\rangle$, then d is objectively impermissible if and only if b is objectively impermissible

Permutation Principle over Act Outcomes

For any act pair $\langle a, b\rangle$ and any bijection g from the population of the outcomes of a and b onto any population, there exists an act pair $\langle c, d\rangle$ that is a qualitative duplicate of $\langle a, b\rangle$ but whose outcome population has been permuted by g

Independence of Irrelevant Alternatives (Objective)

If b is objectively impermissible if an agent's choice set is $\{a, b\}$ then b is objectively impermissible under any expansion of these options

The Qualitativeness of Objective Permissibility just says that objective permissibility should not vary across acts that are qualitative duplicates. The properties of actions or the relations between actions that ground their objective permissibility and impermissibility must be qualitative properties and relations. Although theories of objective permissibility may depend on the relationships between agents, for example, they surely should not depend on which particular agents exist in a given world. Among other things, theories of objective permissibility that appeal to haecceitistic facts seem unacceptably inequitable. I believe that many of the reasons we presented for accepting the Qualitativeness of $\succcurlyeq$ give us reasons to accept this analogous principle for objective permissibility.[249]

The Permutation Principle over Act Outcomes claims that we can find a qualitative duplicate of an act pair such that the population affected by those actions has been permuted. Suppose that act a involves pulling a lever to divert a train in order to save the lives of Alice and Bob and an act b involves failing to do so, the Permutation Principle over Act Outcomes says we can find some possible pair of actions c and d that are qualitatively identical to a and b. In this case, act c involves pulling a lever in exactly the same way and in exactly the same circumstances as in act a, while act d involves failing to do so.[250] Act c and d have qualitatively identical outcomes to acts a and b respectively, but the two people affected by these actions have been permuted. For example, if f(Alice)=Carl and f(Bob)=Dave, then in the outcome of acts c and d Carl plays the qualitative role that Alice plays in the outcome of acts a and b and in the outcome of acts c and d Dave plays the qualitative role that Bob plays in the outcome of acts a and b. As with the original Permutation Principle, the act pair that is a qualitative duplicate under a permutation in the Permutation Principle over Act Outcomes could be logically possible but not metaphysically possible. As with the previous principle, I believe that many of the reasons we presented for accepting Permutation

[249] I also believe that we should accept the analogous principle to the Ethically Relevant Qualitativeness of $\succcurlyeq$ for objective permissibility, but I do not utilize such a principle here.

[250] Since c and d are qualitative duplicates of a and b, any morally important properties that the pair a and b has, the pair c and d also has. For example, if act a involves rights violations while act b does not then act c involves rights violations while act d does not.

Principle give us reasons to accept this analogous principle for objective permissibility.

Finally, the Independence of Irrelevant Alternatives is a widely accepted principle which says that if an act b is impermissible when we can only perform a or b, then adding a further option c cannot render b permissible.[251]

Consider, once more, the example in which the agent must choose between curing a small population of people or leaving them sick. Here I will divide the agents in the original set Y_2 into three sets: two infinite sets of agents, Y_3 and Y_4, and one set of 50 agents, Y_2 (not to be confused with the original). I will also divide the agents in the original set Z_1 into three sets: two infinite sets of agents, Z_2 and Z_3, and one set of 50 agents, Z_1 (again, not to be confused with the original). Under this way of carving up agents, the example is as follows:

	S								
	X 50	Y_1 ∞	Y_2 50	Y_3 ∞	Y_4 ∞	Z_1 50	Z_2 ∞	Z_3 ∞	Z_4 ∞
a	1	1	0	0	0	–	–	–	–
b	0	–	–	–	–	1	1	1	0

Figure 93: Subsets of agents in the cure vs. leave sick case

By the principle 'if you can easily alleviate the suffering of nearby agents at little or no cost to yourself, you are permitted to do so and forbidden from failing to do so', act b is objectively impermissible and act a is objectively permissible. However, by the Permutation Principle over Act Outcomes, for any bijection g from the population of the outcomes of a and b onto any population, there exists an act pair $\langle c, d\rangle$ that is a qualitative duplicate of $\langle a, b\rangle$ but whose outcome population has been permuted by g. Let g be as follows, where J_1, J_2, and J_3 are three infinite sets of agents that do not exist in the outcomes of act a or b:

[251] Rulli and Worsnip [193] argue that non-consequentialist beliefs about permissibility support violations of the Independence of Irrelevant Alternatives. I believe that many will find this principle acceptable in the cases I discuss, but I will formulate a variant of the problem below that does not require it.

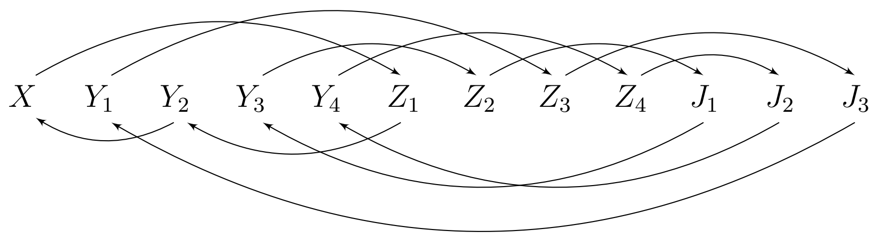


Figure 94: A permutation of subsets of agents in the cure vs. leave sick case

Let the pair of actions $\langle c, d\rangle$ be a qualitative duplicate of $\langle a, b\rangle$ but whose outcome population has been permuted by g. And let the pair of actions $\langle e, f\rangle$ be a qualitative duplicate of $\langle c, d\rangle$ but whose outcome population has been permuted by g (i.e. $\langle e, f\rangle$ is a qualitative duplicate of $\langle a, b\rangle$ under $g \circ g$). Suppose we were to add $\langle c, d\rangle$ and $\langle e, f\rangle$ to the agent's choice set:

	S											
	X	Y_1	Y_2	Y_3	Y_4	Z_1	Z_2	Z_3	Z_4	J_1	J_2	J_3
	50	∞	50	∞	∞	50	∞	∞	∞	∞	∞	∞
a	**1**	1	0	0	0	–	–	–	–	–	–	–
b	**0**	–	–	–	–	1	1	1	0	–	–	–
c	0	–	–	–	–	**1**	0	1	0	–	–	–
d	–	–	1	–	–	**0**	–	–	–	1	0	1
e	–	–	**1**	–	–	0	–	–	–	0	0	1
f	1	1	**0**	1	0	–	–	–	–	–	–	–

Figure 95: The outcomes of $\langle a, b\rangle$ and its duplicates under a permutation

By the principle 'if you can easily alleviate the suffering of nearby agents at little or no cost to yourself, you are permitted to do so and forbidden from failing to do so', act b is impermissible when our choice set is $\{a, b\}$. If our choice set is extended to include c and d then b remains impermissible by the Independence of Irrelevant Alternatives, c is impermissible by Infinite Act Pareto, and d is impermissible by the Qualitativeness of Objective Permissibility.[252] If

[252] Depending on our theory of act individuation, it may be claimed that the pair $\langle c, d\rangle$ cannot be part of the same choice set as the pair $\langle a, b\rangle$ while remaining a qualitative duplicate of this pair. The problem I pose below does not depend on the pair $\langle c, d\rangle$ being part of the same choice set as $\langle a, b\rangle$.

we include e and f in our choice set then b, c and d remain impermissible by the Independence of Irrelevant Alternatives, e is impermissible by Infinite Act Pareto and f is impermissible by the Qualitativeness of Objective Permissibility. But now that f is in the choice set, a is *also* impermissible by Infinite Act Pareto. So none of the acts are permissible!

Therefore if we accept the four axioms above and we claim that act a is permissible and act b is impermissible then, in a choice situation like this one involving two qualitatively identical pairs of acts, all of the actions are objectively impermissible. Note that this will be true regardless of which ethical principle we appeal to when we claim that it is permissible to cure the small population and impermissible to fail to do so. For any such principle, we can find a qualitatively identical pair of actions such that, when faced with the choice between these actions, all of the actions available to us are objectively impermissible.

It might be objected that the existence of moral dilemmas in choice scenarios is not so concerning if agents will rarely face the choice scenarios in question. While the argument above shows that all principles that a is permissible and b is impermissible when the outcomes of a and b are incomparable will in principle lead to moral dilemmas like the one above, agents will rarely face such dilemmas. I believe that the existence of such choice scenarios should cause us to be more wary of principles which says that a is permissible and b is impermissible, even if agents will rarely be faced with scenarios in which an actual conflict arises. We can, however, generate a more direct problem for the claim that such principles exist if we replace the Independence of Irrelevant Alternatives with an analogue of Transitivity for objective permissibility, such as the following:

Strong Transitivity of Objective Permissibility

If act a is permissible when an agent can perform a or b and act b is permissible when the agent's choice set is $\{b, c\}$ then act a is permissible when the agent's choice set is $\{a, c\}$

The Strong Transitivity of Objective Permissibility says that the relation 'is a morally permissible alternative to' is a transitive relation. If a is permissible when b is the alternative and b is permissible when c is the alternative then a is permissible when c is the alternative. Many have taken this principle to be plausible because what it is morally permissible, obligatory, or impermissible for us to do is generally thought to depend on the moral reasons we have for acting. If we have more moral reasons to do a than to do b (because doing a will easily alleviate the suffering of nearby agents while b will not) and we have more moral reasons to do b than to do c (because b will make infinitely many agents better off than c will), then it seems to follow that we must have more moral reasons to do a than to do c.

Kamm [123] and Dorsey [72] both raise a problem for this principle in cases that involve choice sets containing a supererogatory action and a permissible action, however. Let me offer my own version of this problem.[253] Suppose that if you have the choice between donating half of a pile of money and burning the other half (act a) or keeping the pile of money for yourself (act b), it is morally permissible to do either because donating half of the money is supererogatory and keeping it is permissible. Suppose that if you have the choice between keeping the pile of money for yourself (act b) or donating all of the money (act c), it is morally permissible to do either because donating all of the money is supererogatory and keeping it is permissible. Now suppose that you have the choice between donating half of the pile of money and burning the other half (act a) or donating all of the money (act c). It would be impermissible to donate half of the pile of money and burn the other half in this scenario. This seems to violate the Strong Transitivity of Objective Permissibility.

There have been several responses to the 'transitivity problem' for supererogation.[254] One key thing to notice is that it is your self-interested reasons that make it permissible for you to keep the $100 when your other option is to donate all or half of the money. In the final

[253] A problem with the case discussed in Dorsey [72, p. 365-6] is that it does not involve identical actions.
[254] See, for example, Portmore [183] and Archer [5].

choice scenario, however, it is impermissible to choose the action that you have fewer moral reasons to perform because there are no self-interested reasons at play.[255] Importantly, for supererogatory acts to generate a violation of the Strong Transitivity of Objective Permissibility, the agent must have self-interested reasons that make b permissible when the choice set is $\{a, b\}$ and self-interested reason that make b permissible when the choice set is $\{b, c\}$.[256] We can therefore avoid the problems generated by choice sets containing supererogatary acts by endorsing a weaker principle, which is as follows:[257]

Transitivity of Objective Permissibility

If a is objectively permissible and b is objectively impermissible when an agent's choice set is $\{a, b\}$ *and act b is objectively permissible and c is objectively impermissible when the agent's choice set is* $\{b, c\}$ *then a is objectively permissible when the agent's choice set is* $\{a, c\}$

The Transitivity of Objective Permissibility says that if a is permissible and b is impermissible when these are our only two options, then a must be permissible relative to any action that is an impermissible alternative to b. It is consistent with the claim that permissibility is intransitive across choice sets in which both actions are permissible.

The Transitivity of Objective Permissibility is in tension with contrastivism about objective permissibility. Contrastivists about objective permissibility hold that what an agent is objectively permitted to do is relative to a comparison set: I am objectively permitted to save one child if the comparison is saving no children, but not if the comparison is saving three children. Some contrastivists about reasons, such as Snedegar [213], argue that something can be a reason for A when the comparison is B and a reason for B when the comparison

[255] This is unsurprising if we think that whether an action is supererogatory or not is a function of the moral reasons for performing the action and the self-interested reasons for failing to perform the action.

[256] Interestingly, we cannot generate a similar problem for the Strong Transitivity of Impermissibility using suberogation, since suberogation states that an action is wrong but not impermissible.

[257] Alternatively, we could deny that non-moral reasons can make actions that one has fewer moral reasons to perform permissible or we restrict this principle to cases that involve no non-moral reasons.

is C, but can fail to be a reason for A when the comparison is C. This causes Snedegar to reject the following transitivity principle: 'If there is more reason to choose A than to choose B, and more reason to choose B than to choose C, then there is more reason to choose A than to choose C' [213, p. 94]. If it is objectively permissible to perform a if and only if one has more moral reasons for performing a than for performing b, then the Transitivity of Objective Permissibility is in tension with the claim that moral reasons have this structure.

The reasons that Snedegar gives for rejecting the Transitivity of 'More Reasons' mirror the reasons that Temkin gives for rejecting the Transitivity of $\succcurlyeq$. Snedegar argues that different 'reason-providing objectives' such as happiness or justice [213, p. 68] – 'vary in importance across comparisons' [213, p. 109] and that this variation in importance results in violations of transitivity across choice sets. If we believe that there are different moral objectives that vary in importance across choice sets, then we can reject the Transitivity of Objective Permissibility.[258] If, on the other hand, we believe that moral objectives do not vary in importance depending on what is being compared, then the Transitivity of Objective Permissibility looks like a highly plausible principle of objective permissibility.

Suppose we find the Transitivity of Objective Permissibility more compelling than contrastivism about moral reasons. When our choice is between curing the small population (act a) and failing to do so (act b), act a is permissible and act b is impermissible by hypothesis. When our choice is between act b and act c, act b is permissible and act c is impermissible by Infinite Act Pareto. When our choice is between act c and act d, act c is permissible and act d is impermissible by the Qualitativeness of Objective Permissibility (QOP). When our choice is between act d and act e, act d is permissible and act e is impermissible by Infinite Act Pareto. When our choice is between act e and act f, act e is permissible and act f

[258] In the case in which we can cure the small population, contrastivists can claim that a is permissible when b is the alternative and b is permissible when c is the alternative and c is permissible when d is the alternative and d is permissible when a is the alternative. An important outstanding problem for this view is to provide an account of what it is objectively permissible or impermissible to do in these circumstances.

is impermissible by the Qualitativeness of Objective Permissibility. And when our choice is between act f and act a, act f is permissible and act a is impermissible by Infinite Act Pareto (IAP). Let $a \longrightarrow b$ mean that act a is permissible and act b is impermissible if the agent's choice set is $\{a, b\}$. This situation is as follows:

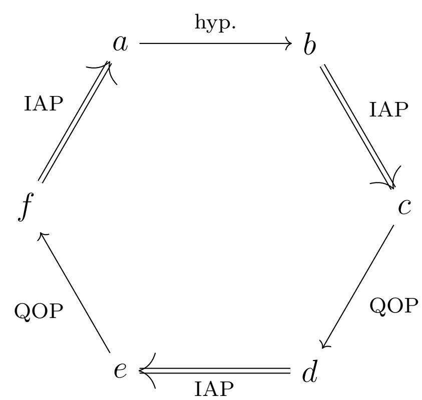


Figure 96: Violation of the transitivity of objective permissibility

Act a is impermissible when the agent's choice set is $\{a, f\}$, which violates the Transitivity of Objective Permissibility. Therefore it cannot be the case that a is permissible and b is impermissible. It is easy to see that if act a were impermissible and act b were permissible by hypothesis, we could produce a violation of the Transitivity of Objective Permissibility by the same reasoning.[259] Therefore we cannot use *any* principle to say that it is permissible to cure the small population and impermissible to fail to do so unless we reject Infinite Act Pareto, the Qualitativeness of Objective Permissibility, the Permutation Principle over Act Outcomes, or the Transitivity of Objective Permissibility.

If we accept Infinite Act Pareto, the Qualitativeness of Objective Permissibility, the Permutation Principle over Act Outcomes, and the Transitivity of Objective Permissibility then we must conclude that if there is a bidirectional upgrade from the population of the outcome of

[259]This direction is much easier: we can show it using an argument that involves just two act pairs. For the structure this will take, see the four world example immediately preceding Result 10 of Chapter 3.

act a to the population of the outcome of act b and there is a bidirectional upgrade from the population of the outcome of act b to the population of the outcome of act a, then it cannot be the case that one of these actions is objectively permissible and the other is objectively impermissible. As I argued above, if the world is infinite then most of the actions available to an agent in a given choice scenario will have incomparable outcomes. This view therefore leads to either ubiquitous permissibility or ubiquitous impermissibility in infinite worlds.

In this section I have presented a new argument for the claim that in cases where the outcomes of our actions differ over infinite populations, there is either ubiquitous objective permissibility across actions or there is ubiquitous objective impermissibility across actions. This result holds regardless of whether we adopt a consequentialist or non-consequentialist theory of objective permissibility. In the next section I will demonstrate that a similar problem also affects all theories of subjective permissibility.

5.4.2 Incomparability and Subjective Permissibility

Actions are subjectively permissible or impermissible if they are permissible given an agent's imperfect information. As we saw in the miners puzzle, actions that are objectively permissible may be subjectively impermissible and actions that are subjectively permissible may be objectively impermissible. In this section I will show that existing outcomes-based accounts of subjectively permissible actions are inadequate in cases where the outcomes of some of our actions are incomparable. I will then show that the puzzle of permissibility outlined in the last section can be applied to subjective permissibility also.

In this section I will use Savage's framework for decision making under uncertainty to model subjective ethical decisions. I assume that agents assign credences to different possible ways that the world might be.[260] Each action available to the agent leads to some outcome in

[260] Credences are subjective probabilities in the $[0, 1]$ interval that obey the standard probability axioms.

each possible state of the world. As before, actions are assumed to be functions from states to outcomes, where the outcome of an action is the entire world that will result if the action is performed, including both its immediate and long-term consequences.

As was noted in the previous section, some theories of permissibility – consequentialist theories in particular – posit a strong connection between the outcome of an action and the objective permissibility of an action. Unsurprisingly, these theories also tend to posit a strong connection between the outcomes that an agent expects her actions to produce and the subjective permissibility of those actions in finite worlds. For example, maximizing utilitarians claim that agents are permitted to perform an action a if and only if there is no action b available to the agent that produces more expected utility than action a does, where the expected utility of an action is the sum of the total utility of each possible outcome of the action multiplied by the agent's credence that the action will yield that outcome.

One natural constraint on subjective permissibility for utilitarian moral theories is a subjective variant of WPCO, which is as follows:[261]

Weak People Criterion for Subjective Permissibility (WPCS)

If, summing over all epistemically possible agents x, the sequence of differences in expected utility for x conditional on act a being performed rather than act b being performed converges and is greater than zero or diverges to ∞, then (i) act b is impermissible and (ii) if a and b are the only actions available to the agent, then act a is permissible

We could add an equality condition to this principle which says that if the sequence of

[261] This is this is modeled on Arntzenius' 'Weak People Criterion' [6, p. 55] which says that the expected utility of act A is greater than the expected utility of act B iff '$\sum_P(EU(P|A) - EU(P|B))$ is absolutely convergent and > 0, where we are summing over all (epistemically possible) people p.' Artnzenius extends the definition of 'absolutely convergent' to include cases where the sum diverges to positive infinity or negative infinity but not both. To this we could add that the expected utility of act A is equal to the expected utility of act B iff '$\sum_P(EU(P|A) - EU(P|B))$ is absolutely convergent and equal to 0. Utilitarians that endorse this account of the expected utility of actions in infinite worlds are likely to endorse WPCS as a principle of subjectively permissible and impermissible actions.

differences in expected utility conditional on act a being performed rather than act b being performed is equal to zero then (i) act a is permissible [impermissible] iff act b is permissible [impermissible], and (ii) if a and b are the only actions available to the agent, act a and act b are both permissible. If we were to add this equality condition then in finite worlds WPCS would entail the maximizing utilitarian criterion of subjective permissibility: act a is subjectively permissible if and only if there is no action b available to the agent that produces more expected utility than act a does. We will not need this condition here, however.

By utilitarian lights, WPCS has some fairly plausible results for actions that have infinite outcomes. Suppose that the same agents in set X will exist regardless of what we do. These agents are unhappy (utility 0) by default but the agent knows that either act a will make all future agents happy or act b will. She has a credence of 0.1 that act a will make them happy (state S_1) and a credence of 0.9 that act b will make them happy (state S_2):

	$S_1(0.1)$	$S_2(0.9)$
	X	X
	∞	∞
a	1	0
b	0	1

Figure 97: High probability improvement vs. low probability improvement

For each agent in set X, the expected utility of their lives conditional on a being performed minus the expected utility of their lives conditional on b being performed is -0.8 and the expected utility of their lives conditional on b being performed minus the expected utility of their lives conditional on a being performed is 0.8. So, summing over all epistemically possible agents (all $x \in X$), the sequence of the difference in expected utility for x if act a rather than act b is performed diverges to $-\infty$ and the sequence of the difference in expected utility for x if act b rather than act a is performed diverges to ∞. Therefore act a

is subjectively impermissible and act b is subjectively permissible according to WPCS. This is what we would expect, since in this example act b is more likely to make all agents happy.

Although it produces plausible results in the cases to which it applies, WPCS places an extremely weak restriction on subjective permissibility if agents have a non-zero credence that the world is infinite. For example, suppose that instead of the same agents being affected by her choice, the agent has a credence of 0.9 that if she performs act a then infinitely many agents in set Y will be unhappy and if she performs act b then those same agents will be happy. She also has a credence of 0.1 that if she performs act a then infinitely many agents in set X will be happy and if she performs act b then those same agents will be unhappy:

	$S_1(0.1)$	$S_2(0.9)$
	X	Y
	∞	∞
a	1	0
b	0	1

Figure 98: Disjoint population high probability improvement vs. low probability improvement

We might think that act b is permissible and act a is impermissible in this case since act b makes infinitely many people happy in the state of the world that is the most likely given the agent's evidence. But WPCS does not entail that act a is subjectively impermissible in this case. For each agent in set X, the expected utility of their lives conditional on a being performed minus the expected utility of their lives conditional on b being performed is 0.1. For each agent in set Y, the expected utility of their lives conditional on a being performed minus the expected utility of their lives conditional on b being performed is -0.9. So the sequence of the difference in expected utility for all possible agents if act a rather than act b is performed does not converge: it diverges to ∞ under some orderings and to $-\infty$ under other orderings. Therefore neither of these actions is subjectively impermissible by WPCS.

More generally, if there is some infinite set of agents that have more expected utility conditional on a being performed than b being performed and there is some infinite set of agents that have more expected utility conditional on b being performed than a being performed (where agents who don't exist are treated as thought they have utility zero lives) then neither action a nor action b will be impermissible by WPCS. Given this, very few actions will be subjectively impermissible by WPCS. For any pair of actions $\langle a, b \rangle$ that an agent can perform, it is extremely likely that the agent will have a rational non-zero credence in some possible state S_i in which an infinite set of agents are better off if a is performed than if b is performed and that she will also have a rational non-zero credence in some possible state S_i in which some different infinite set of agents are better off if b is performed than if a is performed. The evidence required to justify a non-zero credence in such a state is very low.[262] Such states may be highly unlikely given our evidence but they are consistent with our current best theories in physics and we will almost never be in a position entirely rule them out.[263] Therefore WPCS will deem very few actions subjectively impermissible.

To give an example, WPCS will fail to deliver a verdict in the subjective variant of the case given in the previous section. Suppose that an agent is confident that the world is finite (state S_1) but has a non-zero credence that it is infinite (state S_2). She can cure a small population of 50 agents (act a) or she can fail to do so (act b). If the world is finite then this will result in these agents having utility 1 rather than utility 0 lives. If the world is infinite then the action will still result in these agents having utility 1 rather than utility 0 lives, but curing the population will also affect the utility levels of infinitely many future agents. The

[262] The claim that agents can even have a rational credence of zero in logical and mathematical falsehoods is a controversial one (see Garber [90]). We might worry that there are infinitely many possible worlds consistent with the laws of physics and so the agent ought to have an infinitesimal credence in each of them (see Pruss [184]). But is seems likely that even if we have credence 0 in particular possible worlds consistent with our current best theories in physics, a probability density function would assign non-zero credence to the set of worlds that, for any given act pair, result in neither act being impermissible by WPCS.

[263] Even if they were inconsistent with our current best theories in physics, it would seem irrational to have credence zero in such states since we cannot be certain that our current best theories in physics are true.

outcomes of acts a and b in the two states that the agent is uncertain over are as follows:

	$S_1(0.9)$	$S_2(0.1)$	
	X	Y	Z
	50	∞	∞
a	1	1	0
b	0	0	1

Figure 99: Subjective variant of cure vs. leave sick case

It is intuitively better to cure the small population of agents than to fail to do so, but neither of these actions is subjectively impermissible by WPCS.

In the previous section it was demonstrated that principles stronger than WPCO will violate one of four plausible axioms for objective permissibility. We might think that subjective permissibility principles will succeed where objective permissibility principles did not. Here I will show that we can generate a similar problem for subjective permissibility principles that entail a is permissible and b is impermissible in the case above.

There are several kinds of principles that we could use to restrict the set of actions that are subjectively permissible if some of the actions available to an agent produce incomparable outcomes. For example, consider the following intuitively plausible dominance principle for subjective permissibility if the outcomes of some acts in our choice set may be incomparable:

Naïve Dominance

If, for every possible state S the outcome of act a in S is not worse than the outcome of act b in S, and for some possible state S the outcome of a in S is strictly better than the outcome of b in S, then (i) b is subjectively impermissible and (ii) if a and b are the only options available to the agent, then a is subjectively permissible

Naïve Dominance entails it is better to cure the small population of agents than to fail to do so in the case given earlier, because the outcome of curing the population (act a) is better in state than failing to cure the population (act b) in state S_1, and the outcomes of both acts are incomparable in state S_2. I will now show that any principle that says it is better to cure the small population of agents than to fail to do so in the case above will violate one of four plausible principles of subjective permissibility.

The first principle of subjective permissibility that I take to be plausible is a subjective analogue of Infinite Act Pareto. Like infinite Infinite Act Pareto, it posits a weak connection between the outcomes of actions and the subjective permissibility of actions:[264]

No Infinite Risks

If an agent has a non-zero credence that she is in a state in which infinitely many people will be worse off by some positive finite amount if act b is performed than if act a is performed, and the agent has credence zero that she is in a state in which infinitely many people will be worse off by some positive finite amount if act a is performed than if act b is performed, then (i) act b is subjectively impermissible and (ii) if a and b are the only options available to the agent, then act a is permissible

No Infinite Risks says that we should not risk undertaking an action that has some chance of making infinitely many agents worse off and no chance of making infinitely many agents better off, even if there are some states of the world in which this action benefits finitely many agents. I believe that many ethicists will be inclined to endorse this weak link between the outcomes of actions and the subjective permissibility of actions.

If we accept Naïve Dominance and No Infinite Risks, then we will end up violating a further

[264] We may believe that we can generate this puzzle using an even weaker principle than No Infinite Risks that does not involve trade-offs. The key problem with using principles that do not involve trade-offs when we are dealing with decisions under uncertainty is that we must preserve the qualitative properties of each outcome under a permutation. Since some outcomes (the finite ones) will not admit of bidirectional upgrades in both directions, we cannot show that acts with a non-zero chance of producing these outcomes are better than a principle that is more like Pareto: i.e. one that does not involve any trade-offs.

plausible principle of subjective permissibility. To show this, consider the following case:

	$S_1(0.9)$	$S_2(0.1)$		
	X	Y	Z_1	Z_2
	50	∞	∞	∞
a	2	1	0	0
b	0	0	1	1
c	0	1	1	0

Figure 100: Problem case for joint acceptance of Naïve Dominance and No Infinite Risks

If the agent's choice set were $\{a, b\}$ then a would be permissible and b would be impermissible by Naïve Dominance. If the agent's choice set were $\{b, c\}$ then b would be permissible and c would be impermissible by Naïve Dominance. But if the agent's choice set were $\{a, c\}$ (or $\{a, b, c\}$) then c would be permissible and a would be impermissible by No Infinite Risks. This violates the subjective variant of the Transitivity of Objective Permissibility:

Transitivity of Subjective Permissibility

If, relative to a single credal distribution that an agent has across possible states, act a is subjectively permissible and b is subjectively impermissible when the agent's choice set is $\{a, b\}$ *and b is subjectively permissible and c is subjectively impermissible when the agent's choice set is* $\{b, c\}$ *then a is subjectively permissible when the agent's choice set is* $\{a, c\}$

The Transitivity of Subjective Permissibility is restricted to a single credal distribution across states because it is possible that a is subjectively permissible and b is subjectively impermissible relative to one credal distribution and b is subjectively permissible and c is subjectively impermissible relative to a different credal distribution. In this case it would not follow that a is subjectively permissible relative to the second credal distribution. This subjective variant of the Transitivity of Subjective Permissibility is intuitively plausible, especially if we believe that different moral reasons should not be given different weights across choice scenarios.[265] The case above shows that if we accept the Transitivity of Subjective Permissibility and No Infinite Risks then we must reject Naïve Dominance.

In the case above I simply assumed that we could find three acts a, b, and c that generate a conflict between Naïve Dominance and No Infinite Risks. We can generalize this problem to *any* principle that entails that it is better to cure the small population of agents than to fail to do so in the case given at the beginning of this section. To do so, we simply need to endorse the Permutation Principle over Act Outcomes from the previous section, and the following subjective variant of the Qualitativeness of Objective Permissibility:

Qualitativeness of Subjective Permissibility

If $\langle c, d, Pr' \rangle$ is a qualitative duplicate of $\langle a, b, Pr \rangle$, where Pr and Pr' are probability functions, then d is subjectively impermissible if and only if b is subjectively impermissible

If we accept No Infinite Risks, the Transitivity of Subjective Permissibility, the Permutation Principle over Act Outcomes, and the Qualitativeness of Subjective Permissibility then it cannot be permissible to cure the small population of agents and impermissible to fail to do so in the subjective version of this case. To show this, let $\langle c, d, Pr' \rangle$ be a qualitative duplicate of $\langle a, b, Pr \rangle$ but whose outcome population has been permuted by g:

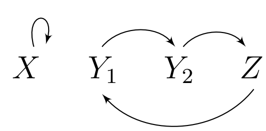


Figure 101: A permutation of the outcome population of $\langle a, b, Pr \rangle$

The outcomes $\langle a, b, Pr \rangle$ and its qualitative duplicate $\langle c, d, Pr' \rangle$ are as follows:

[265]As with the Transitivity of Objective Permissibility, the Transitivity of Subjective Permissibility is in tension with contrastivism about moral reasons.

	$S_1(0.9)$	$S_2(0.1)$		
	X	Y_1	Y_2	Z
	50	∞	∞	∞
a	1	0	0	1
b	0	1	1	0
c	1	1	0	0
d	0	0	1	1

Figure 102: The outcomes of $\langle a, b, Pr \rangle$ and $\langle c, d, Pr' \rangle$

If the agent's option set is $\{a, b\}$ then act a is permissible and act b is impermissible by hypothesis. If the agent's option set is $\{b, c\}$ then act b is permissible and act c is impermissible by No Infinite Risks. If the agent's option set is $\{c, d\}$ then act c is permissible and act d is impermissible by the Qualitativeness of Subjective Permissibility (QSP). And if the agent's option set is $\{a, d\}$ then act d is permissible and act a is impermissible by No Infinite Risks (NIR). This violates the Transitivity of Subjective Permissibility:

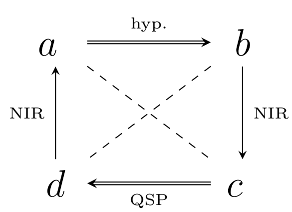


Figure 103: Transitivity violation from NIR and QSP

It is easy to see that if act a were impermissible and act b were permissible by hypothesis, we could produce a violation of the Transitivity of Subjective Permissibility by the same argument. We therefore cannot use *any* principle to say that a is permissible and b is impermissible unless we reject one of four plausible axioms of subjective permissibility: No Infinite Risks, the Qualitativeness of Subjective Permissibility, the Permutation Principle over Act Outcomes, or the Transitivity of Subjective Permissibility.[266]

[266]This applies to non-consequentialist principles such as 'if you are confident that you can easily alleviate

This argument shows that if we accept No Infinite Risks, the Qualitativeness of Subjective Permissibility, the Permutation Principle over Act Outcomes, and the Transitivity of Subjective Permissibility then it cannot be the case that it is permissible to cure the small population and that it is impermissible to fail to do so. This is highly troubling when consider the fact that we can replace 'fail to cure the small population' with 'fail to torture an otherwise happy small population' or any other action that is intuitively impermissible and whose outcomes have the same utility profiles as the actions given in this example. If we want to say that such actions are impermissible then we must reject one of the four plausible principles for subjective permissibility formulated above.

I will not attempt to establish a generalization of the puzzle for subjective permissibility that I have constructed in this section. I believe it is sufficiently troubling to show that we cannot endorse intuitive claims about the subjective impermissibility of failing to cure a small population of sick agents (or of torturing a small population of otherwise happy agents) if we have a non-zero credence that the outcomes of these acts are incomparable.[267]

It may be worth conjecturing about whether and how this result will generalize, however. It seems plausible that we could establish the following modest generalization of the case above: if we accept No Infinite Risks, the Qualitativeness of Subjective Permissibility, the Permutation Principle over Act Outcomes, and the Transitivity of Subjective Permissibility then if the agent's choice set is $\{a, b\}$ and (i) she has a non-zero credence in at least one

the suffering of nearby agents at little or no cost to yourself, you are permitted to do so and forbidden from failing to do so'. It also applies to more consequentialist principles that are stronger than WPCS, such as discounted utilitarianism or an ethical variant of Hare's [104, p. 242-3] 'Prospectism' or principles that appeal to the accumulation intervals of expected utilities. It also applies to principles that tell us to simply act as if we had credence zero in all states that produce incomparable outcomes.

[267] A distinct problem for theories of subjective permissibility if act outcomes are infinite but *comparable* is the so-called 'fanaticism' problem. If act a is guaranteed to produce some very large finite amount of positive value and act b will almost certainly produce some very large finite negative amount of value but has a minuscule non-zero chance of producing a positively infinite amount of value, then 'fanatical' moral theories entail that act b is subjectively permissible and act a is subjectively impermissible. Beckstead [26, 6.4, 6.5; 8.2] argues that the alternatives to fanaticism in such cases have similarly undesirable results.

possible state S_j such that the outcome of a in S_j is incomparable with the outcome of b in S_j by the results of Chapter 3, and (ii) there are finitely many other states that she has a non-zero credence in and all of these states are finite, it cannot be the case that one of the actions in her choice set is subjectively permissible and the other is subjectively impermissible.

To see why I believe that such a generalization of the puzzle formulated above can be established, consider two acts a and b. Suppose that an agent's choice set is $\{a, b\}$ and there is some state $S_{n'}$ such that (a) the agent has a non-zero credence in $S_{n'}$ and (b) the outcomes of a and b in $S_{n'}$ are incomparable by a result from Chapter 3. Let all of the other states that the agent has a non-zero credence in be finite and let a and b have any properties whatsoever in these states: the acts can result in any number of agents with any utility levels, can have any ethically relevant properties like 'violates rights' or 'is virtuous' and so on. For example:

	$S_1, ..., S_n(0.999)$	$S_{n'}(0.001)$		
	$X_1, ..., X_n$	Y_1	Y_2	Z
	n	∞	∞	∞
a	i	0	0	1
b	j	1	1	0

Figure 104: Case generalizing the subjective permissibility puzzle

For any such act pair, we can permute each of the outcomes of the acts in finite states of the world by themselves. In other words, if S_i is a state that results in a finite outcome (an outcome in which only finitely many agents have different utility conditional on a being performed rather than b being performed) then let g be a trivial permutation of the outcomes of a and b in S_i. For all states S_j that can be shown to be incomparable by a result from Chapter 3, there is either a permutation of the outcomes of a and b in S_j that results in a four world violation of the Transitivity of Subjective Permissibility (given No Infinite Risks) or there is a permutation of the outcomes of a and b in S_j that results in a cyclic violation of the Transitivity of Subjective Permissibility. Let us permute each of the outcomes of a and b by these permutations. This means that in the option set $\{b, g(a)\}$, b is permissible and $g(a)$ is impermissible by No Infinite Risks, and in the option set $\{g(b), g(g(a))\}$, $g(b)$ is permissible and $g(g(a))$ is permissible by infinite risks, and there is some $g^n(b)$ such that in the option set $\{g^n(b), a\}$, $g^n(b)$ is permissible and a is impermissible by No Infinite Risks. This violates the Transitivity of Subjective Permissibility.

This is a sketch of the sort of argument we would use to show that it cannot be the case that a is permissible and b is impermissible in these conditions. Since the outcomes of a and b can be shown to be incomparable by a result from Chapter 3, we will always be able to show that it cannot be the case that b is permissible and a is impermissible by the same type of argument. Therefore if the agent's choice set is $\{a, b\}$ then it cannot be the case that one of these actions is permissible and the other is impermissible.

In the case above, we can establish this using a four world argument since the outcomes of a and b in $S_{n'}$ are incomparable by Result 8. Given this, we can establish that it cannot be the case that a is permissible and b is impermissible by permuting the populations of the outcome of a and b in $S_{n'}$ by g:

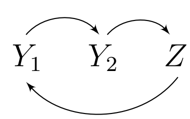


Figure 105: A permutation of the outcome population of $\langle a, b, Pr \rangle$

The outcomes of $\langle c, d, Pr' \rangle$, a qualitative duplicate of $\langle a, b, Pr \rangle$ under g, are as follows:

	$S_1, ..., S_n (0.999)$	$S_{n'} (0.001)$		
	$X_1, ..., X_n$	Y_1	Y_2	Z
	n	∞	∞	∞
a	i	0	0	1
b	j	1	1	0
c	i	1	0	0
d	j	0	1	1

Figure 106: The outcome of $\langle a, b, Pr \rangle$ and its qualitative duplicate under g

If a is permissible and b is impermissible when the agent's choice set is $\{a, b\}$ then c is permissible and d is impermissible when the agent's choice set is $\{c, d\}$ by the Qualitativeness of Subjective Permissibility. We can see that if the agent's choice set is $\{b, c\}$ then b is permissible and c is impermissible by No Infinite Risks (since i and j are finite) and if the agent's choice set is $\{d, a\}$ then d is permissible and a is impermissible by No Infinite Risks (since i and j are finite). Therefore the claim that a is permissible and b is impermissible violates the Transitivity of Subjective Permissibility.

This is a sketch of a modest generalization of the puzzle that I have formulated in this chapter. An important outstanding question is how much this puzzle can be generalized beyond choice situations with these properties. For example, can the puzzle be extended to actions that produce outcomes that are incomparable in at least one state of the world if there are also some states of the world in which one act produces an outcome that is better for infinitely many agents and there are some states of the world in which the other act produces an outcome that is better for infinitely many agents? What if there are some states of the world in which one act produces an outcome that is better for infinitely many agents but there are no states of the world in which the other act produces an outcome that is better for infinitely many agents? Finally, what if there is no state such that the outcome of a and the outcome of b are incomparable in that state, but the outcome of act a in some state S_i is incomparable with the outcome of act b in a different state S_j?

I will not attempt to answer these questions here. I believe that the puzzle for subjective permissibility that I have introduced in this section gives us sufficient reason to believe that accepting incomparability between infinite worlds will have important ramifications for both consequentialist and non-consequentialist ethics. Discovering the extent to which this puzzle can be generalized must be the subject of future research.

This leaves us in a difficult position, since many of the reasons that have to avoid rejecting any of the four axioms outlined at the beginning of this chapter – Transitivity, the Permutation Principle, the Qualitativeness of $\succcurlyeq$, and Pareto – provide us with analogous reasons to want to avoid the variants of these for both objective and subjective permissibility.

The Transitivity of Subjective Permissibility seems highly plausible if we do not believe that the permissibility of actions is highly dependent on our choice set. If b is subjectively impermissible when our choice set if $\{a, b\}$ and c is subjectively impermissible when our choice set is $\{b, c\}$, then surely a cannot be permissible when our choice set is $\{a, c\}$. If we reject the Transitivity of Subjective Permissibility then we must conclude that, in the case that has been the focus of this section, it is impermissible to fail to cure the small population if our choice set is $\{a, b\}$, it is impermissible to cure the small population if our option set is $\{b, c\}$, it is impermissible to fail to cure the small population if our option set is $\{c, d\}$, and it is impermissible to cure the small population if our option set is $\{a, d\}$. This doesn't appear to be a plausible account of which actions are permissible and impermissible.

The Permutation Principle over Act Outcomes merely says that we can permute the populations of the outcomes of our actions. In doing so we can assess how well our ethical principles work across scenarios that only differ by who is affected by our choices. This too seems like a relatively unobjectionable axiom. If we reject this axiom on metaphysical grounds then it also seems likely that, as with the original Permutation Principle, we can still derive the problematic results detailed above by weakening our qualitativeness axiom.

If we deny the Qualitativeness of Subjective Permissibility then we must conclude that even if one act pair and probability function $\langle c, d, Pr' \rangle$ is a duplicate of another act pair and probability function $\langle a, b, Pr \rangle$ in all qualitative respects, it can nonetheless be the case that b is impermissible when our choice set is $\{a, b\}$ but d is not impermissible when our choice set is $\{c, d\}$. It is difficult to see how we could accept this conclusion unless we believe that the permissibility of actions is determined not only by qualitative facts about those who are affected by our actions, such as the nature of our relationship with them, but also haecceitistic facts like 'Obama is affected by this action'. Surely most ethicists would want to reject the claim that haecceitistic facts can play such a role in the permissibility of actions.

One option available to us is to deny No Infinite Risks. If we reject this principle then we accept that it can be permissible to undertake actions that have some chance of causing harm to infinitely many people and that have no chance of benefiting infinitely many people. Denying this principle seems to truly sever even the most basic connection between the permissibility of our actions and the outcomes of our actions. Suppose we believe, as many non-consequentialists do, that we have a duty not to steal from others. We also believe that it can be impermissible to steal from others even if stealing from them will lead to a better outcome. For example, it may be impermissible to throw someone's mobile phone from a moving car in order to prevent them from texting while driving, even if the outcome in which the person stops texting while driving and must buy a new phone is better than the outcome in which they continue to endanger others by texting while driving. It is difficult to maintain this judgment regardless of what is at stake, however. Suppose that you know someone intends to commit a mass shooting. The only way that you can prevent them from doing so is to steal the key to the cabinet in which they keep their guns. Surely stealing this key is not only permissible but is morally required in these circumstances. This case only involves finite trade-offs, however. If we deny No Infinite Risks then we must conclude that an action can remain impermissible even if, by undertaking it, we have some chance of preventing harms not just to a large finite number of people, but to infinitely many people.

If we do not reject one of these axioms, however, then we will be forced to conclude that it is permissible to fail to cure a population of sick agents rather than curing them, and that it is permissible to torture a population of agents rather than refraining from doing so. And these actions are permissible only because we have a non-zero credence that the world is infinite. This constitutes a kind of moral nihilism that most of us would consider reprehensible.

Summary

At the beginning of this chapter I characterized the incomparability results of Chapters 2 and 3 as an impossibility result: we cannot accept the Transitivity of $\succcurlyeq$, the Permutation Principle, the Qualitativeness of $\succcurlyeq$, and the Pareto principle without concluding that almost all infinite world pairs are incomparable. I have argued that the Transitivity of $\succcurlyeq$, the Permutation Principle, the Qualitativeness of $\succcurlyeq$, and the Pareto principle are all highly plausible principles and that denying any of them will come at a great cost. I then demonstrated that accepting ubiquitous incomparability between infinite world pairs creates analogous problems for accounts of both objective and subjective permissibility: problems that create great difficulties for both consequentialist and non-consequentialist ethics. The incomparability results in infinite ethics that I have formulated should therefore not be dismissed as little more than abstract puzzles. Regardless of which axiom or axioms we ultimately reject, the possibility that the world is infinite clearly has a profound effect on ethics.

Conclusion

In this thesis I have argued that if we accept four highly plausible axioms: the transitivity of $\succcurlyeq$, the Permutation Principle, the qualitativeness of $\succcurlyeq$, and Pareto, there is ubiquitous incomparability between infinite worlds. In this concluding section I am going to offer a review of the prior results in infinite ethics and provide an overview of the many ways in which the results of this thesis improve the framework for infinite ethics and offer novel results that take important steps beyond these prior results.

Infinite ethics has primarily been discussed in two domains: in the literature on intergenerational equity within economics and by aggregative ethicists within philosophy. The most relevant major results within intergenerational equity are the Diamond-Basu-Mitra results[268] which shows that any complete, transitive ranking of possible streams of future generations cannot satisfy Generational Sensitivity and Finite Generational Anonymity while also being numerically representable. Generational Sensitivity says that if we make one generation strictly better off without making any generations worse off, the resulting infinite stream of generations is better. The Finite Generational Anonymity axiom says that we can can switch the utilities of finitely many generations in a given infinite stream of generations without making the infinite stream in question better or worse. The motivation for Generational Sensitivity is that we should always treat local increases in utility as valuable even if there are infinitely many generations. The motivation for Finite Generational Anonymity is that fairness requires that we do not favor increases in welfare at particular generations.

The Diamond-Basu-Mitra result shows that if we try to satisfy both Generational Sensitivity and Finite Generational Anonymity then we will not be able to generate a ranking from possible streams of generations onto the real numbers. As I noted in Chapter 1, we might think

[268] Diamond (1965); Basu and Mitra (2007).

that this is not such a troubling result since the benefits of numerical representability may not be that great if we are able to provide a complete, transitive ranking of infinite worlds: a social welfare ordering rather than a social welfare function, to use existing terminology.

The Lauwers-Zame result creates worries for this approach. Lauwers (2007) and Zame (2006) show that if we accept Finite Generational Anonymity then any transitive and complete social welfare ordering will strictly rank very few infinite streams of generations. This means that such a social welfare ordering will entail that a social welfare ordering that satisfies Finite Generational Anonymity will entail that most infinite worlds are as good as each other. Finally, Lauwers and Zame show that describing a social welfare ordering that satisfies Generational Sensitivity and Finite Generational Anonymity requires the axiom of choice. This is an important result since the use of non-constructive mathematics in ethics and economics is considered highly suspect.

Within the literature in ethics, Cain (1995) crucially identified a conflict between agent-based sensitivity principles and time-based sensitivity principles. Philosophers have also constructed various methods for ranking infinite world pairs or build upon existing solutions developed within economics. Many of these are surveyed in the first chapter, but for the purposes of this thesis the most important theories have been Arntzenius's (2014) extension of the Expansionist theory developed by Kagan and Vallentyne (1997) and the characterization of the Weak Catching-Up principle offered by Lauwers and Vallentyne (2004).

I believe that in this thesis I have developed a general framework for discussing infinite ethics that offers significant advantages over the frameworks employed by those working in intergenerational equity and in philosophy. Within intergenerational equity there are several assumptions made about the framing of the problem that have restricted the scope of the results that within this literature. First, it has generally been assumed that we should consider the utility levels of wellbeing of generations of agents. As I demonstrated in Chapter 1, the standard interpretation of generations seems to be whatever collection of agents exist during a given temporal period. This means that Generational Sensitivity can be shown to conflict with Pareto, our agent-based sensitivity principle, when agents can be born in different orders. This use of generations has also led to a reliance on order-dependent principles. If agents are basic locations of value and they lack any natural order, then we should not expect theories in infinite ethics to appeal to order-dependent principles.

The use of generations within the intergenerational equity literature has also led to a failure to identify whether future generations are assumed to be identical, overlapping, or disjoint, or to recognize the moral significance of these differences. If generations are disjoint then Generational Sensitivity may seem less plausible, since it is not clear that there is any sense in which we are 'improving' finitely many generations rather than simply bringing different generations with different levels of wellbeing into existence. Philosophers have been somewhat more sensitive to the ethical significance of whether the populations of worlds are identical, disjoint, or overlapping, but this has generally only resulted in restricting principles to identical population cases and noting the difficulties of comparing worlds that have disjoint or overlapping populations. This is especially troubling when we consider the fact that, given how easy it is to entirely change the population of the future, it is likely that disjoint and overlapping population worlds will be the norm and not the exception.

In this thesis I set aside Generational Sensitivity and focused on Pareto, the agent-based sensitivity principle. I showed that recent objections to Pareto given by Hamkins and Montero (2000) are based on the implicit idea that $\succcurlyeq$ is a qualitative internal relation. I argued that we have no reason to believe that $\succcurlyeq$ is a qualitative internal relation, but that we do have strong reasons to believe that it is a qualitative relation. If $\succcurlyeq$ is a qualitative relation then the relation holds between the worlds in the pair $\langle w_2, w_2 \rangle$ if and only if it holds between all qualitative duplicates of this pair. In Chapter 2 I argued that the qualitativeness of $\succcurlyeq$, in combination with Pareto, constitutes a novel principle that can capture the basic 'equity' desideratum that our theory does not favor increases in utility for particular people that Finite Anonymity axioms were intended to capture. The qualitativeness of $\succcurlyeq$ entails that even if our ethical theories are sensitive to patterns of identities between worlds – as the must be if our theory is to be Paretian – they should not be sensitive to haecceitistic facts. An ethical theory that gives priority to increases in utility based on facts like 'this agent is Obama' seem to be clearly inequitable. When we combine the qualitativeness of $\succcurlyeq$ with Pareto it follows that if two worlds are not Pareto comparable and the utility levels of the agents in the first world is a permutation of the utility levels of the agents in the second world, then neither world is strictly better than the other.

The introduction and defense of the qualitativeness of $\succcurlyeq$ over Finite Anonymity axioms represents an important step in the literature on infinite ethics. Finite Anonymity axioms can be denied by anyone who believes that permuting the utility levels of finitely many agents may sometimes make a world better or worse. They are also too weak to rule our intuitively inequitable theories like temporal discounting theories. The qualitativeness of $\succcurlyeq$ can only be denied by those willing to defend the claim that haecceitistic facts are of ethical significance. Moreover, when combined with Pareto and our auxiliary premises, the qualitativeness of $\succcurlyeq$ conflicts with intuitively inequitable theories like temporal discounting.

In Chapters 2 and 5 I defend the Permutation Principle, which allows us to permute the populations of world pairs, rather than just the utilities of agents in worlds. In Chapter 3 I show that if we accept Pareto, the qualitativeness of $\succcurlyeq$, the Permutation Principle, and transitivity, then we can show that many infinite world pairs are incomparable. I identified novel properties of utility distributions – the existence of bidirectional upgrades between worlds, for example – that entail that any theory that can weakly rank such worlds will violate transitivity. My results go beyond the results within the intergenerational equity literature in three important respects.[269]

First, my results show not only that a complete, transitive ordering of infinite worlds requires the use of non-constructive mathematics to be explicitly described. My results show that if we accept the four axioms listed above then such a complete, transitive ordering of infinite worlds does not exist. Either we must reject one of these axioms or we must give up on the project of finding a complete and transitive ordering of infinite worlds.

Second, because I appeal to the qualitativeness of $\succcurlyeq$ rather than to anonymity principles, my results rule out the possibility of there being any method of ranking infinite worlds with the utility distributions that I describe in this Chapter 3. We can contrast this with Zame's results, for example. Although Zame shows that no social welfare ordering that supervenes only on the utilities of agents can produce a strict ranking over most infinite worlds, his results do not conflict with theories like Expansionism, which appeal to the qualitative properties of infinite worlds – in the case of Expansionism, the distribution of utility over spatiotemporal regions – to produce a ranking of those worlds. By appealing to the qualitativeness of $\succcurlyeq$, my results show not only that can we not construct a complete ranking of infinite worlds based solely on the utilities of agents in those worlds, but that we cannot construct a complete ranking of these worlds by appealing to any other qualitative properties of these worlds.

Third, my results are sensitive to the importance of the identities of agents across worlds. I establish that my results hold not only for identical population worlds: I also extend these results to disjoint and overlapping population worlds by formulating conditions for incomparability that are sensitive to the identities of agents across worlds. This represents a large step forward in the literature on infinite ethics, where disjoint and overlapping population worlds have generally been unexplored, despite the fact that most physically plausible worlds

[269]Unlike many of the results in the intergenerational equity literature, my results also allow for agents to have any finite utility levels rather than utility 1 or 0, or utilities in the [0,1] interval. Such restrictions of possible utility levels are not universal however, and so I do not include this in the list of major contributions.

will have populations that have few agents in common. Worlds that contain infinitely many agents that have identical populations are highly idealized, and the results in infinite ethics should be extended beyond this idealized set of worlds.

In the final two chapters of this thesis I built upon the results of the first three chapters. In Chapter 4 I demonstrated that the results of this thesis can be extended to more infinite world pairs if we introduce further axioms from both ethics and population ethics. For example, I demonstrate that if we accept the Neutral Addition principle and Weak Catching-Up then we are able to both expand the set of worlds that are comparable as well as the set of worlds that can be shown to be incomparable by the four world or cyclic arguments of Chapter 3.

Finally, in the second half of Chapter 5 I considered the impact that accepting that many infinite worlds are incomparable will have on the ethics of objective and subjective permissibility. The impact of infinite ethics on objective and subjective permissibility has been surprisingly underexplored.[270] In this final chapter I show that the principles that generated the incomparability results of this thesis have permissibility analogs that generate important problems for both objective and subjective permissibility. Although infinite ethics has often been the focus of aggregative ethicists, I show that the problems that infinite worlds generate for objective and subjective permissibility are not restricted to aggregative ethical theories. All ethicists must deny at least one of the plausible subjective variants of the principles I have formulated in this thesis. I believe that there is much room for future work on the puzzles for permissibility that I formulate in this chapter.

This thesis offers important improvements the framework that we ought to use within infinite ethics. I have produced results that are sensitive to the identities of agents and to the

[270] Beckstead [26] discusses the 'fanaticism problem'. Others, like Christiano [57] and Arntzenius [6] attempt to construct subjective ethical theories for agents that are uncertain about whether the world is infinite. But the impact of conflicts between principles in infinite ethics on subjective decision-making has not been the subject of much discussion prior to this thesis.

possibility of using qualitative properties other than agent utilities to rank infinite worlds. I believe that these aspects of my framework ought to be replicated by future theorists working in this area. I have also demonstrated important new results using new axioms that I believe represent significant improvements on those appealed to within the literature.

Prior to this thesis, those writing in the literature on intergenerational equity demonstrated that a complete, transitive social welfare ordering that supervenes only on the utilities of each generation and satisfies Generational Sensitivity and Finite Generational Anonymity, then it cannot provide a mapping from infinite worlds onto the real numbers. They also established that any complete, transitive social welfare that supervenes only on the utilities of each generation and satisfies Finite Generational Anonymity produces a strict ranking of very few possible infinite streams of generations.

I have demonstrated that if we accept Pareto, the claim that if $\succcurlyeq$ holds between a world pair then it also holds between any qualitative duplicate world pair, and the claim that we can permute the populations of worlds[271] then it is not possible to generate a complete, transitive ranking on the basis of any qualitative properties of infinite worlds. These results extend to worlds with identical, disjoint, and overlapping populations. Finally, they generate puzzles for both objective and subjective permissibility that apply to both aggregative and non-aggregative ethical theories. I believe these constitute significant steps beyond the prior literature on the problems that infinite worlds pose within ethics.

[271] This is an alternative principle to the principle implicit in the intergenerational equity literature: namely that it is possible to permute the utilities of finitely or finitely many generations in any manner.

Bibliography

[1] Intelligent agent foundations forum — infinite ethics comparisons. https://agentfoundations.org/item?id=1455. (Accessed on 03/26/2018).

[2] R Adam, PAR Ade, N Aghanim, Y Akrami, MIR Alves, F Argüeso, M Arnaud, F Arroja, M Ashdown, J Aumont, et al. Planck 2015 results-i. overview of products and scientific results. *Astronomy & Astrophysics*, 594:A1, 2016.

[3] Anthony Aguirre. Eternal inflation, past and future. *arXiv preprint arXiv:0712.0571*, 2007.

[4] Andreas Albrecht and Paul J Steinhardt. Cosmology for grand unified theories with radiatively induced symmetry breaking. *Physical Review Letters*, 48(17):1220, 1982.

[5] Alfred Archer. Moral obligation, self-interest and the transitivity problem. *Utilitas*, 28(4):441–464, 2016.

[6] Frank Arntzenius. Utilitarianism, decision theory and eternity. *Philosophical Perspectives*, 28(1):31–58, 2014.

[7] G B Asheim and S Zuber. A complete and strongly anonymous leximin relation on infinite streams. *Soc. Choice Welfare*, 2013.

[8] Geir B Asheim. Intergenerational equity. *Annu. Rev. Econom.*, 2(1):197–222, August 2010.

[9] Geir B Asheim and Kuntal Banerjee. Fixed-step anonymous overtaking and catching-up. *Int. J. Econ. Theory*, 6(1):149–165, 2010.

[10] Geir B Asheim, Wolfgang Buchholz, and Bertil Tungodden. Justifying sustainability. *Journal of environmental economics and management*, 41(3):252–268, 2001.

[11] Geir B Asheim, Claude d'Aspremont, and Kuntal Banerjee. Generalized time-invariant overtaking. *J. Math. Econ.*, 46(4):519–533, July 2010.

[12] Geir B Asheim and Bertil Tungodden. Resolving distributional conflicts between generations. *Econom. Theory*, 24(1):221–230, July 2004.

[13] Geir B Asheim and Bertil Tungodden. Resolving distributional conflicts between generations. In *Justifying, Characterizing and Indicating Sustainability*, Sustainability, Economics, and Natural Resources, pages 53–62. Springer, Dordrecht, 2007.

[14] Geir B Asheim and Stéphane Zuber. A complete and strongly anonymous leximin relation on infinite streams. *Soc. Choice Welfare*, 41(4):819–834, October 2013.

[15] Hiroshi Atsumi. Neoclassical growth and the efficient program of capital accumulation. *The Review of Economic Studies*, 32(2):127–136, 1965.

[16] Robert J Aumann and Lloyd S Shapley. Long-term competition – a game-theoretic analysis. In *Essays in Game Theory*, pages 1–15. Springer, 1994.

[17] Kuntal Banerjee. On the extension of the utilitarian and Suppes–Sen social welfare relations to infinite utility streams. *Soc. Choice Welfare*, 27(2):327–339, October 2006.

[18] P Bartha. Taking stock of infinite value: Pascal's wager and relative utilities. *Synthese*, 154(1):5–52, 2007.

[19] P Bartha Synthese and 2007. Taking stock of infinite value: Pascal's wager and relative utilities. *Springer*, 2007.

[20] R.G. Bartle and D.R. Sherbert. *Introduction to real analysis*. John Wiley & Sons Canada, Limited, 2000.

[21] Kaushik Basu and Tapan Mitra. Aggregating infinite utility streams with intergenerational equity: the impossibility of being paretian. *Econometrica*, 71(5):1557–1563, 2003.

[22] Kaushik Basu and Tapan Mitra. Aggregating infinite utility streams with intergenerational equity: The impossibility of being paretian. *Econometrica*, 71(5):1557–1563, 2003.

[23] Kaushik Basu and Tapan Mitra. Utilitarianism for infinite utility streams: A new welfare criterion and its axiomatic characterization. Technical report, 2003.

[24] Kaushik Basu and Tapan Mitra. Possibility theorems for equitably aggregating infinite utility streams. In *Intergenerational Equity and Sustainability*, International Economic Association Series, pages 69–84. Palgrave Macmillan, London, 2007.

[25] Kaushik Basu and Tapan Mitra. Utilitarianism for infinite utility streams: A new welfare criterion and its axiomatic characterization. *J. Econ. Theory*, 133(1):350–373, March 2007.

[26] Bicholas Beckstead. *On the overwhelming importance of shaping the far future*. PhD thesis, Rutgers, 2013.

[27] Vieri Benci, Emanuele Bottazzi, and Mauro Di Nasso. Elementary numerosity and measures. December 2012.

[28] Vieri Benci, Emanuele Bottazzi, and Mauro Di Nasso. Some applications of numerosities in measure theory. November 2014.

[29] Vieri Benci and Mauro Di Nasso. Numerosities of labelled sets: a new way of counting. *Advances in Mathematics*, 173(1):50–67, 2003.

[30] Vieri Benci and Mauro Di Nasso. Numerosities of labelled sets: a new way of counting. *Adv. Math.*, 173(1):50–67, January 2003.

[31] Vieri Benci, Mauro Di Nasso, and Marco Forti. An aristotelian notion of size. *Ann. Pure Appl. Logic*, 143(1):43–53, November 2006.

[32] Vieri Benci, Mauro Di Nasso, and Marco Forti. An aristotelian notion of size. *Annals of Pure and Applied Logic*, 143(1):43–53, 2006.

[33] CL Bennett, D Larson, JL Weiland, N Jarosik, G Hinshaw, N Odegard, KM Smith, RS Hill, B Gold, M Halpern, et al. Nine-year wilkinson microwave anisotropy probe (wmap) observations: final maps and results. *The Astrophysical Journal Supplement Series*, 208(2):20, 2013.

[34] Jeremy Bentham. *The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation*. Clarendon Press, 1996.

[35] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer, 1985.

[36] Selim Berker. Epistemic teleology and the separateness of propositions. *Philosophical Review*, 122(3):337–393, 2013.

[37] Charles Blackorby, Walter Bossert, and David Donaldson. Critical-level utilitarianism and the population-ethics dilemma. *Economics and Philosophy*, 13(2):197–230, 1997.

[38] Charles Blackorby, Walter Bossert, and David Donaldson. Variable-population extensions of social aggregation theorems. *Social Choice and Welfare*, 28(4):567–589, Jun 2007.

[39] Charles Blackorby, David Donaldson, and John A Weymark. Social choice with interpersonal utility comparisons: A diagrammatic introduction. *Int. Econ. Rev.*, 25(2):327–356, 1984.

[40] Walter Bossert, Yves Sprumont, and Kotaro Suzumura. Ordering infinite utility streams. *Journal of Economic Theory*, 135(1):579–589, 2007.

[41] Nick Bostrom. Are we living in a computer simulation? *The Philosophical Quarterly*, 53(211):243–255, 2003.

[42] Nick Bostrom. Infinite ethics. *Analysis and Metaphysics*, 10:9–59, 2011.

[43] Rachael Briggs and Daniel Nolan. Utility monsters for the fission age. *Pacific Philosophical Quarterly*, 96(3):392–407, 2015. PPQ4428.

[44] J. Broome. *Weighing Lives*. Oxford University Press, 2006.

[45] John Broome. *Weighing goods : equality, uncertainty, and time*. B. Blackwell, Oxford, [1. paperback edition], [reprint] edition, 1991.

[46] John Broome. *Ethics Out of Economics*. Cambridge University Press, 2009.

[47] Krister Bykvist. Moral uncertainty. *Philosophy Compass*, 12(3):e12408–n/a, 2017. e12408 PHCO-1006.R2.

[48] James Cain. Infinite utility. *Australas. J. Philos.*, 73(3):401–404, September 1995.

[49] Donald E Campbell. Impossibility theorems and infinite horizon planning. *Social Choice and Welfare*, 2(4):283–293, 1985.

[50] Donald Edward Campbell. Impossibility theorems and infinite horizon planning. *Soc. Choice Welfare*, 2(4):283–293, 1985.

[51] Juan Carlos Candeal, Esteban Induráin, and José Ramón Uriarte. Some issues related to the topological aggregation of preferences. *Soc. Choice Welfare*, 9(3):213–227, July 1992.

[52] Christopher P Chambers. Intergenerational equity: sup, inf, lim sup, and lim inf. *Soc. Choice Welfare*, 32(2):243–252, 2009.

[53] Ruth Chang. The possibility of parity. *Ethics*, 112(4):659–688, 2002.

[54] Eddy Chen and Daniel Rubio. Surreal decisions. https://www.eddykemingchen.net/uploads/4/6/1/3/46137503/surreal_decisions_ppr.pdf, forthcoming. (Accessed on 04/03/2018).

[55] Graciela Chichilnisky. Social aggregation rules and continuity. *Q. J. Econ.*, 97(2):337–352, May 1982.

[56] Graciela Chichilnisky. An axiomatic approach to sustainable development. *Social choice and welfare*, 13(2):231–257, 1996.

[57] Paul Christiano. The absolute self-selection assumption - less wrong. http://lesswrong.com/lw/572/the_absolute_selfselection_assumption/. (Accessed on 03/07/2018).

[58] Sidney Coleman and Frank De Luccia. Gravitational effects on and of vacuum decay. *Physical Review D*, 21(12):3305, 1980.

[59] Mark Colyvan. No expectations. *Mind*, 115(459):695–702, July 2006.

[60] J.H. Conway. *On Numbers and Games*. Ak Peters Series. Taylor & Francis, 2000.

[61] Tyler Cook. Deontologists can be moderate. *The Journal of Value Inquiry*, Sep 2017.

[62] T. Cowan and D. Parfit. Against the social discount rate. In P. Laslett and J.S. Fishkin, editors, *Justice Between Age Groups and Generations*, Philosophy, politics, and society, pages 144–61. Yale University Press, 1992.

[63] Tyler Cowen. Resolving the repugnant conclusion. *The Repugnant Conclusion*, pages 81–97, 2004.

[64] Juan Alfonso Crespo, Carmelo Nuñez, and Juan Pablo Rincón-Zapatero. On the impossibility of representing infinite utility streams. *Economic Theory*, 40(1):47–56, 2009.

[65] Fernando G De Maio. Income inequality measures. *Journal of Epidemiology & Community Health*, 61(10):849–852, 2007.

[66] Gerard Debreu. Representation of a preference ordering by a numerical function. *Decision processes*, 3:159–165, 1954.

[67] David DeGrazia. *Human identity and bioethics*. Cambridge University Press, 2005.

[68] Peter A Diamond. The evaluation of infinite utility streams. *Econometrica*, 33(1):170–177, 1965.

[69] Jim Dolmas. Time-additive representations of preferences when consumption grows without bound. *Economics Letters*, 47(3-4):317–325, 1995.

[70] Cian Dorr. Propositions and counterpart theory. *Analysis*, 65(287):210–218, 2005.

[71] Cian Dorr and Frank Arntzenius. Self-locating priors and cosmological measures. In Khalil Chamcham, Joseph Silk, John D. Barrow, and SimonEditors Saunders, editors, *The Philosophy of Cosmology*, pages 396–428. Cambridge University Press, 2017.

[72] Dale Dorsey. The supererogatory, and how to accommodate it. *Utilitas*, 25(3):355–382, 2013.

[73] Ram Sewak Dubey. Fleurbaey–michel conjecture on equitable weak paretian social welfare order. *Journal of Mathematical Economics*, 47(4-5):434–439, 2011.

[74] Ram Sewak Dubey and Tapan Mitra. Combining monotonicity and strong equity: construction and representation of orders on infinite utility streams. *Social Choice and Welfare*, 43(3):591–602, 2014.

[75] Bhaskar Dutta. Some remarks on the ranking of infinite utility streams. In *Arguments for a Better World: Essays in Honor of Amartya Sen: Volume I: Ethics, Welfare, and Measurement*, page Ch. 8. Oxford University Press, 2008.

[76] Boris A Efimov and Gleb A Koshevoy. A topological approach to social choice with infinite populations. *Math. Soc. Sci.*, 27(2):145–157, April 1994.

[77] George FR Ellis and GB Brundrit. Life in the infinite universe. *Quarterly Journal of the Royal Astronomical Society*, 20:37–41, 1979.

[78] GFR Ellis and WR Stoeger. The evolution of our local cosmic domain: effective causal limits. *Monthly Notices of the Royal Astronomical Society*, 398(3):1527–1536, 2009.

[79] Larry G Epstein. Intergenerational preference orderings. *Soc. Choice Welfare*, 3(3):151–160, 1986.

[80] Delia Graff Fara. Dear haecceitism. *Erkenntnis*, 70(3):285–297, May 2009.

[81] Kit Fine. Varieties of necessity. In Tamar Szabo Gendler and John Hawthorne, editors, *Conceivability and Possibility*, pages 253–281. Oxford Up, 2002.

[82] Peter C Fishburn. A study of independence in multivariate utility theory. *Econometrica*, 37(1):107–121, 1969.

[83] Philip Fisher, Amy Gutmann, Judith Jarvis Thomson, Martha C Nussbaum, JB Schneewind, and Barbara Herrnstein Smith. Goodness and advice. 2009.

[84] D J Fishkind, B Hamkins, and Montero. New inconsistencies in infinite utilitarianism: Is every world good, bad or neutral? *Australas. J. Philos.*, 80(2):178, 2002.

[85] Donniell Fishkind, Joel David Hamkins, and Barbara Montero. New inconsistencies in infinite utilitarianism. *Australasian Journal of Philosophy*, 80(2):178–190, 2002.

[86] Marc Fleurbaey and Philippe Michel. Transfer principles and inequality aversion, with an application to optimal growth. *Math. Soc. Sci.*, 42(1):1–11, July 2001.

[87] Marc Fleurbaey and Philippe Michel. Intertemporal equity and the extension of the ramsey criterion. *J. Math. Econ.*, 39(7):777–802, September 2003.

[88] Philippa Foot. The problem of abortion and the doctrine of double effect. *Oxford Review*, 5:5–15, 1967.

[89] David Gale. On optimal development in a multi-sector economy. *The Review of Economic Studies*, 34(1):1–18, 1967.

[90] Daniel Garber. Old evidence and logical omniscience in bayesian confirmation theory. 1983.

[91] J L A Garcia and M T Nelson. The problem of endless joy: Is infinite utility too much for utilitarianism? *Utilitas*, 6(2):183–192, November 1994.

[92] Peter Gärdenfors and Nils-Eric Sahlin. Unreliable probabilities, risk taking, and decision making. *Synthese*, 53(3):361–386, 1982.

[93] Alan H Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23(2):347, 1981.

[94] Alan H Guth. Eternal inflation and its implications. *Journal of Physics A: Mathematical and Theoretical*, 40(25):6811, 2007.

[95] Alan Hájek. Dutch book arguments.

[96] Alan Hájek. Waging war on pascal's wager. *Philos. Rev.*, 112(1):27–56, 2003.

[97] Joel David Hamkins and Barbara Montero. Utilitarianism in infinite worlds. *Utilitas*, 12(1):91–96, March 2000.

[98] Joel David Hamkins and Barbara Montero. Utilitarianism in infinite worlds. *Utilitas*, 12(01):91–96, 2000.

[99] Joel David Hamkins and Barbara Montero. With infinite utility, more needn't be better. *Australas. J. Philos.*, 78(2):231–240, 2000.

[100] Joel David Hamkins and Barbara Montero. With infinite utility, more needn't be better. *Australasian Journal of Philosophy*, 78(2):231–240, 2000.

[101] Chiaki Hara, Tomoichi Shinotsuka, Kotaro Suzumura, and Yongsheng Xu. Continuity and egalitarianism in the evaluation of infinite utility streams. *Soc. Choice Welfare*, 31(2):179–191, August 2008.

[102] C. Hare. *The Limits of Kindness*. OUP Oxford, 2013.

[103] Caspar Hare. The ethics of morphing. *Philosophical Studies*, 145(1):111–130, 2009.

[104] Caspar Hare. Take the sugar. *Analysis*, 70(2):237–247, 2010.

[105] John C. Harsanyi. Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, 61(5):434–435, 1953.

[106] John C. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4):309–321, 1955.

[107] Brian Hedden. Options and the subjective ought. *Philosophical Studies*, 158(2):343–360, 2012.

[108] Mark Heller. The immorality of modal realism, or: How i learned to stop worrying and let the children drown. *Philosophical Studies*, 114(1-2):1–22, 2003.

[109] Michael Heller and W Hugh Woodin. *Infinity: New Research Frontiers*. Cambridge University Press, February 2011.

[110] A.E. Hurd and P.A. Loeb. *An Introduction to Nonstandard Real Analysis.* Pure and Applied Mathematics. Elsevier Science, 1985.

[111] T. Irwin. *Nicomachean Ethics (Second Edition):.* Hackett Classics Series. Hackett Publishing Company, 2000.

[112] R. Isaac. *The Pleasures of Probability.* Undergraduate Texts in Mathematics. Springer New York, 2013.

[113] Thomas Jech. *Set theory.* Springer Science & Business Media, 2013.

[114] Adam Jonsson. An axiomatic basis for blackwell optimality. January 2017.

[115] Adam Jonsson and Mark Voorneveld. The limit of discounted utilitarianism. *Theoretical Economics*, 13(1):19–37, 2018.

[116] Adam L Jonsson. The limit of discounted utilitarianism. August 2013.

[117] Y K. Ng. Should we be very cautious or extremely cautious on measures that may involve our destruction? *Soc. Choice Welfare*, 8:79–88, 1991.

[118] Shelly Kagan. Normative ethics. *Mind*, 109(434):373–377, 2000.

[119] Shelly Kagan. The costs of transitivity: Thoughts on larry temkin's rethinking the good. *journal of moral philosophy*, 12(4):462–478, 2015.

[120] K Kamaga and T Kojima. On the leximin and utilitarian overtaking criteria with extended anonymity. *Soc. Choice Welfare*, 2010.

[121] Kohei Kamaga and Takashi Kojima. On the leximin and utilitarian overtaking criteria with extended anonymity. 2008.

[122] Kohei Kamaga and Takashi Kojima. On the leximin and utilitarian overtaking criteria with extended anonymity. *Soc. Choice Welfare*, 35(3):377–392, September 2010.

[123] Frances Myrna Kamm. Supererogation and obligation. *Journal of Philosophy*, 82(3):118–138, 1985.

[124] I. Kant and J.W. Ellington. *Grounding for the Metaphysics of Morals: with On a Supposed Right to Lie because of Philanthropic Concerns.* Hackett Classics. Hackett Publishing Company, Incorporated, 1993.

[125] M. Kline. *Mathematics: The Loss of Certainty.* A Galaxy book. Oxford University Press, 1982.

[126] Tjalling C Koopmans. Stationary ordinal utility and impatience. *Econometrica*, 28(2):287–309, 1960.

[127] Tjalling C Koopmans. Representation of preference orderings over time. *Decision and organization*, 57:100, 1972.

[128] Christine M. Korsgaard. The right to lie: Kant on dealing with evil. *Philosophy and Public Affairs*, 15(4):325–349, 1986.

[129] C.M. Korsgaard, M. Gregor, and J. Timmermann. *Kant: Groundwork of the Metaphysics of Morals.* Cambridge Texts in the History of Philosophy. Cambridge University Press, 2012.

[130] Matthew H. Kramer. *Moral Realism as a Moral Doctrine.* Wiley-Blackwell, 2009.

[131] Saul A Kripke. Naming and necessity. In *Semantics of natural language*, pages 253–355. Springer, 1972.

[132] Luc Lauwers. Infinite chichilnisky rules. *Econ. Lett.*, 42(4):349–352, January 1993.

[133] Luc Lauwers. Time-neutrality and linearity. *J. Math. Econ.*, 24(4):347–351, January 1995.

[134] Luc Lauwers. Continuity and equity with infinite horizons. *Soc. Choice Welfare*, 14(2):345–356, April 1997.

[135] Luc Lauwers. Continuity and equity with infinite horizons. In *Topological Social Choice*, pages 199–210. Springer, 1997.

[136] Luc Lauwers. Infinite utility: Insisting on strong monotonicity. *Australas. J. Philos.*, 75(2):222–233, June 1997.

[137] Luc Lauwers. Infinite utility: Insisting on strong monotonicity. *Australas. J. Philos.*, 75(2):222–233, June 1997.

[138] Luc Lauwers. A note on weak ∞-Chichilnisky rules. In *Topological Social Choice*, pages 211–212. Springer, Berlin, Heidelberg, 1997.

[139] Luc Lauwers. Rawlsian equity and generalised utilitarianism with an infinite population. *Econ Theory*, 9(1):143–150, February 1997.

[140] Luc Lauwers. Intertemporal objective functions: Strong pareto versus anonymity. *Math. Soc. Sci.*, 35(1):37–55, January 1998.

[141] Luc Lauwers. A note on chichilnisky's social choice paradox. *Theory Decis.*, 52(3):261–266, May 2002.

[142] Luc Lauwers. Ordering infinite utility streams comes at the cost of a Non-Ramsey set. *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1360649*, February 2009.

[143] Luc Lauwers. Ordering infinite utility streams comes at the cost of a non-ramsey set. *J. Math. Econ.*, 46(1):32–37, January 2010.

[144] Luc Lauwers. Strong anonymity and infinite streams. 2011.

[145] Luc Lauwers. Intergenerational equity, efficiency, and constructibility. *Econ Theory*, 49(2):227–242, February 2012.

[146] Luc Lauwers. The axiomatic approach to the ranking of infinite streams. In *The Economics of the Global Environment*, Studies in Economic Theory, pages 231–255. Springer, Cham, 2016.

[147] Luc Lauwers. Intergenerational equity, efficiency, and constructibility. In *The Economics of the Global Environment*, Studies in Economic Theory, pages 191–206. Springer, Cham, 2016.

[148] Luc Lauwers and Peter Vallentyne. Infinite utilitarianism: more is always better. *Economics & Philosophy*, 20(2):307–330, 2004.

[149] Luc Lauwers and Luc Van Liedekerke. Ultraproducts and aggregation. *J. Math. Econ.*, 24(3):217–237, 1995.

[150] Stephanie Leary. Non-naturalism and normative necessities. *Oxford studies in metaethics*, 12:76–105, 2017.

[151] James Lenman. Consequentialism and cluelessness. *Philosophy & Public Affairs*, 29(4):342–370, 2000.

[152] D. Lewis. *On the Plurality of Worlds*. Wiley, 2001.

[153] David Lewis. Survival and identity. In Amelie Oksenberg Rorty, editor, *The Identities of Persons*, pages 17–40. University of California Press, 1976.

[154] Andrei Linde. Inflationary cosmology. In *Inflationary Cosmology*, pages 1–54. Springer, 2008.

[155] Andrei D Linde. A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108(6):389–393, 1982.

[156] Andrei D Linde. Eternal chaotic inflation. *Modern Physics Letters A*, 1(02):81–85, 1986.

[157] Tadeusz Litak. Infinite populations, choice and determinacy. pages 1–31, 06 2017.

[158] John Locke. *An essay concerning human understanding.* T. Tegg and Son, 1836.

[159] Jean-Pierre Luminet. The status of cosmic topology after planck data. *Universe*, 2(1):1, 2016.

[160] Jordi Massó. Undiscounted equilibrium payoffs of repeated games with a continuum of players. *J. Math. Econ.*, 22(3):243–264, January 1993.

[161] Tim Maudlin. *The metaphysics within physics.* Oxford University Press on Demand, 2007.

[162] Jeff McMahan. Causing people to exist and saving people's lives. *The Journal of ethics*, 17(1-2):5–35, 2013.

[163] J.S. Mill. *Utilitarianism.* Cosimo Classics, 2008.

[164] Tapan Mitra and Kaushik Basu. On the existence of paretian social welfare relations for infinite utility streams with extended anonymity. *Intergenerational equity and sustainability. Palgrave, London (forthcoming)*, 2006.

[165] Tapan Mitra and Kaushik Basu. On the existence of paretian social welfare Quasi-Orderings for infinite utility streams with extended anonymity. In *Intergenerational Equity and Sustainability*, International Economic Association Series, pages 85–99. Palgrave Macmillan, London, 2007.

[166] Tapan Mitra and Kaushik Basu. On the existence of paretian social welfare quasi-orderings for infinite utility streams with extended anonymity. In *IEA CONFERENCE VOLUME SERIES*, volume 143, page 85. Springer, 2007.

[167] Tim Mulgan. Future people: A moderate consequentialist account of our obligations to future generations. *OUP Catalogue*, 2008.

[168] Jacob M Nebel. The good, the bad, and the transitivity of better than. *Noûs*, April 2017.

[169] Mark T Nelson. Utilitarian eschatology. *Am. Philos. Q.*, 28(4):339–347, 1991.

[170] Mark T Nelson. Utilitarian eschatology. *Am. Philos. Q.*, 28(4):339–347, 1991.

[171] Yew-Kwang Ng. Welfarism and utilitarianism: A rehabilitation. *Utilitas*, 2(2):171–193, November 1990.

[172] Yew-Kwang Ng. Infinite utility and van liedekerke's impossibility: A solution. *Australasian Journal of Philosophy*, 73(3):408–412, 1995.

[173] Yew-Kwang Ng. Infinite utility and van liedekerke's impossibility: A solution. *Australas. J. Philos.*, 73(3):408–412, September 1995.

[174] Yew-Kwang Ng. Infinite utility and van liedekerke's impossibility: A solution. *Australas. J. Philos.*, 73(3):408–412, September 1995.

[175] Alastair Norcross. Comparing harms: Headaches and human lives. *Philos. Public Aff.*, 26(2):135–167, April 1997.

[176] Alastair Norcross. Good and bad actions. *Philos. Rev.*, 106(1):1–34, 1997.

[177] R. Nozick. *Anarchy, State, and Utopia*. Harper Torchbooks. Basic Books, 1974.

[178] Michael Otsuka. Prioritarianism and the measure of utility. *Journal of Political Philosophy*, 23(1):1–22, 2015.

[179] D. Parfit. *Reasons and Persons*. Oxford paperbacks. Clarendon Press, 1984.

[180] Derek Parfit. What we together do.

[181] Henrik Petri. Asymptotic properties of welfare relations. *Economic Theory*, pages 1–22, 2017.

[182] A.C. Pigou. *The Economics of Welfare*. Palgrave Classics in Economics. Palgrave Macmillan, 2013.

[183] DOUGLAS W. PORTMORE. Transitivity, moral latitude, and supererogation. *Utilitas*, 29(3):286–298, 2017.

[184] Alexander R. Pruss. Probability, regularity, and cardinality. *Philosophy of Science*, 80(2):231–240, 2013.

[185] Wlodek Rabinowicz. The size of inequality and its badness some reflections around temkin's inequality. *Theoria*, 69(1-2):60–84, 2003.

[186] WLODEK RABINOWICZ. The size of inequality and its badness some reflections around temkin's inequality. *Theoria*, 69(1-2):60–84, 2003.

[187] Stuart Rachels. Counterexamples to the transitivity of better than. *Australas. J. Philos.*, 76(1):71–83, March 1998.

[188] F P Ramsey. A mathematical theory of saving. *Econ. J. Nepal*, 38(152):543–559, 1928.

[189] John Rawls. *A theory of justice: Revised edition*. Harvard university press, 2009.

[190] Donald H. Regan. *Utilitarianism and Co-Operation*. Oxford University Press, 1980.

[191] Domain Restrictions and Weak Pareto axioms. Possibility theorems for aggregating infinite utility streams equitably.

[192] W. D. Ross. *The Right and the Good*. Clarendon Press, 1930.

[193] Tina Rulli and Alex Worsnip. Iia, rationality, and the individuation of options. *Philosophical Studies*, 173(1):205–221, Jan 2016.

[194] T Sakai. Intergenerational equity and an explicit construction of welfare criteria. 2008.

[195] T Sakai. Intergenerational equity and an explicit construction of welfare criteria. *Soc. Choice Welfare*, 2010.

[196] Toyotaka Sakai. A characterization and an impossibility of finite length anonymity for infinite generations. *J. Math. Econ.*, 46(5):877–883, September 2010.

[197] Toyotaka Sakai. Intergenerational equity and an explicit construction of welfare criteria. *Soc. Choice Welfare*, 35(3):393–414, Sep 2010.

[198] L.J. Savage. *The Foundations of Statistics*. Dover Books on Mathematics Series. Dover Publications, 1972.

[199] Samuel Scheffler. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Oxford University Press, 1994.

[200] Maik Schneider, Christian Traeger, and Ralph Winkler. Trading off generations: infinitely lived agent versus olg. 2012.

[201] Krister Segerberg. A neglected family of aggregation problems in ethics. *Noûs*, 10(2):221–244, 1976.

[202] Teddy Seidenfeld. A contrast between two decision rules for use with (convex) sets of probabilities: γ-maximin versus e-admissibilty. *Synthese*, 140(1-2):69–88, 2004.

[203] A. Sen. *Collective Choice and Social Welfare*. Mathematical Economics Texts Series. Holden-Day, 1970.

[204] Saharon Shelah. Cardinal arithmetic for skeptics. *American mathematical society*, 26(2):197–210, 1992.

[205] David Shoemaker. Personal identity and ethics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

[206] Robert K. Shope. The conditional fallacy in contemporary philosophy. *Journal of Philosophy*, 75(8):397–413, 1978.

[207] Theodore Sider. *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford University Press, 2001.

[208] H. Sidgwick and J. Rawls. *The Methods of Ethics*. Hackett Classics Series. Hackett, 1907.

[209] Peter Singer. All animals are equal. In Tom Regan and Peter Singer, editors, *Animal Rights and Human Obligations*, pages 215–226. Oxford University Press, 1989.

[210] J. J. C. Smart and Bernard Williams. *Utilitarianism: For and Against*. Cambridge University Press, 1973.

[211] J.J.C. Smart and B. Williams. *Utilitarianism: For and Against*. Cambridge University Press, 1973.

[212] Saul Smilansky. Can deontologists be moderate? *Utilitas*, 15(1):71–75, 2003.

[213] Justin Snedegar. *Contrastive Reasons*. Oxford University Press, 2017.

[214] Paul Joseph Steinhardt. Natural inflation. *The Very Early Universe*, page 251, 1982.

[215] P.J. Steinhardt and N. Turok. *Endless Universe: Beyond the Big Bang*. Crown/Archetype, 2007.

[216] Paul J Steinhaxdt and Neil Turok. The cyclic universe: An informal introduction. *Nuclear Physics B-Proceedings Supplements*, 124:38–49, 2003.

[217] GUNTER Stephan, Georg Müller-Fürstenberger, and PASCAL Previdoli. Overlapping generations or infinitely-lived agents: intergenerational altruism and the economics of global warming. *Environmental and Resource Economics*, 10(1):27–40, 1997.

[218] Patrick Suppes. Some formal models of grading principles. In *Studies in the Methodology and Foundations of Science*, pages 148–169. Springer, 1969.

[219] Lars-Gunnar Svensson. Equity among generations. *Econometrica*, 48(5):1251–1256, 1980.

[220] Lars-Gunnar Svensson. Equity among generations. *Econometrica*, 48(5):1251–1256, 1980.

[221] Max Tegmark. The interpretation of quantum mechanics: Many worlds or many words? *Fortschritte der Physik: Progress of Physics*, 46(6-8):855–862, 1998.

[222] Larry Temkin. Equality, priority, and the levelling-down objection. In Matthew Clayton and Andrew Williams, editors, *The Ideal of Equality*, pages 126–61. Macmillan, 2000.

[223] Larry S Temkin. Intransitivity and the mere addition paradox. *Philos. Public Aff.*, 16(2):138–187, 1987.

[224] Larry S Temkin. A continuum argument for intransitivity. *Philos. Public Aff.*, 25(3):175–210, July 1996.

[225] Larry S Temkin. Rationality with respect to people, places, and times. *Canadian Journal of Philosophy*, 45(5-6):576–608, 2015.

[226] L.S. Temkin. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford Ethics Series. Oxford University Press, USA, 2012.

[227] Bertil Tungodden. Egalitarianism: is leximin the only option? *Economics & Philosophy*, 16(2):229–245, 2000.

[228] Bertil Tungodden. Equality and priority. In Paul Anand, Prasanta Pattanaik, and Clemens Puppe, editors, *Handbook of Rational and Social Choice*. Oxford University Press, 2009.

[229] Bertil Tungodden and Peter Vallentyne. On the possibility of paretian egalitarianism. *The Journal of Philosophy*, 102(3):126–154, 2005.

[230] Peter Vallentyne. Utilitarianism and infinite utility. *Australas. J. Philos.*, 71(2):212–217, June 1993.

[231] Peter Vallentyne. Infinite utility and temporal neutrality. *Utilitas*, 6(2):193–199, November 1994.

[232] Peter Vallentyne. Infinite utility: Anonymity and person-centredness. *Australas. J. Philos.*, 73(3):413–420, September 1995.

[233] Peter Vallentyne and Shelly Kagan. Infinite value and finitely additive value theory. *J. Philos.*, 94(1):5–26, 1997.

[234] Peter Vallentyne and Shelly Kagan. Infinite value and finitely additive value theory. *J. Philos.*, 94(1):5–26, 1997.

[235] Peter Vallentyne and Luc Lauwers. Infinite utilitarianism: More is always better. October 2004.

[236] Luc Van Liedekerke. Should utilitarians be cautious aboutan infinite future? *Australas. J. Philos.*, 73(3):405–407, September 1995.

[237] Luc Van Liedekerke and Luc Lauwers. Sacrificing the patrol: Utilitarianism, future generations and infinity. *Economics & Philosophy*, 13(2):159–174, October 1997.

[238] Alexander Vilenkin. Birth of inflationary universes. *Physical Review D*, 27(12):2848, 1983.

[239] Carl Christian Von Weizsäcker. Existence of optimal programs of accumulation for an infinite time horizon. *The Review of Economic Studies*, 32(2):85–104, 1965.

[240] Alex Voorhoeve. Vaulting intuition: Temkin's critique of transitivity. *Economics & Philosophy*, 29(3):409–423, 2013.

[241] David Wallace. Decoherence and ontology. In S. Saunders, J. Barrett, A. Kent, and D. Wallace, editors, *Many Worlds?: Everett, Quantum Theory, & Reality*, chapter 1. OUP Oxford, 2010.

[242] Philippe Weil. Overlapping families of infinitely-lived agents. *Journal of Public Economics*, 38(2):183 – 198, 1989.

[243] Hermann Weyl. Mathematics and logic. *The American Mathematical Monthly*, 53(1):2–13, 1946.

[244] Hermann Weyl. Mathematics and logic. *Am. Math. Mon.*, 53(1):2–13, 1946.

[245] Yongsheng Xu. Pareto principle and intergenerational equity: Immediate impatience, universal indifference and impossibility. In *Intergenerational Equity and Sustainability*, pages 100–109. Springer, 2007.

[246] Richard Yetter Chappell. Philosophy, et cetera: Objective and subjective oughts. http://www.philosophyetc.net/2010/05/objective-and-subjective-oughts.html. (Accessed on 02/27/2018).

[247] William Zame. Can intergenerational equity be operationalized? *Theoretical Economics*, 2(2), June 2007.

[248] Stéphane Zuber and Geir B Asheim. Justifying social discounting: The rank-discounted utilitarian approach. *J. Econ. Theory*, 147(4):1572–1601, July 2012.

[249] Stéphane Zuber and Geir B Asheim. Justifying social discounting: The rank-discounted utilitarian approach. *J. Econ. Theory*, 147(4):1572–1601, July 2012.