Infinite Ignorance

Loren K. Fryxell*

August 26, 2024

(Latest Version Here)

Abstract

I consider the problem of normative decision making in a world of potentially infinite value across space and time when every feasible action causes a finite change in value. It is well-known that many normative theories break down when applied to such worlds. In particular, all aggregative consequentialist theories suffer from *infinite paralysis*: "If there is any positive probability the universe contains infinite moral value, then we should be morally indifferent among all our feasible actions." But many, if not all, aggregative consequentialist theories, including expected utility theory, are not grounded axiomatically in a framework which allows for infinitely good alternatives in the first place. Applying these theories in infinite worlds is applying them outside the scope in which they are grounded. I return to these foundations and construct a normative theory of decision making under risk and uncertainty in worlds which may contain infinitely good and bad alternatives. This approach uncovers a positive result, reversing the statement of infinite paralysis: "If there is any positive probability that the universe contains finite moral value, then we should evaluate our feasible actions conditional on the universe containing finite moral value." As this prescription does not require that we actually know how to rank infinite worlds, I call this *infinite ignorance*.

*Department of Economics and Global Priorities Institute, University of Oxford (e-mail: loren.fryxell@economics.ox.ac.uk).

1 Introduction

Infinite ethics, or intergenerational equity, is the study of ethics, or normative decision making, in a universe which is potentially infinite.[1] In particular, infinite ethics takes seriously the possibility that there may be infinitely many happy and/or unhappy beings throughout space and/or time. Taking such infinities seriously is difficult and breaks a lot of standard intuitions. In this paper, I propose a pragmatic solution to the problem. I do not propose a complete theory for how to rank infinite worlds. Instead, I propose a theory for how to rank *feasible actions*, which give rise to probability distributions over finite and infinite worlds (Section 3), or more generally, to sets of finite and infinite worlds (Section 4).

Much of the existing literature has focused on ranking infinite worlds themselves or probability distributions over infinite worlds.[2] Of the literature which focuses on ranking feasible actions which give rise to the possibility of both finite and infinite worlds, a certain negative result stands out: all aggregative consequentialist theories suffer from *infinite paralysis*.

> **Infinite Paralysis.** Suppose every feasible action causes a finite change in value. If there is any positive probability the universe contains infinite moral value, then we should be morally indifferent among all our feasible actions (Bostrom, 2011).

The basic premise of this result is simple. If we

1. assign moral values to universes and
2. assign moral values to probability distributions over universes by their expected moral value,

then the moral value of any probability distribution which places positive probability on universes with infinite moral value (e.g., with infinitely many happy beings) is itself infinite or undefined.

I take a different approach. I posit that there is some, possibly incomplete, moral ranking $\succeq^*$ over actions which give rise to probability distributions over possible universes (Section 3), or more generally, to sets of possible universes

[1] In philosophy, this area is generally referred to as "infinite ethics". In economics, this is generally referred to as "intergenerational equity".

[2] See for instance Koopmans (1960), Diamond (1965), Basu and Mitra (2003), Basu and Mitra (2007), Jonsson and Voorneveld (2018), Askell (2018), Pivato (2022), Pivato (2023), and Wilkinson (2023).

(Section 4). We may not fully know or understand $\succeq^*$, but we can write down some properties we expect that it should satisfy, narrowing the space of possibilities considerably.

This line of reasoning uncovers a positive result. Indeed, this result effectively *reverses* the statement of infinite paralysis and, moreover, implies that when infinite moral value is on the table, we should *not* simply assign moral values to universes and take expectations. As this result implies that we do not need to rank infinite worlds themselves, I call this *infinite ignorance*.

> **Infinite Ignorance.** Suppose every feasible action causes a finite change in value. If there is any positive probability that the universe contains finite moral value, then we should evaluate our feasible actions conditional on the universe containing finite moral value.

The intuition for this result can be explained with a simple example, which I will use throughout both the informal and formal exposition. It turns out we can think about this problem with or without explicit probabilities (Sections 3 and 4, respectively). For simplicity, I will use explicit probabilities here.

Suppose there are two islands: Island $\mathscr{A}$ and Island $\mathscr{B}$. There are only people on the islands, and the people are either happy or unhappy. We have agreed that happy people additively contribute $+1$ moral value to the world, unhappy people additively contribute -1 moral value to the world, and our moral ranking should coincide with that of the total moral value. We are undecided about what to do if there are infinitely many happy and/or unhappy people.

We live on $\mathscr{A}$. It is finite in size, and so, in isolation, we know how to morally rank all possible states of affairs on the island. Island $\mathscr{B}$ is far away. Our actions have no consequences on the island. It is also of unknown size. In particular, it may be infinitely large with infinitely many happy and/or unhappy people. An action gives rise to a probability distribution over states of affairs across both islands. How should we morally rank such probability distributions?

First, suppose we were to consider a ranking which coincides with the expected moral value across the two islands. This leads to infinite paralysis. If there is any positive probability that there is infinite or undefined moral value on $\mathscr{B}$, then the expected moral value across the two islands is infinite or undefined.

Instead, let us take a step back and write down some properties that we believe the world satisfies and the moral ranking $\succeq^*$ ought to satisfy. Suppose

1. our actions have no effect on what happens in $\mathscr{B}$,
2. $\mathscr{B}$ is finite with positive probability,
3. if one state of affairs is strictly morally preferred to another when $\mathscr{B}$ is finite, then this preference is not flipped when $\mathscr{B}$ is infinite, and
4. $\succeq^*$ satisfies the independence axiom.

If these four properties are satisfied, then the moral ranking between any two actions coincides with the moral ranking between those actions conditional on the event that $\mathscr{B}$ is finite. In other words, all humanly possible actions should be evaluated conditional on the universe containing finite moral value. This is infinite ignorance.

The third property is what I call *cautious ignorance*. Let (a, b) denote a happy people on Island $\mathscr{A}$ and b happy people on Island $\mathscr{B}$. Cautious ignorance implies that if (3, 3) $\succ^*$ (2, 3), then it may be that (3, ∞) $\succ^*$ (2, ∞) or that (3, ∞) $\sim^*$ (2, ∞), but it should not be that (3, ∞) $\prec^*$ (2, ∞). If we are comfortable ranking what happens on $\mathscr{A}$ for some finite $\mathscr{B}$, then, when $\mathscr{B}$ is infinite, we should either have the same ranking or be morally indifferent.

The fourth property is the independence axiom from von Neumann-Morgenstern expected utility (von Neumann and Morgenstern, 1944). Suppose we must choose between actions L and R. With probability α, L brings about some finite world p—e.g., $(3, 3)$—and R brings about another finite world q—e.g., $(2, 3)$. With the remaining probability $1 - \alpha$, we have no choice in the matter and must accept some other possibly infinite world r—e.g., $(3, \infty)$. The independence axiom states that our moral preference between L and R should not depend on α or r. In particular, if $p \succ^* q$, then the distribution which brings about p with probability α and r with probability $1 - \alpha$ should also be strictly morally preferred to the distribution which brings about q with probability α and r with probability $1 - \alpha$.

The intuition, and indeed the complete proof, for infinite ignorance is simple. If the universe is finite, suppose eating cake is strictly better than not eating cake. If the universe is infinite, not eating cake must not be strictly better than eating cake by cautious ignorance. There is some probability α the universe is finite. Eating cake results in a strictly better universe than not eating cake with probability α and a weakly better universe with probability $1 - \alpha$, so by independence, eating cake is strictly better than not eating cake.

Finally, the analogy of the islands $\mathscr{A}$ and $\mathscr{B}$ can be interpreted in various ways. In general, $\mathscr{A}$ represents what our actions can influence, and $\mathscr{B}$ represents everything else—what occurs in the possibly infinite background. But more definitively, $\mathscr{A}$ can be taken to represent everything within our Hubble sphere[3] until the heat death of the universe, and $\mathscr{B}$ can be taken to represent everything else.[4]

The rest of the paper is organized as follows. Section 2 lays out a fundamental challenge for infinite ethics—why, in my view, many approaches in the literature may be starting off on the wrong foot. Section 3 presents the formal model and results with explicit probabilities. Section 4 presents the formal model and results without explicit probabilities. Note that Sections 3 and 4 contain exactly analogous expositions under two different frameworks (Definition/Axiom 1.n is analogous to Definition/Axiom 2.n) and should also be fully self-contained.

2 A Fundamental Challenge for Infinite Ethics

In this section, I present a fundamental challenge for infinite ethics. In the subsequent sections, I propose a solution which circumvents it along with the more widely known challenges within infinite ethics.

When thinking about how to make sense of moral value in an infinite universe, it is natural to conceptualize the universe as an infinite sequence of locations (e.g., blocks of spacetime or human lives), each of which has some moral value, and to ask how one ought to morally rank such objects. In particular, it is natural to reduce the question of infinite ethics to the question of how to rank infinite sequences of real numbers $\mathbb{R}^{\mathbb{N}}$.

Two natural axioms one might seek to impose are Pareto and anonymity. The weakest (least constraining) versions of such axioms are strong Pareto and finite anonymity. Strong Pareto says that, for any $x, y \in \mathbb{R}^{\mathbb{N}}$, if every element in x is strictly larger than in y, then $x \succeq y$. Finite anonymity states that, for any $x, y \in \mathbb{R}^{\mathbb{N}}$, if y is a finite permutation of x, then $x \sim y$. It turns out that

[3]The Hubble sphere is a spherical region of the universe surrounding an observer beyond which objects recede faster than the speed of light due to the expansion of the universe.

[4]Note that we may be uncertain about the radius of the Hubble sphere and when the heat death of the universe will occur. Moreover, these beliefs may, in principle, have unbounded support. This changes nothing about the model (except opening the door for pathological St. Petersburg style actions to potentially not satisfy Definitions 1.2 and 2.2 and hence for which the main result does not apply).

there exists a complete and transitive ordering $\succeq$ on $\mathbb{R}^{\mathbb{N}}$ which satisfies strong Pareto and and finite anonymity. However, it cannot be explicitly defined. For a thorough and fantastic review of this literature, see Asheim (2010).

Be that as it may, I believe that modeling universes as infinite sequences of value is itself problematic. In particular, many seemingly desirable axioms are dependent on precisely *how* we choose to represent the universe as an infinite sequence of locations. But how we choose to mathematically represent the physical world shouldn't affect how we ought to behave within it. To illustrate this phenomena, I present two examples in which universe A can be seen to Pareto dominate B given one representation and B can be seen to Pareto dominate A given another.[5]

Example 1 shows that the decision to segment the universe by lives or by spacetime influences which of two universes Pareto dominates the other.

Example 1 (Lives and Spacetime)**.** *Suppose universe A contains infinitely many individuals who each live for 100 years and whose life contains 4 units of value, and universe B contains infinitely many individuals who each live for 50 years and whose life contains 3 units of value.*[6]

If locations are individual lives, then A, represented by $(4, 4, \ldots)$, Pareto dominates B, represented by $(3, 3, \ldots)$. If locations are 50-year blocks of spacetime, then B, represented by $(3, 3, \ldots)$, Pareto dominates A, represented by $(2, 2, \ldots)$. □

Example 2 shows that even if we were to decide that it is morally correct to segment the universe by spacetime, how we segment the universe by spacetime influences which of two universes Pareto dominates the other.

Example 2 (Two Segmentations of Spacetime)**.** *Suppose locations are blocks of spacetime. Suppose under one segmentation of the universe into sequential blocks, we have that universe A is represented by $(1, 1, \ldots)$ and universe B is*

[5]Note that this is a distinct concept from anonymity. Anonymity fixes a single representation of the universe and says that, if a sequence y can be formed by moving value around from x, then $x \sim y$. For example, $(2, 1, 1, \ldots) \sim (1, 2, 1, 1, \ldots)$. Examples 1 and 2 show that two different representations of universes lead us to conclude that universe A Pareto dominates B in one representation and B Pareto dominates A in the other. Note that we only compare universes within the same representation.

[6]This captures the fact that one would prefer to live a 100-year life in universe A over a 50-year life in universe B (since $4 > 3$), but that one would prefer to live two 50-year lives in universe B over one 100-year life in universe A (since $3 + 3 > 4$).

represented by $(-1,-1,\ldots)$*. Hence, A Pareto dominates B. At every location in the universe, A contains more value than B.*

Suppose that each of the blocks in A can be split into two sub-blocks, the first with value 2 *and the second with value* -1*. Similarly, suppose that each of the blocks in B can be split into two sub-blocks, the first with value* -2 *and the second with value* 1*. So A is represented by* $((2,-1),(2,-1),\ldots)$ *and B is represented by* $((-2,1),(-2,1),\ldots)$*.*

Construct a new segmentation of the universe such that each location consists of one of the first sub-blocks (from one block of the former segmentation) and three of the second sub-blocks (across three different blocks of the former segmentation). A is then represented by $((-1,-1,-1,2),(-1,-1,-1,2),\ldots)$ *and B is represented by* $((-1,-1,-1,2),(-1,-1,-1,2),\ldots)$*. Reducing the blocks to their total value, we have that A is represented by* $(-1,-1,\ldots)$ *and B is represented by* $(1,1,\ldots)$*, so B Pareto dominates A. At every location in the universe, B contains more value than A.*[7] □

Nearly all papers in infinite ethics and intergenerational equity assume that the universe can be modeled by an infinite sequence of value and impose the Pareto axiom on that domain.[8] However, showing that a universe contains more value than another at every location (Pareto dominates by location) is arguably not sufficient to conclude that it is in fact morally preferred, since for a different set of locations, that universe contains *less* value at every location than the other (is Pareto dominated by location). If so, then such approaches may be a non-starter. In the following sections, I do not assume a universe can be represented by an infinite sequence of value nor do I assume Pareto.

3 Infinite Ignorance with Probabilities

Let Ω be a set of complete histories of the universe across all of spacetime (past and future). Decompose Ω into $\Omega = \mathbb{A} \times \mathbb{B}$, where $\mathbb{A}$ describes aspects of the universe that the actions under consideration can affect and $\mathbb{B}$ describes

[7]Note that this example does not rely on having both infinite positive and negative value. An example with only positive value is given by the following. In the first segmentation, $A = (4,4,\ldots) = ((3,1),(3,1),\ldots)$ and $B = (3,3,\ldots) = ((1,2),(1,2),\ldots)$. As before, construct a new segmentation such that each location consists of one of the first sub-blocks and three of the second sub-blocks. Then $A = ((1,1,1,3),(1,1,1,3),...) = (6,6,...)$ and $B = ((2,2,2,1),(2,2,2,1),\ldots) = (7,7,\ldots)$.

[8]See for instance Basu and Mitra (2003), Basu and Mitra (2007), Jonsson and Voorneveld (2018), Askell (2018), Pivato (2022), Pivato (2023), and Wilkinson (2023).

aspects of the universe that the actions under consideration cannot affect. Let $\mathcal{F}$, $\mathcal{A}$, and $\mathcal{B}$ be σ-algebras on Ω, $\mathbb{A}$, and $\mathbb{B}$, respectively, such that $A \in \mathcal{A}$ and $B \in \mathcal{B}$ implies $A \times B \in \mathcal{F}$. Let $\Delta\Omega$ be the set of all probability measures $\mathbb{P} : \mathcal{F} \to [0, 1]$ on Ω. For any $X, Y \subseteq \Omega$ and $p \in \Delta\Omega$ such that $p(Y) > 0$, let $p(X \mid Y) = p(X \cap Y)/p(Y)$.

An action gives rise to a probability measure $p \in \Delta\Omega$, which we call a prospect. We say that two prospects have no (ex ante) effect on $\mathbb{B}$ (relative to each other) if, for any $B \in \mathcal{B}$, the probability that B occurs is the same.

Definition 1.1. *Two prospects $p, q \in \Delta\Omega$* have no (ex ante) effect on $\mathbb{B}$ *if, for any $B \in \mathcal{B}$, $p(\mathbb{A} \times B) = q(\mathbb{A} \times B)$.*[9]

Let $\succeq$ be a transitive relation on $\Delta\Omega$ reflecting the (moral and potentially incomplete) preferences over $\Delta\Omega$ that a decision maker feels confident in. Let $\succeq^*$ be a transitive relation on $\Delta\Omega$ reflecting the (moral and potentially incomplete) idealized preferences over $\Delta\Omega$ that the decision maker aspires to but may presently be ignorant of.

Axiom 1.1 states that $\succeq$ is compatible with $\succeq^*$, i.e., any preference $\succeq$ that a decision maker feels confident in is aligned with her idealized preference $\succeq^*$.

Axiom 1.1 (Compatibility)**.** *For any $p, q \in \Delta\Omega$, $p \succ q \implies p \succ^* q$ and $p \sim q \implies p \sim^* q$.*

Let $\mathbb{B}_0 \in \mathcal{B}$ be a measurable set of aspects of the universe which the actions under consideration cannot affect such that, when conditioning on $\mathbb{A} \times \mathbb{B}_0$ (the set of histories consisting of any element of $\mathbb{A}$ and any element of $\mathbb{B}_0$), the decision maker can confidently rank prospects p and q. Let $\mathbb{B}_1 = \mathbb{B} \setminus \mathbb{B}_0$.

In the island example, $\mathbb{B}_0$ reflects any set of states of affairs on Island $\mathscr{B}$ such that, if the decision maker were to condition on $\mathbb{A} \times \mathbb{B}_0$, she could confidently morally rank prospects p and q. For example, $\mathbb{B}_0$ be might be the aspect that there are no people on $\mathscr{B}$ (or the aspect that there are no more than a million people on $\mathscr{B}$).

If this is true, we say that p and q are comparable given $\mathbb{B}_0$.

Definition 1.2. *For any $\mathbb{B}_0 \in \mathcal{B}$, two prospects $p, q \in \Delta\Omega$ are* comparable given $\mathbb{B}_0$ *if the decision maker can rank p and q conditional on $\mathbb{A} \times \mathbb{B}_0$, i.e.,*

[9] $\mathbb{A} \times B$ is the set of states in which B occurs.

$p(\cdot \mid \mathbb{A} \times \mathbb{B}_0) \succeq q(\cdot \mid \mathbb{A} \times \mathbb{B}_0)$ *or* $q(\cdot \mid \mathbb{A} \times \mathbb{B}_0) \succeq p(\cdot \mid \mathbb{A} \times \mathbb{B}_0)$.

We will say that a decision maker is cautiously ignorant about p and q given $\mathbb{B}_1$ if her idealized preference over p and q conditional on $\mathbb{A} \times \mathbb{B}_1$ is either the same as her preference over p and q conditional on $\mathbb{A} \times \mathbb{B}_0$ or indifference.

Definition 1.3. *Given any two prospects $p, q \in \Delta\Omega$ which are comparable given $\mathbb{B}_0$, a decision maker is* cautiously ignorant *about p and q given $\mathbb{B}_1$ if $p(\cdot \mid \mathbb{A} \times \mathbb{B}_0) \succeq q(\cdot \mid \mathbb{A} \times \mathbb{B}_0) \implies p(\cdot \mid \mathbb{A} \times \mathbb{B}_1) \succeq^* q(\cdot \mid \mathbb{A} \times \mathbb{B}_1)$.*

Cautious ignorance reflects the idea that if a decision maker is confident about how to morally rank two prospects p and q given $\mathbb{A} \times \mathbb{B}_0$ (which her actions have no effect upon), then although she may still be uncertain about how to rank p and q given $\mathbb{A} \times \mathbb{B}_1$ (which her actions also have no effect upon), she should be confident that her preference will not flip.

For example, let p represent an action which brings about 3 happy people on $\mathscr{A}$ and q represent an action which brings about 2 happy people on $\mathscr{A}$. In both cases, there is background uncertainty about whether there are 3 happy people ($\mathbb{B}_0$) or infinitely many happy people ($\mathbb{B}_1$) on $\mathscr{B}$, each occurring with .5 probability.[10] Let (a, b) denote a happy people on $\mathscr{A}$ and b happy people on $\mathscr{B}$. Then p is a prospect which assigns probability .5 to each of $(3, 3)$ and $(3, \infty)$, and q is a prospect which assigns probability .5 to each of $(2, 3)$ and $(2, \infty)$. Cautious ignorance states that if a decision maker feels confident ranking $(3, 3) \succ (2, 3)$, then it may be that $(3, \infty) \succ^* (2, \infty)$ or that $(3, \infty) \sim^* (2, \infty)$, but it should not be that $(3, \infty) \prec^* (2, \infty)$.

Note that, in principle, a decision maker may update about what she can causally influence ($\mathbb{A}$) based on what she cannot ($\mathbb{B}$). Suppose a decision maker is deciding whether or not to build a human colony on Mars. She might believe that the probability that this goes well is higher conditional on the universe having infinitely many beings living in unreachable galaxies than it is conditional on the universe having merely finitely many beings living in unreachable galaxies.[11] Cautious ignorance rules out that the difference in her posterior beliefs, if any, is significant enough to flip her original preference.

[10] Since p and q have no effect on $\mathscr{B}$, the background uncertainty about what happens on $\mathscr{B}$ is the same.

[11] For example, conditional on the universe having merely finitely many beings living in unreachable galaxies, building a colony on Mars might be worse ex ante than not ($p(\cdot \mid \mathbb{A} \times \mathbb{B}_0) \prec^* q(\cdot \mid \mathbb{A} \times \mathbb{B}_0)$), while, conditional on the universe having infinitely many beings living in unreachable galaxies, building a colony on Mars might be better ex ante than not ($p(\cdot \mid \mathbb{A} \times \mathbb{B}_1) \succ^* q(\cdot \mid \mathbb{A} \times \mathbb{B}_1)$).

Axiom 1.2 applies independence from classical expected utility theory to the idealized preference $\succeq^*$. For any $F \in \mathcal{F}$, let $\alpha p(F) + (1-\alpha) r(F) = (\alpha p + (1-\alpha) r)(F) \in \Delta\Omega$.

Axiom 1.2 (Independence)**.** *For any $p, q, r \in \Delta X$ and $\alpha \in (0, 1]$,*

$$p \succ^* q \implies \alpha p + (1-\alpha) r \succ^* \alpha q + (1-\alpha) r$$

and

$$p \sim^* q \implies \alpha p + (1-\alpha) r \sim^* \alpha q + (1-\alpha) r.$$

Suppose that you must choose between action L and R. With some probability α, your action matters and prospect p realizes if you chose L and prospect q realizes if you chose R. With the remaining probability $1-\alpha$, prospect r realizes regardless of which action you originally chose. The independence axiom states that whether you choose L or R should only depend on p and q, and not on α and r.

For example, p might be an apple for sure, q a banana for sure, α the probability that fruit is in stock, and r nothing for sure, which is given when no fruit is in stock. Independence states that you should choose L over R if and only if you prefer an apple to a banana, regardless of the probability $(1-\alpha)$ that no fruit is in stock and you will instead receive nothing at all (r).

We are now ready to state the main result. If p and q have no effect on $\mathbb{B}$, p and q are comparable given $\mathbb{B}_0$, and the decision maker is cautiously ignorant about p and q given $\mathbb{B}_1$, then $p \succeq^* q$ if and only if $p(\cdot \mid \mathbb{A} \times \mathbb{B}_0) \succeq q(\cdot \mid \mathbb{A} \times \mathbb{B}_0)$. That is, when comparing p and q, it suffices to compare the simpler $p(\cdot \mid \mathbb{A} \times \mathbb{B}_0)$ vs $q(\cdot \mid \mathbb{A} \times \mathbb{B}_0)$. This is infinite ignorance.

Theorem 1. *Suppose Axioms 1.1 and 1.2 hold. For any $p, q \in \Delta\Omega$, if p and q have no effect on $\mathbb{B}$, p and q are comparable given $\mathbb{B}_0$, and the decision maker is cautiously ignorant about p and q given $\mathbb{B}_1$, then*

$$p \succeq^* q \iff p(\cdot \mid \mathbb{A} \times \mathbb{B}_0) \succeq q(\cdot \mid \mathbb{A} \times \mathbb{B}_0).$$

Proof. Since p and q have no effect on $\mathbb{B}$, $p(\mathbb{A} \times \mathbb{B}_0) = q(\mathbb{A} \times \mathbb{B}_0)$. Suppose $p(\cdot \mid \mathbb{A} \times \mathbb{B}_0) \succ q(\cdot \mid \mathbb{A} \times \mathbb{B}_0)$. Then $p(\cdot \mid \mathbb{A} \times \mathbb{B}_1) \succeq^* q(\cdot \mid \mathbb{A} \times \mathbb{B}_1)$ by cautious ignorance, and by Axioms 1.1 and 1.2,

$$\begin{aligned} p &= p(\mathbb{A} \times \mathbb{B}_0) p(\cdot \mid \mathbb{A} \times \mathbb{B}_0) + p(\mathbb{A} \times \mathbb{B}_1) p(\cdot \mid \mathbb{A} \times \mathbb{B}_1) \\ &\succeq^* p(\mathbb{A} \times \mathbb{B}_0) p(\cdot \mid \mathbb{A} \times \mathbb{B}_0) + p(\mathbb{A} \times \mathbb{B}_1) q(\cdot \mid \mathbb{A} \times \mathbb{B}_1) \\ &\succ^* q(\mathbb{A} \times \mathbb{B}_0) q(\cdot \mid \mathbb{A} \times \mathbb{B}_0) + q(\mathbb{A} \times \mathbb{B}_1) q(\cdot \mid \mathbb{A} \times \mathbb{B}_1) = q. \end{aligned}$$

Suppose $p(\cdot \mid \mathbb{A}\times\mathbb{B}_0) \sim q(\cdot \mid \mathbb{A}\times\mathbb{B}_0)$. Then $p(\cdot \mid \mathbb{A}\times\mathbb{B}_1) \sim^* q(\cdot \mid \mathbb{A}\times\mathbb{B}_1)$ by cautious ignorance, and by Axioms 1.1 and 1.2,

$$\begin{aligned} p &= p(\mathbb{A}\times\mathbb{B}_0)p(\cdot \mid \mathbb{A}\times\mathbb{B}_0) + p(\mathbb{A}\times\mathbb{B}_1)p(\cdot \mid \mathbb{A}\times\mathbb{B}_1) \\ &\sim^* p(\mathbb{A}\times\mathbb{B}_0)p(\cdot \mid \mathbb{A}\times\mathbb{B}_0) + p(\mathbb{A}\times\mathbb{B}_1)q(\cdot \mid \mathbb{A}\times\mathbb{B}_1) \\ &\sim^* q(\mathbb{A}\times\mathbb{B}_0)q(\cdot \mid \mathbb{A}\times\mathbb{B}_0) + q(\mathbb{A}\times\mathbb{B}_1)q(\cdot \mid \mathbb{A}\times\mathbb{B}_1) = q. \end{aligned}$$

Since p and q are comparable given $\mathbb{B}_0$, at least one of $p(\cdot \mid \mathbb{A}\times\mathbb{B}_0) \succ q(\cdot \mid \mathbb{A}\times\mathbb{B}_0)$, $p(\cdot \mid \mathbb{A}\times\mathbb{B}_0) \sim q(\cdot \mid \mathbb{A}\times\mathbb{B}_0)$, and $q(\cdot \mid \mathbb{A}\times\mathbb{B}_0) \succ p(\cdot \mid \mathbb{A}\times\mathbb{B}_0)$ is true, completing the proof. ■

Notice that the result holds if, instead of assuming that p and q have no effect on $\mathbb{B}$, we simply assume that p and q don't change the probability that the universe is finite ($p(\mathbb{A}\times\mathbb{B}_0) = q(\mathbb{A}\times\mathbb{B}_0)$).

4 Infinite Ignorance without Probabilities

Let Ω be a set of complete histories of the universe across all of spacetime (past and future). Decompose Ω into $\Omega = \mathbb{A}\times\mathbb{B}$, where $\mathbb{A}$ describes aspects of the universe that the actions under consideration can affect and $\mathbb{B}$ describes aspects of the universe that the actions under consideration cannot affect. Let $\mathcal{F}$, $\mathcal{A}$, and $\mathcal{B}$ be σ-algebras on Ω, $\mathbb{A}$, and $\mathbb{B}$, respectively, such that $A \in \mathcal{A}$ and $B \in \mathcal{B}$ implies $A \times B \in \mathcal{F}$.[12]

An action $E \in \mathcal{F}$ is a set of complete histories of the universe—namely, those histories in which the action was taken. In Section 3, we had a formal notion of objective probabilities. In this section, we have only an informal notion of "subjective likelihood". We say that two actions have no (ex ante) effect on $\mathbb{B}$ (relative to each other) if, for any $B \in \mathcal{B}$, the subjective likelihood that B occurs is the same whichever action we take.

Definition 2.1. *Two actions $E, F \in \mathcal{F}$* have no (ex ante) effect on $\mathbb{B}$ *if, for any $B \in \mathcal{B}$, the subjective likelihood of $E \cap (\mathbb{A}\times B)$ conditional on E is the same as the subjective likelihood of $F \cap (\mathbb{A}\times B)$ conditional on F.*[13]

Let $\succeq$ be a transitive relation on $\mathcal{F}$ reflecting the (moral and potentially incomplete) preferences over $\mathcal{F}$ that a decision maker feels confident in. Let

[12]This framework is known as the Bolker-Jeffrey framework, due to Bolker (1966, 1967) and Jeffrey (1983). See Broome (1990) for an excellent introduction and overview.

[13]$E \cap (\mathbb{A}\times B)$ is the set of states in E in which B occurs.

$\succeq^*$ be a transitive relation on $\mathcal{F}$ reflecting the (moral and potentially incomplete) idealized preferences over $\mathcal{F}$ that the decision maker aspires to but may presently be ignorant of.

Axiom 2.1 states that $\succeq$ is compatible with $\succeq^*$, i.e., any preference $\succeq$ that a decision maker feels confident in is aligned with her idealized preference $\succeq^*$.

Axiom 2.1 (Compatibility)**.** *For any $E, F \in \mathcal{F}$, $E \succ F \implies E \succ^* F$ and $E \sim F \implies E \sim^* F$.*

Let $\mathbb{B}_0 \in \mathcal{B}$ be a measurable set of aspects of the universe which the actions under consideration cannot affect such that, when conditioning on $\mathbb{A} \times \mathbb{B}_0$ (the set of histories consisting of any element of $\mathbb{A}$ and any element of $\mathbb{B}_0$), the decision maker can confidently rank actions E and F. Let $\mathbb{B}_1 = \mathbb{B} \setminus \mathbb{B}_0$.

In the island example, $\mathbb{B}_0$ reflects any set of states of affairs on Island $\mathscr{B}$ such that, if the decision maker were to condition on $\mathbb{A} \times \mathbb{B}_0$, she could confidently morally rank actions E and F. For example, $\mathbb{B}_0$ be might be the aspect that there are no people on $\mathscr{B}$ (or the aspect that there are no more than a million people on $\mathscr{B}$).

If this is true, we say that E and F are comparable given $\mathbb{B}_0$.

Definition 2.2. *For any $\mathbb{B}_0 \in \mathcal{B}$, two actions $E, F \in \mathcal{F}$ are* comparable given $\mathbb{B}_0$ *if the decision maker can rank E and F conditional on $\mathbb{A} \times \mathbb{B}_0$, i.e., $E \cap (\mathbb{A} \times \mathbb{B}_0) \succeq F \cap (\mathbb{A} \times \mathbb{B}_0)$ or $F \cap (\mathbb{A} \times \mathbb{B}_0) \succeq E \cap (\mathbb{A} \times \mathbb{B}_0)$.*

We will say that a decision maker is cautiously ignorant about E and F given $\mathbb{B}_1$ if her idealized preference over E and F conditional on $\mathbb{A} \times \mathbb{B}_1$ is either the same as her preference over E and F conditional on $\mathbb{A} \times \mathbb{B}_0$ or indifference.

Definition 2.3. *Given any two actions $E, F \in \mathcal{F}$ which are comparable given $\mathbb{B}_0$, a decision maker is* cautiously ignorant *about E and F given $\mathbb{B}_1$ if $E \cap (\mathbb{A} \times \mathbb{B}_0) \succeq F \cap (\mathbb{A} \times \mathbb{B}_0) \implies E \cap (\mathbb{A} \times \mathbb{B}_1) \succeq^* F \cap (\mathbb{A} \times \mathbb{B}_1)$.*

Cautious ignorance reflects the idea that if a decision maker is confident about how to morally rank two actions E and F given $\mathbb{A} \times \mathbb{B}_0$ (which her actions have no effect upon), then although she may still be uncertain about how to rank E and F given $\mathbb{A} \times \mathbb{B}_1$ (which her actions also have no effect upon), she should be confident that her preference will not flip.

For example, let E be an action which brings about 3 happy people on $\mathscr{A}$

and F be an action which brings about 2 happy people on $\mathscr{A}$. In both cases, there is subjective background uncertainty about whether there are 3 happy people ($\mathbb{B}_0$) or infinitely many happy people ($\mathbb{B}_1$) on $\mathscr{B}$.[14] Let (a, b) denote a happy people on $\mathscr{A}$ and b happy people on $\mathscr{B}$. Then $E = \{(3,3), (3,\infty)\}$ and $F = \{(2,3), (2,\infty)\}$. Cautious ignorance states that if a decision maker feels confident ranking $\{(3,3)\} \succ \{(2,3)\}$, then it may be that $\{(3,\infty)\} \succ^* \{(2,\infty)\}$ or that $\{(3,\infty)\} \sim^* \{(2,\infty)\}$, but it should not be that $\{(3,\infty)\} \prec^* \{(2,\infty)\}$.

Note that, in principle, a decision maker may update about what she can causally influence ($\mathbb{A}$) based on what she cannot ($\mathbb{B}$). Suppose a decision maker is deciding whether or not to build a human colony on Mars, which may either go well or not go well. She might believe that the subjective likelihood that this goes well is higher conditional on the universe having infinitely many beings living in unreachable galaxies than it is conditional on the universe having merely finitely many beings living in unreachable galaxies.[15] Cautious ignorance rules out that the difference in her subjective likelihoods, if any, is significant enough to flip her original preference.

Axiom 2.2 is a weaker but qualitatively similar counterpart to Axiom 1.2. It can be seen as an analog to the independence axiom within the von Neumann-Morgenstern framework (where the primitive is an ordering over probability distributions over states) for the Bolker-Jeffrey framework (where the primitive is an ordering over sets of states) applied to the idealized preference $\succeq^*$.

Axiom 2.2 (Independence). *For any two actions $E, F \in \mathcal{F}$ and any two partitions*[16] *$\{E_0, E_1\}$ and $\{F_0, F_1\}$ of E and F, respectively, if the subjective likelihood of E_0 conditional on E is the same as the subjective likelihood of F_0 conditional on F, then*

$$E_0 \succeq^* F_0 \textit{ and } E_1 \succeq^* F_1 \implies E \succeq^* F$$

and, if at least one of the relations in the antecedent is strict, $E \succ^* F$.[17]

[14]Since E and F have no effect on $\mathscr{B}$, the subjective background uncertainty about what happens on $\mathscr{B}$ is the same.

[15]For example, conditional on the universe having merely finitely many beings living in unreachable galaxies, building a colony on Mars might be worse ex ante than not ($E \cap (\mathbb{A} \times \mathbb{B}_0) \prec^* F \cap (\mathbb{A} \times \mathbb{B}_0)$), while, conditional on the universe having infinitely many beings living in unreachable galaxies, building a colony on Mars might be better ex ante than not ($E \cap (\mathbb{A} \times \mathbb{B}_1) \succ^* F \cap (\mathbb{A} \times \mathbb{B}_1)$).

[16]That is, $E_0 \cap E_1 = \emptyset$ and $E_0 \cup E_1 = E$.

[17]Note that, unlike in Section 3 where Axiom 1.2 is taken directly from von Neumann-

Suppose $E_0 \succeq F_0$. Now suppose additional possible outcomes E_1 are added to E_0 (E_0 and E_1 disjoint) and likewise additional possible outcomes F_1 are added to F_0 (F_0 and F_1 disjoint), and suppose that $E_1 \succeq F_1$. May we conclude that $E_0 \cup E_1 \succeq F_0 \cup F_1$? Not quite. Suppose you feel that E_1 is subjectively extremely likely relative to E_0, and F_1 is subjectively extremely unlikely relative to F_0. Then, intuitively, you may view $E_0 \cup E_1$ as relatively similar to E_1 and $F_0 \cup F_1$ as relatively similar to F_0, and it may well be that $E_1 \prec F_0$.

Axiom 2.2 says that *if* you consider the subjective likelihood of E_1 relative to E_0 equal to the subjective likelihood of F_1 relative to F_0, then, in fact, $E_0 \cup E_1 \succeq F_0 \cup F_1$. In other words, there are two qualities that may be relevant to your preference—what happens in E_1 and F_1 (i.e., how you morally rank E_1 and F_1) and how likely you perceive them to be conditional on taking each action ($E_0 \cup E_1$ and $F_0 \cup F_1$, respectively). Axiom 2.2 says that if, conditional on taking each action, the subjective likelihoods are the same, then what happens is all that matters.

Note that Axiom 2.2 states directly what Axiom 1.2 would imply if we could apply it. In particular, if we could construct an action E that results in E_0 with probability α and E_1 with probability $1-\alpha$ and an action F that results in F_0 with probability α and F_1 with probability $1-\alpha$, then Axiom 1.2 gives us that $E_0 \succeq F_0$ and $E_1 \succeq F_1$ implies $E \succeq F$. However, in this framework, we cannot simply assign probabilities to events. The events stand alone, and come with any subjective probabilities a decision maker may or may not associate with them. Axiom 2.2 assumes that *if* E_1 and F_1 have the same subjective likelihood relative to $E_0 \cup E_1$ and $E_0 \cup E_1$, respectively, then what we would have concluded from Axiom 1.2 if these subjective likelihoods were objective probabilities holds.

We are now ready to state the main result. If E and F have no effect on $\mathbb{B}$, E and F are comparable given $\mathbb{B}_0$, and the decision maker is cautiously ignorant about E and F given $\mathbb{B}_1$, then $E \succeq^* F$ if and only if $E \cap (\mathbb{A} \times \mathbb{B}_0) \succeq F \cap (\mathbb{A} \times \mathbb{B}_0)$. That is, when comparing E and F, it suffices to compare the simpler $E \cap (\mathbb{A} \times \mathbb{B}_0)$ vs $F \cap (\mathbb{A} \times \mathbb{B}_0)$. This is infinite ignorance.

Theorem 2. *Suppose Axioms 2.1 and 2.2 hold. For any $E, F \in \mathcal{F}$, if E and F have no effect on $\mathbb{B}$, E and F are comparable given $\mathbb{B}_0$, and the decision*

Morgenstern expected utility theory, Axiom 2.2 is not part of the Bolker-Jeffrey axiomatization.

maker is cautiously ignorant about E and F given $\mathbb{B}_1$, then

$$E \succeq^* F \iff E \cap (\mathbb{A} \times \mathbb{B}_0) \succeq F \cap (\mathbb{A} \times \mathbb{B}_0).$$

Proof. Since E and F have no effect on $\mathbb{B}$, the subjective likelihood of $E \cap (\mathbb{A} \times \mathbb{B}_0)$ conditional on E is the same as the subjective likelihood of $F \cap (\mathbb{A} \times \mathbb{B}_0)$ conditional on F. Suppose $E \cap (\mathbb{A} \times \mathbb{B}_0) \succ F \cap (\mathbb{A} \times \mathbb{B}_0)$. Then $E \cap (\mathbb{A} \times \mathbb{B}_1) \succeq^* F \cap (\mathbb{A} \times \mathbb{B}_1)$ by cautious ignorance, and by Axioms 2.1 and 2.2, $E \succ^* F$.

Suppose $E \cap (\mathbb{A} \times \mathbb{B}_0) \sim F \cap (\mathbb{A} \times \mathbb{B}_0)$. Then $E \cap (\mathbb{A} \times \mathbb{B}_1) \sim^* F \cap (\mathbb{A} \times \mathbb{B}_1)$ by cautious ignorance, and by Axioms 2.1 and 2.2, $E \sim^* F$.

Since E and F are comparable given $\mathbb{B}_0$, at least one of $E \cap (\mathbb{A} \times \mathbb{B}_0) \succ F \cap (\mathbb{A} \times \mathbb{B}_0)$, $E \cap (\mathbb{A} \times \mathbb{B}_0) \sim F \cap (\mathbb{A} \times \mathbb{B}_0)$, and $F \cap (\mathbb{A} \times \mathbb{B}_0) \succ E \cap (\mathbb{A} \times \mathbb{B}_0)$ is true, completing the proof. ■

Notice that the result holds if, instead of assuming that E and F have no effect on $\mathbb{B}$, we simply assume that E and F don't change the subjective likelihood that the universe is finite (that the subjective likelihood of $E \cap (\mathbb{A} \times \mathbb{B}_0)$ conditional on E is the same as the subjective likelihood of $F \cap (\mathbb{A} \times \mathbb{B}_0)$ conditional on F).

References

Asheim, Geir B. 2010. "Intergenerational Equity." *Annual Review of Economics*, 2: 197–222.

Askell, Amanda. 2018. "Pareto Principles in Infinite Ethics." PhD diss. New York University.

Basu, Kaushik, and Tapan Mitra. 2003. "Aggregating Infinite Utility Streams with Intergenerational Equity: The Impossibility of Being Paretian." *Econometrica*, 71(5): 1557–1563.

Basu, Kaushik, and Tapan Mitra. 2007. "Utilitarianism for infinite utility streams: A new welfare criterion and its axiomatic characterization." *Journal of Economic Theory*, 133(1): 350–373.

Bolker, Ethan D. 1966. "Functions Resembling Quotients of Measures." *Transactions of the American Mathematical Society*, 124(2): 292–312.

Bolker, Ethan D. 1967. "A Simultaneous Axiomatization of Utility and Subjective Probability." *Philosophy of Science*, 34(4): 333–340.

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

Broome, John. 1990. "Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism." *Review of Economic Studies*, 57(3): 477–502.

Diamond, Peter A. 1965. "The Evaluation of Infinite Utility Streams." *Econometrica*, 33(1): 170–177.

Jeffrey, Richard C. 1983. *The Logic of Decision.* . 2nd ed., The University of Chicago Press.

Jonsson, Adam, and Mark Voorneveld. 2018. "The limit of discounted utilitarianism." *Theoretical Economics*, 13: 19–37.

Koopmans, Tjalling C. 1960. "Stationary Ordinal Utility and Impatience." *Econometrica*, 28(2): 287–309.

Pivato, Marcus. 2022. "A characterization of Cesàro average utility." *Journal of Economic Theory*, 201: 1–26.

Pivato, Marcus. 2023. "Population ethics in an infinite universe." *Philosophical Studies*, 180: 3383–3414.

von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior.* . 1st ed., Princeton University Press.

Wilkinson, Hayden. 2023. "Infinite Aggregation and Risk." *Australasian Journal of Philosophy*, 101: 340–359.