Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 176 (2018) 444–478

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Planning for the long run: Programming with patient, Pareto responsive preferences ☆

Urmee Khan [a,*], Maxwell B. Stinchcombe [b]

[a] *Department of Economics, U. C. Riverside, United States*
[b] *Department of Economics, U. T. Austin, United States*

Received 21 April 2017; final version received 30 November 2017; accepted 1 April 2018
Available online 5 April 2018

Abstract

Respect for first order distributional overtaking guarantees that social welfare functions for intergenerational problems treat present and future people equally and respect the Pareto criterion, modulo null sets. For weakly ergodic optimization problems, this class of social welfare functions yields solutions that respect welfare concerns, sharply contrasting with extant patient criteria. For problems in which the evolution of future paths hinges on early events and decisions, the curvature of our social welfare functions determines the risks that society is willing to undertake and leads to a variant of the precautionary principle.

JEL classification: C6; D6; D8

Keywords: Intergenerational equity; Pareto responsiveness; Long-run optimality in stochastic dynamic problems

☆ Many thanks to: Aloisio Araujo for showing us the connections between this work and general equilibrium theory; to Svetlana Boyarchenko for many discussions about irreversible decisions; to Hector Chade, Amanda Friedenberg and Alejandro Manelli for ganging up and forcing one of the authors to abandon a silly insistence on radical pragmatism; to Paolo Piacquadio, Marcus Pivato, Federico Echenique, Chris Chambers and Matthew Jackson for early encouragement and illuminating conversations; to V. Bhaskar for asking just the right question; to seminar audiences at USC, ASU, UT Austin, Carlos III, Iowa, Kansas State, Bielefeld, Bonn and UC Irvine, and to participants at SSCW 2014, ERG Winter Workshop 2015, SAET 2016, Stony Brook Conference 2016, ASU Theory Conference 2016 and LA Theory Workshop 2017 for many helpful questions and discussions. Special thanks to Adam Jonsson for pointing out infelicities in our naming of various overtaking criteria. An anonymous referee provided highly detailed and insightful comments that have improved the paper.

* Corresponding author.
E-mail address: urmeek@ucr.edu (U. Khan).

https://doi.org/10.1016/j.jet.2018.04.001

> ... intergenerational solidarity is not optional, but rather a basic question of justice, since the world we have received also belongs to those who will follow us. (Pope Francis, 2015)

> As we peer into society's future, we — you and I, and our government — must avoid the impulse to live only for today, plundering for our own ease and convenience the precious resources of tomorrow. We cannot mortgage the material assets of our grandchildren without risking the loss also of their political and spiritual heritage. We want democracy to survive for all generations to come .... (Dwight David Eisenhower, Farewell Address, January 17, 1961)

1. Introduction

We study long-term intertemporal optimization for a society made up of an infinite number of generations. In this setting, current decisions can have lasting, even irreversible, future consequences. We develop a theoretical framework for analyzing decisions a society makes when current decisions generate externalities, positive or negative, for later generations. We aim to characterize optima that a patient society, as a whole, might choose.

To address such intertemporal problems with generational externalities, we need to first answer two questions.

(1) How patient does the society want to be while evaluating the effects of current actions? In other words, what is the society's view on intergenerational ethics?
(2) What kind of efficiency criteria could we use?

While each of the two questions separately has a ready-made answer, taken together they pose significant operational difficulties, or it would appear so. As regards to the first question, the idea of treating generations equally, conditional on the resource endowments, is central to the economic study of intergenerational social welfare. But modern formulations of intergenerational equity have come up against what seems to be an incompatibility with the ethical imperative of the Pareto criterion — the commonly espoused answer to the second question. We put forward a resolution to the apparent incompatibility of the two ethical criteria, and then connect the resulting patient, ethical intergenerational social welfare functions to the dynamic programming tools for long-run optimization, thereby offering a route to applications.

Our motivation stems from a set of observations about the state of the art, both substantive and technical, in the literature on long-run planning at the societal level. A significant part of the literature shares the premise that long-run planning must, of necessity, discount the well-being of future generations. The argument runs as follows: consumption foregone and well-invested grows at some rate r, and this leads to an inequitable $1 : \frac{1}{1+r}$ trade-off between the consumption of present and future people.

To the contrary, it appears to us that in a wide class of societal investment problems with intergenerational implications, arguments based on opportunity cost (as measured by the present marginal rates of return on investment) suffer from an important limitation: a neglect of both durability and non-exclusivity. Knowledge, once created and disseminated, is a durable and mostly non-exclusive good. A reliable understanding of causal structures, once found, is available for the use and benefit of each present person, and, for each of them, it is also available for roughly 10^7 additional future people.[1]

There may be a lag of several generations between breakthroughs in basic science (typically funded by the government for the benefit of society as a whole), and their embodiment in commercial products. Examples abound: more than a generation passed between the discovery of the physics behind touch screens and their extensive commercialization; genetic sequencing techniques began to yield benefits in about two generations; there was a three generation lag between theoretical understanding of relativity effects and their use in GPS systems; we do not yet know how long the lag will be between the initial understanding of quantum physics in the first half of the 20'th century and, say, quantum computing.

Suppose that we expect a knowledge intensive society, recognizably related to the present day, to continue to exist till the end of the current millennium. If knowledge produced today and used over this entire period is discounted to the present day at even the lowest of the currently suggested discount rates for public investment (Gollier and Hammitt, 2014, Table 2), the weight given to the benefits of that knowledge at the end of the period is 0 to five decimal places.

Discounting thus fails to capture the hugely disproportionate ratio of future to present people in a reasonably long-lived society, and the arguments for discounting in effect become arguments against resources being allocated to public goods that benefit the generations yet to be born. To capture this notion of each generation being infinitesimally small within the appropriate time horizon for a society, we use a limit model of infinite number of generations. We introduce a new social welfare criterion that we call *first order distributional overtaking* that captures patience in this setting, and examine the behavior of the resultant social welfare functions in standard economic models.

The next sections discuss several strands of literature we seek to unite and extend. There is a long history of economic thought about intergenerational equity, and an extensive set of results — both axiomatic and applied — on characterizing various welfare functions to evaluate the long-run value of infinite streams of measures of well-being. We cannot do full justice to the rich and varied literature; we focus on the works that have most influenced our thinking on the issues and provided the foundations that we build on.

1.1. Public goods for future people

Economists have long observed that impatience at the societal level, and the accompanying short planning horizons, can be detrimental to the future generations in many aspects. The idea of equal treatment of future generations entered mainstream economics with Pigou's early 1900's work on welfare (Pigou, 1912, 1920). Following Collard (1996), Pigou leaned on Sidgwick's

[1] From Asheim (2010),

> With 500 million years left of acceptable habitat for humans on Earth, population being stable at 10 billion with an average length of life equal to 73 years, the ratio of people who will potentially live in the future to people living now is approximately 10 million to 1.

utilitarianism from the mid- to late 1800's, and in particular the latter's discussion of intergenerational tradeoffs.[2]

Sidgwick was very aware of the free rider issue central to intergenerational allocation problems, and the public good aspects of natural resources and scientific research. He foresaw a failure of individualistic systems to adequately provide for the future.[3]

In the context of state policy with regards to natural resources, as early as the early 1700's, Sèbastien Le Prestre de Vauban, Louis XIV's defense minister, described several aspects of the economics and biology of forests that complicate the analysis of good societal practices.[4] He concluded that the only institutions that could, and should, undertake such projects in society's interest were the government and the church, on the presumption that these entities would have the requisite long planning horizon, as opposed to the impatience of private actors. Our project began as an attempt to build social welfare functions that capture the long planning horizons appropriate at the societal level.

Next we turn to formal treatments of intergenerational welfare. There are two main strands of literature, with some overlap in conceptual and technical aspects. The following sections discuss the two approaches, the common themes and problems, and put our results in context.

1.2. Various criteria for long-run social welfare

The first modern mathematical formulation of long-run intergenerational welfare was by Ramsey (1928). Working on a question posed by Pigou, Ramsey developed a theory that "does not discount later enjoyments in comparison with earlier ones, a practice which is ethically indefensible and arises merely from the weakness of the imagination." His approach posits a "bliss" point, or highest feasible utility, $\widehat{u}$, and then examines the necessary conditions for optimizing the criterion $\sum_{t=0}^{\infty}(u_t - \widehat{u})$. Unfortunately, these conditions are only necessary for paths along which the sum is finite. Chakravarty (1962) showed that, in simple examples, the divergence of the infi-

2 On the weighting of generations, Sidgwick wrote

> How far are we to consider the interests of posterity when they seem to conflict with those of existing human beings? It seems clear ... that the time at which a man exists cannot affect the value of his happiness from a universal point of view: and that the interests of posterity must concern a Utilitarian as much as those of his contemporaries, except in so far as the effect of his actions on posterity — and even the existence of human beings to be affected — must necessarily be more uncertain (Sidgwick, 1874, Book IV, Ch. I, §1).

3 Sidgwick on private investment in forests and scientific discoveries,

> ... the advantage (of forests) is one which private enterprise has no tendency to provide; since no one can appropriate and sell improvements in climate. For a somewhat different reason scientific discoveries, again, however ultimately profitable to industry, have not generally speaking a market value: the inventions in which the discovery is applied can be protected by patents; but the extent to which any given discovery will aid invention is mostly so uncertain, that even if the secret of a law of nature could be conveniently kept, it would not be worth an inventor's while to buy it, in the hope of being able to make something of it (Sidgwick, 1883, Book II, Ch. III, §3).

4 See *Traité de la Culture des Forêts* in de Vauban (1910) or de Vauban (2007). Vauban observed that being a free or easy access resource, forests were systematically over-exploited and rarely re-planted. After replanting, forests start being productive in about 100 years but do not become fully productive for 200 years. He also noted that no private enterprise can have so long and multi-generational a time horizon.

nite horizon sums can lead to the Ramsey's optimality equations being satisfied by feasible plans with minimal long-run utility.

To overcome such problems, Von Weizsäcker (1965) formulated the notion of overtaking optimality and gave sufficient conditions for the existence of optima in deterministic economic growth models (Brock, 1970, axiomatized these preferences). Overtaking is one of several heavily studied partial orders on streams of utilities, denoted $\boldsymbol{u} = (u_0, u_1, u_2, \ldots)$ and $\boldsymbol{v} = (v_0, v_1, v_2, \ldots)$.

- $\boldsymbol{u}$ **Pareto dominates** $\boldsymbol{v}$ if for all T, $(u_T - v_T) > 0$.
- $\boldsymbol{u}$ **catches up to** $\boldsymbol{v}$ if $\liminf_T \sum_{t=0}^{T}(u_t - v_t) \geq 0$.
- $\boldsymbol{u}$ **overtakes** $\boldsymbol{v}$ if $\liminf_T \sum_{t=0}^{T}(u_t - v_t) > 0$.
- $\boldsymbol{u}$ **overtakes** $\boldsymbol{v}$ **on average** if $\liminf_T \frac{1}{T+1}\sum_{t=0}^{T}(u_t - v_t) > 0$.
- $\boldsymbol{u}$ has a higher **patient limit** than $\boldsymbol{v}$ if $\sum_{t=0}^{\infty} u_t\beta^t > \sum_{t=0}^{\infty} v_t\beta^t$ for all β in some interval $(\underline{\beta}, 1)$.

The above criteria capture a notion of patience in the form 'eventually, and then forever after.' They have been extensively used in growth theory, in particular to study the 'turnpike' properties of deterministic optimal paths. An early example is Gale (1967). The turnpike approach to growth theory was surveyed, unified and extended in McKenzie (1976). The first general existence result for overtaking optimality in convex dynamic programming problems was given by Brock and Haurie (1976). A criterion related to 'overtaking on average' was first studied for stochastic dynamic programming by Veinott (1966) and Denardo and Miller (1968).[5] They also studied its relation with patient limits, a concept that has become central to the analysis of repeated games.

The formal criterion we develop is a sharpening of 'overtaking on average'. For distributions p and q with bounded support in $\mathbb{R}$, p strictly first order stochastically dominates q if $\int f(r)\,dp(r) > \int f(r)\,dq(r)$ for all strictly increasing, continuously differentiable $f : \mathbb{R} \to \mathbb{R}$ with $f'(x)$ uniformly bounded away from 0. From this,

- $\boldsymbol{u}$ **first order overtakes** $\boldsymbol{v}$ **in distribution** if $\liminf_T \frac{1}{T+1}\sum_{t=0}^{T}[f(u_t) - f(v_t)] > 0$ for all continuously differentiable, strictly increasing $f : \mathbb{R} \to \mathbb{R}$ with $f'(x)$ uniformly bounded away from 0.

For any bounded sequence of utilities $\boldsymbol{u}$, the cdf $F(x|\boldsymbol{u}, T) := \frac{1}{T+1}\sum_{t=0}^{T} 1_{(-\infty,x]}(u_t)$ is the empirical distribution of the numbers $u_0, \ldots, u_T$ when they are viewed as a random sample. Essentially, first order overtaking in distribution requires that $F(\cdot|\boldsymbol{u}, T)$ strictly first order stochastically dominates $F(\cdot|\boldsymbol{v}, T)$ for large T.

We show that this criterion captures the two ethical precepts we are concerned with, and also has desirable properties in terms of representation via social welfare functions.

1.3. The axiomatics of patience, Pareto and the impossibility results

We now come to the discussion of an axiomatic foundation for patience, or equivalently, intergenerational equity, and the possibility of accommodating the Pareto criterion alongside.

[5] They studied a criterion known as 'average overtaking' in operations research literature; it is a slightly different criterion than the one defined above.

Central to the idea of equitable social preferences is an indifference to permutations of the names of those receiving benefits. Central to the idea of patient preferences is indifference to permutations in the arrival time of benefits. These ideas overlap in social welfare functions defined on streams of generational utilities. To capture the notion of equal treatment of generations, various degrees of permutation invariance have been introduced in ranking infinite streams. Social indifference between permuted sequences of measures of generational well-being has been called, varyingly, "equity," "weak anonymity," or "intergenerational neutrality." Attempts to find functional forms to represent social preferences consistent with such indifference, while also respecting the Pareto criterion, have foundered. There are two types of impossibility results: the non-existence of measurable functions representing preferences satisfying (a version of) the strong Pareto criterion and (versions of) intergenerational neutrality; and the non-measurability of the graph of any relation satisfying these two criteria. These results make it seem difficult to find solutions to optimization problems while simultaneously satisfying these two criteria.

Diamond (1965) showed that there is no sup norm continuous function on the space of sequences of utilities that is simultaneously strongly Pareto and indifferent, for every t, to swapping the utilities of the first and the t'th generations. Basu and Mitra (2003) show that the same is true if "continuous" is replaced by "measurable," Fleurbaey and Michel (2003), Bossert et al. (2007), and Basu and Mitra (2007) contain extensions and further results, both positive and negative, and Asheim (2010) provides an extensive review of this literature.

Our work is closely related to Marinacci (1998), that axiomatizes patience in two ways. The first is indifference between a utility stream $\boldsymbol{u} = (u_0, u_1, u_2, u_3, \ldots)$ and any finite shift permutation $\boldsymbol{u}^F := (u_F, u_{F+1}, u_{F+2}, u_{F+3}, \ldots)$. It is informative to see how this captures patience and how it conflicts with the classical Pareto criterion in the presence of continuity. Suppose that $\boldsymbol{u}$ is a sequence with $4 \leq u_t \leq 7$ for all t. The stream $\boldsymbol{u}$ is the F-shifted stream of, hence is indifferent to, either

$$\boldsymbol{u}(0, F) := (\underbrace{0, 0, \ldots, 0}_{F \text{ times}}, u_0, u_1, u_2, u_3, \ldots) \text{ or} \tag{1}$$

$$\boldsymbol{u}(9, F) := (\underbrace{9, 9, \ldots, 9}_{F \text{ times}}, u_0, u_1, u_2, u_3, \ldots). \tag{2}$$

The indifference between $\boldsymbol{u}(0, F)$ and $\boldsymbol{u}$ captures patience as an immunity to having to wait for rewards. The indifference between $\boldsymbol{u}(9, F)$ and $\boldsymbol{u}$ captures a social willingness to ignore benefits accruing to a finite subset of an infinite population.

As to the conflict with the Pareto criterion, consider $\boldsymbol{r} = (r_0, r_1, \ldots)$ with $r_t \downarrow 0$ and compare $\boldsymbol{u} + \boldsymbol{r}$ to $\boldsymbol{u}$ assuming that preferences are represented by a (uniformly) continuous $S(\cdot)$. We have

$$|S(\boldsymbol{u}) - S(\boldsymbol{u}+\boldsymbol{r})| \leq \underbrace{|S(\boldsymbol{u}) - S(\boldsymbol{u}^F)|}_{=0} + \underbrace{|S(\boldsymbol{u}^F) - S((\boldsymbol{u}+\boldsymbol{r})^F)|}_{\to 0} + \underbrace{|S((\boldsymbol{u}+\boldsymbol{r})^F) - S(\boldsymbol{u}+\boldsymbol{r})|}_{=0}. \tag{3}$$

The "$= 0$" conclusions follow from indifference to finite shifts, and the "$\to 0$" conclusion follows from continuity and $d(\boldsymbol{u}^F, (\boldsymbol{u}+\boldsymbol{r})^F) \to 0$ as $F \uparrow \infty$. In other words, $S(\cdot)$ would not capture such improvements.

Numerical representations of preferences are not the only way to proceed. One could, in principle at least, dispense with them if the graph of the preference ordering had properties useful for optimization. For example, there is no numerical representation for lexicographic preferences on $\mathbb{R}^2_+$, but there is no difficulty in finding the optimal consumption bundle in any compact budget set. Zame (2007) shows that such an approach will not work for intergeneration problems. If $[0, 1]^{\mathbb{N}_0}$ has the natural (product Lebesgue) measure, then any preference relation that respects the Pareto criterion and is indifferent to finite permutations has a graph with inner measure 0 and outer measure 1.

1.4. Our answer to the impossibility results

At the heart of the issue is the measure attached to different-sized sets of generations. The mathematics of intergenerational equity treats the generations as points in a non-atomic measure space. Non-atomic population models contain many null coalitions[6] with infinite cardinality.

In such models, the Pareto optimality of an allocation cannot be over-turned by the existence of possible improvements for null coalitions (Hildenbrand, 1969, Definition 1). In this setting, intergenerational equity (or patience), formalized as permutation invariance, imply that the measure representing the agents is purely finitely additive, and this is the crux of the matter. A measure, η, on $\mathbb{N}_0$ is purely finitely additive if and only if, for all $\boldsymbol{r}$ with $r_t \downarrow 0$, $\int r_t \, d\eta(t) = 0$.[7] Interpreting $\boldsymbol{r}$ as a change to an allocation $\boldsymbol{u}$, the non-negative $\boldsymbol{r}$ delivers, on average, 0 to any coalition B — B's utility in the allocation $\boldsymbol{u} + \boldsymbol{r}$ is $\int_B (u_t + r_t)\, d\eta(t)$, but $\int_B r_t \, d\eta(t) = 0$, which implies that $\int_B (u_t + r_t)\, d\eta(t) = \int_B u_t \, d\eta(t)$.

The apparent irreconcilability of patience and Pareto arises precisely because adding such a mean 0, positive function to an allocation is interpreted as a Pareto improvement. But the function integrating to 0 implies that for any positive ϵ, the part of society receiving ϵ or more extra consumption is a null set in the relevant finitely additive measure. The Pareto improvement interpretation *fails* in a substantive way — only null sets are receiving positive amounts.

It is the observation that purely finitely additive measures and the associated null sets play a key role in the patience versus Pareto conundrum that leads us to our version of the overtaking criterion. We show that respect for first order distributional overtaking guarantees that social welfare functions for intergenerational problems treat present and future people equally and respect the Pareto criterion, *modulo null sets*. The definition of first order distributional overtaking also implicitly contains our definition of null and substantial coalitions: a coalition N is null if changing an allocation $\boldsymbol{u}$ for the generations in N does not effect the set of $\boldsymbol{v}$ that $\boldsymbol{u}$ first order overtakes; a coalition B is substantial if changing the allocation $\boldsymbol{u}$ for the generations in B can change the set of $\boldsymbol{v}$ that $\boldsymbol{u}$ first order overtakes. A social welfare function is *Pareto responsive* if it pays attention to boons to substantial coalitions and ignores boons to null coalitions. We present three main results. The first one shows that patient, Pareto responsive social welfare functions exist. The second and third, respectively, contain information about the tangents to our social welfare functions and to those previously studied in the literature.

---

[6] We use the word 'coalition' to denote sets of generations. This usage is without any strategic connotations, and is in line with the usage in Hildenbrand (1969).

[7] The more frequently used definition of a probability η being purely finitely additive is the existence of a countable, measurable partition into null sets, $\{E_n : n \in \mathbb{N}\}$ with $\eta(E_n) \equiv 0$. If $\{E_n : n \in \mathbb{N}\}$ is a countable partition into null sets and $r_n \downarrow 0$, then $f := \sum_n r_n 1_{E_n}$ is strictly positive and $\int f \, d\eta = 0$. If f is a strictly positive and $\int f \, d\eta = 0$, then $E_n := f^{-1}((r_n, r_{n-1}])$ with $r_0 := \infty$ gives a countable partition into null sets.

- Theorem A shows that a continuous, concave social welfare function that respects first order distributional overtaking is patient, i.e. invariant under a wide class of permutations, and Pareto responsive.
- Theorem B shows that, modulo a caveat about boundary points, a continuous, concave social welfare function is Pareto responsive if and only if its tangents are.
- Theorem C shows that the long-run social welfare functions previously studied in the literature evaluate each stream, $\boldsymbol{u}$, as the minimum of $L(\boldsymbol{u})$ where L belongs to a compact convex set of positive, norm 1, patient linear functionals.

Theorems B and C contrast our social welfare functions and previously studied long-run social welfare functions in terms of the properties of their tangents. Theorem B does not claim that the tangents to our patient social welfare functions are patient, indeed, in the interesting examples, they fail to be patient. This is in fact a useful feature, as Lemma 8 shows that having all tangents be patient forces a particular form of distributional risk neutrality on the social welfare function. In contrast, Theorem C shows that the previously studied long-run social welfare functions in fact have patient tangents.

In general, knowledge about the tangents is useful for the study of optima in applications.

1.5. Applications

We propose the first order distributional overtaking criterion as a way to capture patience, while respecting a reasonable version of the Pareto principle. But one cannot fully understand a class of preferences without knowing their implications in the analysis of problems of interest.[8] We therefore examine how our social welfare functions choose between actions when consequences may be irreversible, and then compare these results to the optima that arise for other forms of patient social welfare functions previously suggested in the literature. Respect for first order distributional overtaking can capture preferences for equity and risk aversion towards random shocks to future well-being in novel ways, and in problems with irreversibility, our class of welfare functions yields a version of the precautionary principle.

We also examine infinite horizon general equilibrium models with dynasties that have preferences respecting first order distributional overtaking. In this case, the first and the second welfare theorems of general equilibrium theory hold. Perhaps more interesting are the parallels between one-period risk sharing with different probability distributions and equilibrium paths with social welfare functions that value future coalitions in different ways.

In many dynamic programming problems, the concern is with the long-run average behavior of the system rather than the dynamics the precede the averages settling down. We investigate how our social welfare functions behave in Markov transition models with the following ingredients:

- the utility at any given time t is a function of the state of the system, x_t, and the present choice of present action, a_t;

[8] The motivation for Asheim (1988, 1991) is to examine the implications of equitable intergenerational preferences by examining their behavior in well-understood models. See also Atkinson (2001, p. 206), “By applying ethical criteria to concrete economic models, we learn about their consequences, and this may change our views about their attractiveness.”

- externalities flowing from the present to the future are encoded in two ways,
  - the set of available actions depends on the current state, $a_t \in A(x_t)$, and
  - the distribution of the future state, X_{t+1}, depends on the present state and present choice of actions.

The tangent characterization of the social welfare functions provides sharp results in strongly ergodic problems. A problem is strongly ergodic when each stationary policy gives rise to a unique ergodic distribution, i.e. a probability μ with the property that $\lim_T \frac{1}{T+1} 1_E(a_t, X_t) = \mu(E)$, and μ is independent of the initial state of the system. For strongly ergodic problems, maximizing the long-run average payoff guarantees catching up to any feasible path and the patient limit criteria is also optimized.[9] However, our social welfare functions allow for more substantive modes of trade offs between generational utilities.

1.6. Organization

The next section specifies the setting in which we work and states the main results. The subsequent section contains an extensive discussion of the results, both in general terms and in specific applications. The penultimate section covers two topics to illustrate the substantive implications of our results. First, a version of the precautionary principle arises from optimizing our social welfare functions while making irreversible decisions. Second, general equilibrium exchange models between dynasties with our class of preferences retain the classical welfare theorems. The last section provides a summary and a sketch of future work to be done.

The three Theorems are the main results. Propositions concern classes of preferences or models. Corollaries and Lemmas concern subsidiary information relevant to interpretations. Probabilities are finitely additive. If they are also countably additive, this will be explicitly noted. We have gathered the proofs in the appendix.

2. Patient, equitable, Pareto responsive social welfare functions

We first specify the formal setting, then motivate and describe the postulates that lead to our class of social welfare functions. The representation result and two results on tangent properties conclude this section.

2.1. The setting and the aim

Intergenerational streams of well-being are normalized to belong to **W**, the non-negative elements of ℓ_∞.[10] The streams of well-being, **W**, is equipped with the ℓ_∞ norm, $\|\boldsymbol{u}\| := \sup_{t\in\mathbb{N}_0} |u_t|$ where $\boldsymbol{u} = (u_0, u_1, \ldots) = (u_t)_{t\in\mathbb{N}_0}$ and $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$. The sup norm distance is $d(\boldsymbol{u}, \boldsymbol{v}) := \|\boldsymbol{u} - \boldsymbol{v}\|$. A stream $\boldsymbol{u}$ belongs to the interior of **W**, denoted $\boldsymbol{u} \in \text{int}(\mathbf{W})$, if and only if $\inf_t u_t \geq r$ for some strictly positive r.

[9] Assumptions on such problems sufficient to guarantee the existence of policies that maximize the expectation of $\lim_T \frac{1}{T+1} \sum_{t=0}^{T} u(a_t, X_t)$ have been extensively studied. Arapostathis et al. (1993) is a slightly dated survey of roughly two hundred contributions spanning three decades, Jaśkiewicz and Nowak (2006) and Feinberg et al. (2012) are extensions that contain short overviews of more recent work. Meyn and Tweedie (2012) is a comprehensive study of the stochastic stability of Markov chains with continuous, locally compact state spaces.

[10] See Blackorby et al. (1995) for a discussion of these normalizations.

The domain for preferences is the mixture set, $\mathcal{M}$, of countably additive Borel probabilities on $\mathbf{W}$ having norm bounded support, $(\exists B)[p(\{\boldsymbol{u} \in \mathbf{W} : \|\boldsymbol{u}\| \le B\}) = 1]$.[11] The set $\mathcal{M}$ is given the weak* topology, that is, $p^n \to p$ if and only if $\int f\, dp^n \to \int f\, dp$ for all norm continuous, bounded $f : \mathbf{W} \to \mathbb{R}$.

Our aim is to characterize social preferences that can be represented $p \succ q$ if and only if $\int_{\mathbf{W}} S(\boldsymbol{u})\, dp(\boldsymbol{u}) > \int_{\mathbf{W}} S(\boldsymbol{u})\, dq(\boldsymbol{u})$ where $S : \mathbf{W} \to [0, \infty)$ is norm continuous, concave, Pareto responsive, and patient.

2.2. Definitions

The following definitions will be used to state our postulates and results.

2.2.1. Pareto responsiveness

Coalitions are null if they form an asymptotically negligible portion of the generations, and they are substantial if they are not a vanishing portion of society at any large time horizon.

Definition 2.1. A coalition $N \subset \mathbb{N}_0$ is **null** if $\limsup_T \frac{1}{T+1} \sum_{t=0}^{T} 1_N(t) = 0$ and $B \subset \mathbb{N}_0$ is **substantial** $\liminf_T \frac{1}{T+1} \sum_{t=0}^{T} 1_B(t) > 0$.

Pareto responsive preferences ignore boons given to null coalitions and pay attention to boons given to substantial coalitions.

Definition 2.2. A function $S : \mathbf{W} \to \mathbb{R}$ is **Pareto responsive** if for all $\boldsymbol{u} \in \mathbf{W}$ and all $r > 0$, $S(\boldsymbol{u} + r1_B) > S(\boldsymbol{u})$ for all substantial $B \subset \mathbb{N}_0$ and $S(\boldsymbol{u} + r1_N) = S(\boldsymbol{u})$ for all null $N \subset \mathbb{N}_0$.

Comments. There are coalitions that are neither null nor substantial, and different social welfare functions can treat such coalitions differently. We specify a compact and convex set of continuous linear functionals, $\mathfrak{U}$, below. It has the property that N is null if and only if $L(1_N) = 0$ for all $L \in \mathfrak{U}$ while B is substantial if and only if $L(1_B) > 0$ for all $L \in \mathfrak{U}$. Pareto responsiveness implies monotonicity with respect to uniform increases because $\mathbb{N}_0$ is a substantial coalition, hence $S(\boldsymbol{u} + r1_{\mathbb{N}_0}) > S(\boldsymbol{u})$ for all $r > 0$. This in turn implies that indifference sets can have no interior as $d(\boldsymbol{u} + \epsilon 1_{\mathbb{N}_0}, \boldsymbol{u}) = \epsilon$.

2.2.2. Asymptotic first order dominance

Our patience postulate is respect for asymptotic first order overtaking of the distribution of utilities. Let $\mathbb{F}$ denote the set of strictly increasing, continuously differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ with derivative uniformly bounded away from 0. For distributions p and q on a bounded subset of $\mathbb{R}$, p first order dominates q if $\int f\, dp > \int f\, dq$ for all $f \in \mathbb{F}$.

Definition 2.3. ***u* first order overtakes *v***, denoted $\boldsymbol{u} \gg_{fo} \boldsymbol{v}$, if for all $f \in \mathbb{F}$,

$$\liminf_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} [f(u_t) - f(v_t)] > 0. \tag{4}$$

[11] This choice of domain facilitates the study of stochastic general equilibrium theory and more general dynamic intergenerational problems. An alternative domain is the set of sequences of distributions of outcomes at t as in Marinacci's (1998) axiomatic treatment of patience. This is less suitable for the study of environments where e.g. different present choices induce different distributions of the random time until a change of state.

There is no level of discounting that respects asymptotic first order overtaking: consider $\boldsymbol{u}^F$ with $u_t^F = 1$ for $t = 0, 1, \ldots, F$ and $u_t^F = 0$ for $t > F$ and $\boldsymbol{v}^F$ that has $v_t^F = 0$ for $t = 0, 1, \ldots, F$ and $v_t^F = 2$ for $t > F$; for any $\beta < 1$, there exist F for which $\sum_t u_t^F \beta^t > \sum_t v_t^F \beta^t$; but for all F, $\boldsymbol{v}^F \gg_{fo} \boldsymbol{u}^F$.

The following will be used to show that respect for asymptotic first order dominance entails Pareto responsiveness.

Lemma 1. *For any substantial coalition B, any $r > 0$, and any $\boldsymbol{u} \in \mathbf{W}$, $(\boldsymbol{u} + r1_B) \gg_{fo} \boldsymbol{u}$. For any null coalition N: for any $r > 0$ and any $\epsilon > 0$, $(\boldsymbol{u} + 2\epsilon 1_{\mathbb{N}_0}) \gg_{fo} (\boldsymbol{u} + r1_N + \epsilon 1_{\mathbb{N}_0}) \gg_{fo} \boldsymbol{u}$; and if $\boldsymbol{u} \gg_{fo} \boldsymbol{v}$, then $\boldsymbol{u} \gg_{fo} (\boldsymbol{v} + r1_N)$.*

Respect for first order overtaking implies Pareto responsiveness, but Pareto responsiveness by itself does not entail indifference to the permutations used to define patience. That is a separate property.

2.2.3. Permutations and patience

For continuous functions $S : \mathbf{W} \to [0, \infty)$, respect for asymptotic first order dominance also implies invariance with respect to a wide class of permutations.

The set $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$ is denoted $\mathbb{Z}$. A **permutation** is a 1-to-1 function $\pi : \mathbb{N}_0 \to \mathbb{Z}$ that is *onto* $\mathbb{N}_0$. Given $\boldsymbol{u} = (u_0, u_1, u_2, \ldots) \in \ell_\infty$ and a permutation π, define $\boldsymbol{u}^\pi$ as $(u_{\pi^{-1}(0)}, u_{\pi^{-1}(1)}, u_{\pi^{-1}(2)}, \ldots)$. In increasing order of generality, we consider the following classes of permutations.

- π is a **shift** permutation if $\pi(T) = T - F$ for some integer F.
- π is a **bounded** permutation if, for all T, $|\pi(T) - T| \le F$ for some integer F.
- π is a **asymptotic** permutation[12] if $\lim_{T\to\infty} \frac{|\pi(T)-T|}{T} = 0$.

Definition 2.4. A function $S : \mathbf{W} \to \mathbb{R}$ is **asymptotically patient**, or just **patient**, if for all $\boldsymbol{u} \in \mathbf{W}$ and all asymptotic permutations π, $S(\boldsymbol{u}) = S(\boldsymbol{u}^\pi)$.

Continuous social welfare functions that respect asymptotic first order dominance must be patient: for any $\boldsymbol{u} \in \mathbf{W}$, any asymptotic π, and any $\epsilon > 0$, it is plausible (and we give a proof in Lemma 9(h)) that $(\boldsymbol{u} + 2\epsilon 1_{\mathbb{N}_0}) \gg_{fo} (\boldsymbol{u}^\pi + \epsilon 1_{\mathbb{N}_0}) \gg_{fo} \boldsymbol{u}$; sending ϵ to 0 and using continuity delivers $S(\boldsymbol{u}) \ge S(\boldsymbol{u}^\pi) \ge S(\boldsymbol{u})$.

2.2.4. Risk aversion and resultants

Apart from patience and Pareto responsiveness, risk aversion is our other key postulate. It stipulates that the expectation of a random variable in $\mathbf{W}$ is preferred to the random variable itself. A probability $p \in \mathcal{M}$ has a **resultant** (or **barycenter** or Pettis **expectation**), $\boldsymbol{r}(p)$, defined as the point $\boldsymbol{r} \in \mathbf{W}$ that satisfies $\int \left[\sum_t v_t y_t\right] dp(\boldsymbol{v}) = \sum_t \boldsymbol{r}_t y_t$ for all $\boldsymbol{y} \in \ell_1$. We postulate risk aversion in terms of the resultant.

[12] In the classic Landau notation for asymptotic analysis, $|\pi(T) - T|$ is $o(T)$. Asymptotic permutations that are also one-to-one and onto mappings from $\mathbb{N}_0$ to $\mathbb{N}_0$ are called "bounded" in Lauwers (1998). Lauwers's characterization of this subclass of asymptotic permutations can be restated as invariance with respect to integration against any weak* accumulation point of uniform distributions on $\{0, 1, \ldots, T\}$. The closed convex hull of these accumulation points, denoted $\mathfrak{U}$, plays an important role in the next section.

For notational simplicity, we identify any $\boldsymbol{u} \in \mathbf{W}$ with the point mass $\delta_{\boldsymbol{u}} \in \mathcal{M}$.

2.2.5. *Asymmetric weak orders*

Our first four postulates on preferences on $\mathcal{M}$ guarantee that they have a continuous, risk averse, expected utility representation. To this we add respect for first order overtaking. We give the postulates in terms of a binary relation $\succ$ on the set of probabilities $\mathcal{M}$. As usual:

- $\succ$ is **asymmetric** if $p \succ q$ implies that it is not the case that $q \succ p$;
- $\sim$ and $\succsim$ are defined by $p \sim q$ if neither $p \succ q$ nor $q \succ p$, and $p \succsim q$ by $p \succ q$ or $p \sim q$;
- $\succ$ is **negatively transitive** if for all $p, q, r \in \mathcal{M}$, $[p \succ r] \Rightarrow [[p \succ q] \vee [q \succ r]]$, and
- a negatively transitive, asymmetric $\succ$ is an **asymmetric weak order**.

2.3. *Five postulates*

With the building blocks in place, we now present our five postulates. We will impose the following.

Postulate I. **Weak Order**. $\succ$ is an asymmetric weak order.
Postulate II. **Independence**. For all $p, q, r \in \mathcal{M}$ and all $\alpha \in (0, 1)$, if $p \succ q$, then $\alpha p + (1 - \alpha)r \succ \alpha q + (1 - \alpha)r$.
Postulate III. **Continuity**. For all $q \in \mathcal{M}$, the sets $\{p \in \mathcal{M} : p \succ q\}$ and $\{p \in \mathcal{M} : p \prec q\}$ are open.
Postulate IV. **Risk aversion/concavity**. For any $p \in \mathcal{M}$, $\boldsymbol{r}(p) \succsim p$.
Postulate V. **Respect for asymptotic first order dominance**. $[\boldsymbol{u} \ggcurly_{fo} \boldsymbol{v}] \Rightarrow [\boldsymbol{u} \succ \boldsymbol{v}]$.

The first four postulates are standard in risk averse expected utility theory: Postulate I is the usual ordering assumption for preference relations; Postulate II is the "linearity in probabilities" assumption for the existence of an expected utility representation for $\succ$; Postulate III guarantees that the representation is continuous; and Postulate IV guarantees risk aversion in the form of concavity of the expected utility function. It is Postulate V that implies asymptotic patience and Pareto responsiveness.

2.4. *Representation*

The first part of the following result, i.e. the existence of a continuous, concave expected utility function on $\mathcal{M}$, depends only on Postulates I–IV, and is a direct consequence of Fishburn's (1982, Theorem 4, Ch. 3) work on expected utility preferences over distributions on convex subsets of vector spaces.

Theorem A. *If an order $\succ$ satisfies Postulates I–V, then there exists a continuous, concave $S : \mathbf{W} \to [0, \infty)$ such that $[p \succ q] \Leftrightarrow \left[\int S(\boldsymbol{u})\, dp(\boldsymbol{u}) > \int S(\boldsymbol{u})\, dq(\boldsymbol{u})\right]$ with $S(\cdot)$ both patient and Pareto responsive, that is, for all $\boldsymbol{u} \in \mathbf{W}$,*

(1) *for all asymptotic permutations π, $S(\boldsymbol{u}) = S(\boldsymbol{u}^{\pi})$, and*
(2) *for all $r > 0$, $S(\boldsymbol{u} + r1_N) = S(\boldsymbol{u})$ for all null N, and $S(\boldsymbol{u} + r1_B) > S(\boldsymbol{u})$ for all substantial B.*

It has proved difficult to show that a continuous, concave $S : \mathbf{W} \to [0, \infty)$ satisfying conditions (1) and (2) in Theorem A represents a preference ordering satisfying respect for first order dominance, Postulate V, but we have no counter-example.

2.5. *Pareto responsive tangents*

Modulo a caveat about boundary points, a concave $S : \mathbf{W} \to [0, \infty)$ is Pareto responsive if and only if its tangents (supdifferentials) are. However, the tangents to a patient $S(\cdot)$ need not be patient, and in §3.5.2, we show that imposing patience on the set of tangents implies that $S(\cdot)$ cannot be sensitive to many forms of intergenerational riskiness.

By definition, a continuous linear $L : \ell_\infty \to \mathbb{R}$ is belongs to the **supdifferential** of a concave function $S : \mathbf{W} \to [0, \infty)$ at a point $\boldsymbol{u} \in \mathbf{W}$ iff for all $\boldsymbol{v} \in \mathbf{W}$,

$$S(\boldsymbol{u}) + L(\boldsymbol{v} - \boldsymbol{u}) \geq S(\boldsymbol{v}). \tag{5}$$

This is denoted $L \in \boldsymbol{\partial} S(\boldsymbol{u})$, and the geometric form of the Hahn–Banach Theorem implies that $\boldsymbol{\partial} S(\boldsymbol{u})$ is non-empty for every interior $\boldsymbol{u}$ in $\mathbf{W}$. However, $\boldsymbol{\partial} S(\boldsymbol{u})$ may be empty at boundary points of $\mathbf{W}$.[13]

Let $\mathbb{V}$ denote the set of Pareto responsive, continuous linear $L : \ell_\infty \to \mathbb{R}$.

Definition 2.5. A continuous, concave $S : \mathbf{W} \to [0, \infty)$ is $\mathbb{V}$**-concave** on $\text{int}(\mathbf{W})$ if for all $\boldsymbol{u} \in \text{int}(\mathbf{W})$, the set of tangent functions at $\boldsymbol{u}$ is a closed, non-empty, norm bounded subset of $\mathbb{V}$. A function is $\mathbb{V}$**-concave** if the same condition holds for all $\boldsymbol{u} \in \mathbf{W}$.

Theorem B. *If* $S : \mathbf{W} \to [0, \infty)$ *is continuous, concave, and Pareto responsive, then it is* $\mathbb{V}$*-concave on* $\text{int}(\mathbf{W})$*, and every* $\mathbb{V}$*-concave* $S(\cdot)$ *is Pareto responsive.*

It is worth stressing here again that these tangents are not necessarily patient. A detailed example of a social welfare function satisfying Postulates I–V with tangents that are not patient follows the proof of Theorem B in the appendix. The intuition is simple: consider the utility stream $\boldsymbol{u} = (3, 1, 3, 1, 3, \ldots) = 3 \cdot 1_E + 1_{E^c}$ where $E = \{0, 2, 4, \ldots\}$ is the set of even numbered generations; it is plausible that a social welfare function $S(\cdot)$ is less sensitive to a small increase in the utilities of the E generations that it is to the same increase in the utilities of the E^c generations; this implies that the tangent function to $S(\cdot)$ at $\boldsymbol{u}$ increases more slowly in the 1_E direction than in the 1_{E^c} direction; but 1_E and 1_{E^c} are related by a shift permutation, and a patient tangent could not treat them differently.

2.6. *Tangents with additional properties*

Knowing the properties of the tangents of a function facilitates analyses of its optima. Beyond this, the pattern in Theorem B is quite general, in §3.4, we will see that if social a welfare function $S(\cdot)$ has tangents in a set $\mathcal{C}$, then $\mathcal{C}$ can determine which coalitions $S(\cdot)$ responds to and which it ignores. Further, in §3.5, we will see that having tangents invariant to bounded permutations can limit the usefulness of a social welfare function. Theorem C shows that the previously studied

[13] One sees the same pattern already in finite dimensional settings, the Cobb–Douglas functions $f(x_1, x_2) = x_1^\alpha x_2^{(1-\alpha)}$ for $0 < \alpha < 1$ and $x_1, x_2 \geq 0$ are a case in point.

patient social welfare functions have tangents with this invariance property. The first step is to reprise the representation results for continuous linear functionals on ℓ_∞.

2.6.1. Representation of tangents

Every continuous linear functional, $L : \ell_\infty \to \mathbb{R}$, has a representation as $L(\boldsymbol{u}) = \langle \boldsymbol{u}, \eta_L \rangle$ where: η_L belongs to $\boldsymbol{ba}(\mathbb{N}_0)$, the set of a bounded, finitely additive measure on the subsets of $\mathbb{N}_0$; and $\langle \cdot, \cdot \rangle$ is the bi-linear function $\langle \boldsymbol{u}, \eta \rangle = \int_{\mathbb{N}_0} u_t \, d\eta(t)$. Further, for given $\eta \in \boldsymbol{ba}(\mathbb{N}_0)$, the mapping $L_\eta(\boldsymbol{u}) := \langle \boldsymbol{u}, \eta \rangle$ is continuous and linear, and the mapping $\eta \leftrightarrow L_\eta$ is an isometric isomorphism.

We give $\boldsymbol{ba}(\mathbb{N}_0)$, equivalently, the set of continuous linear functionals on ℓ_∞, the weak*-topology. This can be defined by specifying, as a neighborhood basis at each η, the class of sets

$$G(\eta; (\boldsymbol{u}_1, \epsilon_1), \ldots, (\boldsymbol{u}_I, \epsilon_i)) = \{\eta' \in \boldsymbol{ba}(\mathbb{N}_0) : |\langle \boldsymbol{u}_i, \eta - \eta' \rangle| < \epsilon_i, i = 1, \ldots I\} \tag{6}$$

where I is finite, each $\boldsymbol{u}_i$ is an element of ℓ_∞, and each ϵ_i is strictly positive. The weak* topology can also be specified by convergence of nets, $\eta^\alpha \to \eta$ iff $\langle \boldsymbol{u}, \eta^\alpha \rangle \to \langle \boldsymbol{u}, \eta \rangle$ for all $\boldsymbol{u} \in \ell_\infty$. By Alaoglu's theorem, every norm bounded subset of $\boldsymbol{ba}(\mathbb{N}_0)$ has compact weak* closure.

2.6.2. Prominent sets of tangents

The following compact convex sets of measures will be used several times below. For $T < T'$, $\eta_{T,T'}$ denotes the uniform distribution on $\{T, T+1, \ldots, T'\}$.

- $\Delta \subset \boldsymbol{ba}(\mathbb{N}_0)$ is the set of probabilities.
- $\Delta^{pfa} \subset \Delta$ is the set of purely finitely additive probabilities, probabilities for which $\eta(F) = 0$ for any finite $F \subset \mathbb{N}_0$.
- $\mathbb{BM} \subset \Delta^{pfa}$ is the set of **Banach–Mazur** limits, the set of purely finitely additive probabilities for which $\langle \boldsymbol{u}, \eta \rangle = \langle \boldsymbol{u}^\pi, \eta \rangle$ for all $\boldsymbol{u}$ and all finite permutations π.
- $\mathcal{P} \subset \mathbb{BM}$ is the intersection, over $\epsilon > 0$, of the closed convex hull of the weak* accumulation point of uniform distributions on $\{(1-\epsilon)T, T\}$ — for $\delta > 0$, let $A(\delta) = \cap_T \overline{\text{co}}\,(\{\eta_{(1-\delta)T',T'} : T' \geq T\})$, for $\epsilon > 0$, let $B(\epsilon) = \overline{\text{co}}\,(\{A(\delta) : 0 < \delta < \epsilon\})$, and set $\mathcal{P} = \cap_{\epsilon > 0} B(\epsilon)$.
- $\mathfrak{U} \subset \mathbb{BM}$ is the closed convex hull of the set of weak*-accumulation points of the uniform distributions $\{\eta_{0,T} : T \in \mathbb{N}_0\}$.

Comments. The sets Δ and Δ^{pfa} contain point masses, while every η in $\mathbb{BM}$, $\mathfrak{U}$, or $\mathcal{P}$ is non-atomic. The sets $\mathfrak{U}$ and $\mathcal{P}$ are very different kinds of Banach–Mazur limits: although for all $\boldsymbol{u} \in \mathbf{Erg}$ and all $\eta, \eta' \in \mathfrak{U} \cup \mathcal{P}$, $\langle \boldsymbol{u}, \eta \rangle = \langle \boldsymbol{u}, \eta' \rangle$, one can show that $\mathfrak{U} \cap \mathcal{P} = \emptyset$. The elements of $\mathfrak{U}$ can also be characterized as the accumulation points of the linear functionals

$$L_\tau(\boldsymbol{u}) := E\, \tfrac{1}{\tau+1} \textstyle\sum_{t=0}^{\tau} u_t \tag{7}$$

as $P(\tau \geq M) \to 1$ for all M where τ is the random time until society ends.[14]

[14] Judging the optimality of policies using a criterion that has uncertainty about the longevity of society also captures Sidgwick's "Utilitarian," who cares "as much (for posterity as for) his contemporaries, except in so far as the effect of his actions on posterity — and even the existence of human beings to be affected — must necessarily be more uncertain." See Sidgwick (1874, Book IV, Ch. I, §1).

2.6.3. Previously studied patient social welfare functions

If $\mathcal{C}$ is a compact and convex set of non-negative measures on $\mathbb{N}_0$, then $S_{\mathcal{C}}(\boldsymbol{u}) := \min\{\int_{\mathbb{N}_0} u_t\, d\eta(t) : \eta \in \mathcal{C}\}$ is a candidate social welfare function. For many classes $\mathcal{C}$, one can give explicit representations of the function $S_{\mathcal{C}}(\cdot)$; each $S_{\mathcal{C}}(\cdot)$ is continuous: if $\|\boldsymbol{u}^n - \boldsymbol{u}\|_\infty \to 0$, then $\int u_t^n\, d\eta(t) \to \int u_t^n\, d\eta(t)$ uniformly on $\mathcal{C}$; it is concave, as the minimum of a set of linear functionals is a concave function. The previously studied patient social welfare functions (that we are aware of) are of the form $S_{\mathcal{C}}(\cdot)$ where $\mathcal{C}$ is a set of probabilities.

The following characterizations will play a central role in comparative analyses of the properties of social welfare functions.

Theorem C. *For all $\boldsymbol{u} \in \ell_\infty$, we have the following equalities.*

(a) $\inf_t u_t = S_\Delta(\boldsymbol{u}) = \min\{\langle \boldsymbol{u}, \eta\rangle : \eta \in \Delta\}$.
(b) $\liminf_t u_t = S_{\Delta^{pfa}}(\boldsymbol{u}) = \min\{\langle \boldsymbol{u}, \eta\rangle : \eta \in \Delta^{pfa}\}$.
(c) $\liminf_{T\uparrow\infty} \inf_{j\geq 0} \frac{1}{T+1}\sum_{t=0}^{T} u_{j+t} = S_{\mathbb{BM}}(\boldsymbol{u}) = \min\{\langle \boldsymbol{u}, \eta\rangle : \eta \in \mathbb{BM}\}$.
(d) $\liminf_{\epsilon\downarrow 0} \liminf_{T\to\infty} \frac{1}{\epsilon T}\sum_{t=(1-\epsilon)T}^{T} u_t = S_{\mathcal{P}}(\boldsymbol{u}) = \min\{\langle \boldsymbol{u}, \eta\rangle : \eta \in \mathcal{P}\}$.
(e) $\liminf_T \frac{1}{T+1}\sum_{t=0}^{T} u_t = S_{\mathfrak{U}}(\boldsymbol{u}) = \min\{\langle \boldsymbol{u}, \eta\rangle : \eta \in \mathfrak{U}\}$.

Comments. The functions in (b)–(e) are patient in the sense that they are invariant to various classes of shifts containing, at least, the bounded ones. The tangents for the functions in (c)–(e) are also invariant to bounded permutations. The function $S_{\mathfrak{U}}(\cdot)$ in (e) has been the most widely studied. It agrees with $S_\varphi^\circ(\cdot)$ on **Erg** for the function $\varphi(x) = x$, and the class $\mathfrak{U}$ is exactly the set of accumulation points of the $L_\tau(\cdot)$'s given in (7). The function $S_{\mathbb{BM}}(\cdot)$ in (c) is Marinacci's *patient* social welfare function, the function $S_{\mathcal{P}}(\cdot)$ in (d) is his *completely patient* social welfare function (Marinacci, 1998).

3. Discussion

We now turn to a number of issues related to the Postulates and results given above. We organize this section as follows.

- In §3.1, we extend patient, Pareto responsive, concave social welfare functions defined on the closed linear set of ergodic sequences to all of **W**. This parallels the development of Banach–Mazur limits and yields an explicit class of social welfare functions satisfying our postulates.
- In §3.2, we study intergenerational problems with irreversible decisions and compare the implications of the degree 1 homogeneity of previously used social welfare functions with the strict curvatures allowed in ours.
- In §3.3, we examine how concave social welfare functions with different classes of tangents behave in weakly and strongly ergodic Markovian decision problems.
- In §3.4, we examine how different sets of tangents give rise to different classes of null and substantial classes, and how this leads to different forms of Pareto responsiveness.
- In §3.5, we end with a comparison with of our set of tangents and medial limits.

3.1. Extension of concave functionals

Let $c \subset \ell_\infty$ denote the closed linear subspace of $\boldsymbol{x} \in \ell_\infty$ for which $\lim_t x_t$ exists. For $\boldsymbol{x} \in c$, define $L^\circ(\boldsymbol{x}) = \lim_t x_t$. Banach–Mazur limits are the finite shift invariant, continuous linear extensions of L° from c to all of ℓ_∞. To develop our social welfare functions as extensions, we start with a closed linear subspace of ℓ_∞ strictly larger than c.

Definition 3.1. $\boldsymbol{u}$ in **W** has an **ergodic limit** or **occupation measure** $\mu(\cdot|\boldsymbol{u})$ if for every continuous $f : \mathbb{R} \to \mathbb{R}$, $\int f(r)\, d\mu(r|\boldsymbol{u}) = \lim_T \frac{1}{T+1} \sum_{t=0}^{T} f(u_t)$. The closed linear subspace of $\boldsymbol{u}$'s with this property is denoted **Erg**.[15]

Utility streams with the same occupation measure must be indifferent for all preferences respecting first order overtaking — if $\mu(\cdot|\boldsymbol{u}) = \mu(\cdot|\boldsymbol{v})$, then for all $\epsilon > 0$,

$$(\boldsymbol{u} + 2\epsilon 1_{\mathbb{N}_0}) \gg_{fo} (\boldsymbol{v} + \epsilon 1_{\mathbb{N}_0}) \gg_{fo} \boldsymbol{u},$$

sending $\epsilon \downarrow 0$ and appealing to continuity completes the argument. One class of our social welfare functions defines $S^{\circ}_{\varphi}(\boldsymbol{u}) = \int_{\mathbb{R}_+} \varphi(r)\, d\mu(r|\boldsymbol{u})$ for $\boldsymbol{u} \in$ **Erg** where $\varphi : [0, \infty) \to [0, \infty)$ is strictly increasing and concave. The patient, concave, Pareto responsive $S^{\circ}_{\varphi}(\cdot)$ functions on **Erg** can be extended to all of **W**.

Proposition 1. *For any strictly increasing, and concave $\varphi : [0, \infty) \to [0, \infty)$, there exists a patient, Pareto responsive $S_\varphi : \mathbf{W} \to [0, \infty)$ with the property that for all $\boldsymbol{u} \in$ **Erg**, $S_\varphi(\boldsymbol{u}) = S^{\circ}_{\varphi}(\boldsymbol{u})$.*

The following gives a sense of the preferences induced by the $S_\varphi(\cdot)$.

Example 3.1. Consider the utility streams $\boldsymbol{u} = (0, 1, 0, 1, 0, 1, \ldots)$, $\boldsymbol{v} = (1, 0, 1, 0, 1, 0, \ldots)$, and the shift permutation $\pi(t) = t - 1$. We have $\boldsymbol{v} = \boldsymbol{u}^{\pi}$, and $\boldsymbol{u}$ and $\boldsymbol{v}$ are indifferent because they have the same occupation measure on $\mathbb{R}_+$, namely half mass on 0 and half mass on 1, $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. By contrast, for any $0 \le \alpha \le 1$, the occupation measure of $\alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}$ is $\frac{1}{2}\delta_\alpha + \frac{1}{2}\delta_{(1-\alpha)}$. If the S_φ arises from a strictly concave φ, then for any $0 < \alpha < 1$, $S_\varphi(\alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}) > S_\varphi(\boldsymbol{u}) = S_\varphi(\boldsymbol{v})$.

3.2. Homogeneity, risk neutrality, and irreversibility

Because **W** is non-separable, there are no strictly concave functions $S : \mathbf{W} \to [0, \infty)$ (Choquet, 1969, Cor. 27.3 *et seq*). This means that there always exist directions of change from an interior stream $\boldsymbol{u}$ along which the social welfare function is linear/risk neutral. The following is an example of this kind of 'flatness' for our social welfare functions.

Example 3.2. For $\mu^\dagger$ a probability distribution on an interval $[0, \overline{u}] \subset \mathbb{R}$ let $\mathcal{L}_{\mu^\dagger}$ denote the set of all $\boldsymbol{u} \in$ **Erg** having $\mu^\dagger$ as their occupation measure. It is a non-empty, infinite dimensional convex subset of **W**. Every S satisfying Postulates I–V is constant on each $\mathcal{L}_{\mu^\dagger}$ — for $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{L}_{\mu^\dagger}$ and any $\epsilon > 0$, $(\boldsymbol{u} + 2\epsilon 1_{\mathbb{N}_0}) \gg_{fo} (\boldsymbol{v} + \epsilon 1_{\mathbb{N}_0}) \gg_{fo} \boldsymbol{u}$, as $\epsilon \downarrow 0$, continuity delivers $S(\boldsymbol{u}) \ge S(\boldsymbol{v}) \ge S(\boldsymbol{u})$.

3.2.1. Directional risk neutrality: test cases

Risk neutrality along the "more unequal/less unequal" and "more than/less than" directions is particularly problematic for social welfare functions. The social welfare functions $S_{\mathcal{C}}(\cdot)$ are

[15] Replacing $\int f(r)\, d\mu(r|\boldsymbol{u})$ by $\int 1_E(r)\, d\mu(r|\boldsymbol{u})$ recovers the usual definition for ergodic processes. Defining $\boldsymbol{u}$ by $u_t = \frac{t}{t+1}$, $\mu(\cdot|\boldsymbol{u})$ is point mass on 1 using our definition. By contrast, $\frac{1}{T+1}\sum_{t=0}^{T} 1_{[0,1)}(u_t) \equiv 1 \neq \mu([0, 1))$ so that $\boldsymbol{u}$ is not ergodic with the usual definition.

homogeneous of degree 1 (Hd1) — for $\boldsymbol{u} \geq 0$ and $\lambda > 0$, $S_{\mathcal{C}}(\lambda \boldsymbol{u}) = \lambda S_{\mathcal{C}}(\boldsymbol{u})$. It is this form of risk neutrality over generational utilities that renders $S_{\mathcal{C}}(\cdot)$ unsuitable for use in many long-run models.

In many dynamic optimization problems, we assume that the well-being at time t takes the form $u_t = u(a_t, X_t)$ where a_t is the current action, X_t is the current state, and the action-state pair determine the distribution of X_{t+1}. The concavity of $u(a, \cdot)$ captures generational risk aversion as well as the tradeoff between present and future changes in the state. Kreps and Porteus (1978), Epstein and Zin (1989) and many others have worked on disentangling generational risk attitudes from intergenerational tradeoffs while preserving the structural stationarity that arises with discounting. With $S_{\mathcal{C}}(\cdot)$ preferences, there can be generational risk aversion, but the intergenerational risk attitude is fixed at neutrality in important ways.[16]

We use three test cases to examine the behavior of the five social welfare functions in Theorem C as well as the $S_{\varphi}(\cdot)$ functions described in Proposition 1. In each of the three cases, we are evaluating a once-and-for-all decision to expose the present and future generations to risk that may increase or decrease utility. What differs across the tests cases is the degree of persistence of the risk and its relation to the utility of the generations.

A bounded non-negative stochastic process is a collection of random variables $\boldsymbol{X} = (X_t)_{t \in \mathbb{N}_0}$, with $0 \leq X_t \leq B$ for some bound B. For a stream of utilities $\boldsymbol{u}$, we denote by $\boldsymbol{u} \odot \boldsymbol{X}$ the random stream $(u_0 X_0, u_1 X_1, u_2 X_2, u_3 X_3, \ldots)$ that arises from multiplying each generation's measure of well-being, u_t, by the corresponding X_t. Throughout, we assume that for all t, $P(X_t < 1) > 0$ and $P(X_t > 1) > 0$. Our three test cases are:

- a single shock, all of the X_t are equal to X_0;
- an i.i.d. sequence of shocks, the X_t are independent and identically distributed; and
- an independent sequence of shocks, the X_t are independent but have different distributions.

If one interprets the u_t as the expected utility of a generation at time t, then generation t is 'risk' neutral with respect to multiplicative shocks to u_t. However, we understand each u_t as a measure of the welfare of the generation at t rather than as an expected utility. When risk attitudes within a generation can differ, the nearest we have to consensus on an aggregate measure of welfare in risky situations is the sum of individual uncertainty equivalents (Chambers, 2012; Chambers and Echenique, 2012). Without additional intra-generational distributional assumptions, multiplicative shocks with mean greater than 1 may not be socially preferred by a generation.

Beyond these considerations, agency is valued by decision makers. This means that society may not wish to allow the present generation to take a decision that affects all succeeding generations, even if those generations would take the same decision on their own.

3.2.2. A single shock

If the social welfare function is Hd1, then preferences for a single multiplicative shock depend only on the expected value of the shock.

Lemma 2. *If all of the X_t are equal to X_0 and $\mathcal{C} \subset \Delta$, then for all $\boldsymbol{u}$, $E\, S_{\mathcal{C}}(\boldsymbol{u} \odot \boldsymbol{X}) = (E\, X_0) \cdot S_{\mathcal{C}}(\boldsymbol{u})$.*

[16] The influence of a timely conversation with V. Bhaskar is quite pronounced in this section.

In this case, the Hd1 property of the $S_{\mathcal{C}}(\cdot)$ preferences implies that the optimal once-and-for-all decision is to accept the risk if $E\,X_0 > 1$ and to reject if $E\,X_0 < 1$. By contrast, whether or not the decision is optimal for one of the $S_\varphi(\cdot)$ depends on the interaction of the occupation measure and the distribution of X_0, as well as the risk aversion encoded in $\varphi(\cdot)$.

Example 3.3. Suppose that the occupation measure for $\boldsymbol{u}$ is $\mu = \frac{1}{2}\delta_5 + \frac{1}{2}\delta_7$, $P(X_0 = 0.1) = \frac{1}{6}$ and $P(X_0 = 1.2) = \frac{5}{6}$. Because $E\,X_0 = \frac{61}{60} > 1$, every Hd1 social welfare function would accept the $\frac{1}{6}$ chance of a 90% decline in every future generation's well-being in return for the $\frac{5}{6}$ chance of a 20% gain. By contrast, $E\,S_\varphi(\boldsymbol{u} \odot \boldsymbol{X}) = \int \varphi\, d\mu'$ where

$$\mu' := \tfrac{1}{6}(\delta_{0.5} + \delta_{0.7}) + \tfrac{5}{6}(\delta_6 + \delta_{8.4}).$$

The optimal once-and-for-all decision depends on the risk aversion of φ viewed as an expected utility function, if $\int \varphi(r)\, d\mu(r) > \int \varphi(r)\, d\mu'(r)$, then it is optimal to accept the risk, and it is optimal to reject the risk if the inequality is reversed.

3.2.3. I.i.d. shocks

In this case, the decision involves subjecting all future generations to a set of utilities that is riskier, but that may have a higher mean. Under a regularity assumption on $\boldsymbol{u}$, three of the $S_{\mathcal{C}}(\cdot)$ preferences forbid this, another two accept/reject as the expected value of the multiplicative shocks is greater than/less than 1, while the $S_\varphi(\cdot)$ preferences again depend on the interaction of the occupation measure, the distribution of X_t's, and the risk aversion encoded in $\varphi(\cdot)$.

Lemma 3. *If $\boldsymbol{u}$ is a probability 1 realization of an i.i.d. sequence of utilities that is independent of $\boldsymbol{X}$ and the X_t are i.i.d. with support $[a, b]$, $a < 1 < b$ and mean m, then*

(a) $E\,S_{\mathcal{C}}(\boldsymbol{u} \odot \boldsymbol{X}) = a \cdot S_{\mathcal{C}}(\boldsymbol{u})$ *for* $\mathcal{C} = \Delta, \Delta^{pfa}, \mathbb{BM}$, *while*
(b) $E\,S_{\mathcal{C}}(\boldsymbol{u} \odot \boldsymbol{X}) = m \cdot S_{\mathcal{C}}(\boldsymbol{u})$ *for* $\mathcal{C} = \mathfrak{U}, \mathcal{P}$.

To see how the interactions work for the $S_\varphi(\cdot)$ preferences, note that the calculations in Example 3.3 carry through to the i.i.d. case.

3.2.4. Independent shocks with different distributions

To fix ideas, suppose that the X_t are independent, have support $[a_t, b_t]$, $a_t < 1 < b_t$, and mean m_t. Further suppose that the supports and the means are functions of u_t, $a_t = f(u_t)$, $b_t = g(u_t)$ and $m_t = h(u_t)$. If $h(\cdot)$ is an increasing function, then on average, the $\boldsymbol{u} \mapsto (\boldsymbol{u} \odot \boldsymbol{X})$ mapping rewards the rich for being rich while plaguing the poor for being poor. By contrast, if $h(\cdot)$ is decreasing, then the mapping $\boldsymbol{u} \mapsto (\boldsymbol{u} \odot \boldsymbol{x})$ acts, in expectation, as a progressive tax rate, equalizing the distribution across generations. For three of the $S_{\mathcal{C}}(\cdot)$ social welfare functions given above, only the lower bound function, $f(\cdot)$ can play a role, for the other two, only the $h(\cdot)$ function can play a role.

Lemma 4. *Suppose that $\boldsymbol{u}$ is a probability 1 realization of an i.i.d. sequence of utilities with distribution μ, that $\boldsymbol{u}$ is independent of $\boldsymbol{X}$, and that the X_t are mutually independent with support $[f(u_t), g(u_t)]$ with $f(u_t) < 1 < g(u_t)$ and mean $h(u_t)$. Suppose further that the X_t have densities uniformly bounded away from 0. Then*

(a) $E\,S_{\mathcal{C}}(\boldsymbol{u} \odot \boldsymbol{X}) = \inf_t u_t \cdot f(u_t)$, *and for* $\mathcal{C} = \Delta, \Delta^{pfa}, \mathbb{BM}$, *and*
(b) $E\,S_{\mathcal{C}}(\boldsymbol{u} \odot \boldsymbol{X}) = [\int h(r)\, d\mu(r)] \cdot S_{\mathcal{C}}(\boldsymbol{u})$ *for* $\mathcal{C} = \mathcal{P}, \mathfrak{U}$.

The quantity $\int h(r)\,d\mu(r)$ is the long-run average multiplication factor for the generational utilities in the sequence $\boldsymbol{u}$. For the $S_{\mathcal{P}}(\cdot)$ and $S_{\mathfrak{U}}(\cdot)$ preferences only this average can matter. By contrast, with the $S_\varphi(\cdot)$ preferences, an increasing $h(\cdot)$ will tend to make the occupation measure more iniquitous, and the concavity of φ will tend to make the decision sub-optimal.

3.3. Optimization for ergodic problems

In general, the different objective functions have different sets of optima. For weakly ergodic Markovian decision problems (MDPs), we show that a large class of Hd1 social welfare functions all have the *same* set of optimal policies that have been studied in the operations research literature. We also characterize how the optima differ for different $S_\varphi(\cdot)$ functions. We begin with the general considerations and then turn to a two-state, strongly ergodic MDP to see how the generalities work.

For MDPs, all of the probabilities are, by assumption, countably additive.

3.3.1. Generalities

An MDP is specified by the quintuple (X, A, K, u, Q) with the following interpretations: at each $t \in \mathbb{N}_0$, the present state, a point $x \in X$, is observed and an action $a \in A(x)$, $A(x) \subset A$, the set of feasible actions, is chosen; the period utility is $u(a, x)$, and the state randomly transitions to a new state according to the transition probability $Q(\cdot|a, x)$. The graph of the correspondence $x \mapsto A(x)$ is $K \subset (A \times X)$.

The minimal regularity assumptions are: for all x, $A(x)$ is non-empty; the period utility function, $u : K \to \mathbb{R}_+$ is measurable and bounded[17]; X and K are measurably isomorphic to measurable subsets of a Polish space; K has measurable selections; and for each measurable $E \subset X$, $(a, x) \mapsto Q(E|a, x)$ is measurable.

A policy, α, is a measurable function from histories to feasible actions. The existence of measurable selections from K guarantees that policies exist. Under the assumptions given, for each starting state, x_0, each policy α gives rise to a unique outcome probability $p_\alpha(\cdot|x_0) \in \mathcal{M}$ that is consistent with the transition probability and the policy. A stationary policy is a policy that only depends on the last observed state. For stationary α, $p_\alpha(\cdot|x_0)$ describes a Markov process.

Definition 3.2. A probability $p \in \mathcal{M}$ is **weakly ergodic** if $p(\mathbf{Erg}) = 1$, it is **strongly ergodic** if there is some occupation measure μ° such that $p(\{\boldsymbol{u} : \mu(\cdot|\boldsymbol{u}) = \mu^\circ\}) = 1$. An MDP (X, A, K, u, Q) is **weakly/strongly ergodic** as the outcome distribution, $p_\alpha(\cdot|x_0)$, associated with any stationary policy α and starting point x_0 is weakly/strongly ergodic.

Every strongly ergodic p is weakly ergodic, and the same holds for MDPs. For $\boldsymbol{u} \in \mathbf{Erg}$, $\mathbf{lra}(\boldsymbol{u})$ denotes the long-run average payoff, $\lim_T \frac{1}{T+1}\sum_{t=0}^{T} u_t$, equivalently, $\mathbf{lra}(\boldsymbol{u}) = \int r\,d\mu(r|\boldsymbol{u})$ where $\mu(\cdot|\boldsymbol{u})$ is $\boldsymbol{u}$'s occupation measure. Assumptions on (X, A, K, u, Q) sufficient to guarantee that it is strongly ergodic and that there exists a stationary deterministic policy maximizing the expected value of $\mathbf{lra}(\boldsymbol{u})$ have been extensively studied (Arapostathis et al., 1993; Jaśkiewicz and Nowak, 2006; Feinberg et al., 2012). For such problems, quantitative and qualitative information about the optimal policies can be found by solving the Poisson equations.

[17] Boundedness is assumed for our social welfare functions, but unbounded rewards are frequently treated in the operations research literature on MDPs.

These are strikingly similar to the Bellman equations for discounted dynamic programming, and they can be interpreted as describing a tradeoff between present utility and the total utility while maximizing the "value" along a "sojourn" in the Markov process that arises from the stationary policy (Meyn, 1997).

Less has been done on weakly ergodic problems where one replaces the search for a stationary policy that maximizes the single number, $\mathbf{lra}(\boldsymbol{u})$, with the search for a policy that maximizes $\int \mathbf{lra}(\boldsymbol{u})\, dp_\alpha(\boldsymbol{u}|x_0)$. To tie optimality for MDPs to optimality for our social welfare functions we need the following class of probabilities on $\mathbb{N}_0$,

$$\mathcal{G} := \{\eta \in \Delta : (\forall \boldsymbol{u} \in \mathbf{Erg})(\forall f \in C(\mathbb{R}))[\int f(u_t)\, d\eta(t) = \int f(r)\, d\mu(r|\boldsymbol{u})]\} \tag{8}$$

where $C(\mathbb{R})$ is the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$. These are the set of measures on $\mathbb{N}_0$ for which the time average of each $\boldsymbol{u} = (u_0, u_1, u_2, \ldots) \in \mathbf{Erg}$ is equal to the occupation measure, $\mu(\cdot|\boldsymbol{u})$. For the sets of tangents discussed in Theorem C, it can be shown that any probability having a positive density with respect to an element of the convex hull of $\mathfrak{U}$ and $\mathcal{P}$ belongs to $\mathcal{G}$, and that $\mathcal{G}$ is a strict subset of $\mathbb{BM}$.

Lemma 5. *If (X, A, K, u, Q) is weakly ergodic, then for every initial state x_0,*

(a) *α solves $\max_\alpha \int \mathbf{lra}(\boldsymbol{u})\, dp_\alpha(\boldsymbol{u}|x_0)$ if and only if, for all non-empty, weak*-compact $\mathcal{C} \subset \mathcal{G}$, it solves $\max_\alpha \int S_{\mathcal{C}}(\boldsymbol{u})\, dp_\alpha(\boldsymbol{u}|x_0)$, and*
(b) *for any continuous $\varphi : \mathbb{R} \to \mathbb{R}$, α solves $\max_\alpha \int \mathbf{lra}(\varphi \circ \boldsymbol{u})\, dp_\alpha(\boldsymbol{u}|x_0)$ if and only if, for all $\eta \in \mathcal{G}$, it solves $\max_\alpha \int S_\varphi(u_t)\, d\eta(t)$.*

Lemma 5(a) shows that choosing between any of the Hd1 social welfare functions that are well-behaved on **Erg** makes no difference, one might as well maximize the expected long-run average. By contrast, Lemma 5(b) shows that the $S_\varphi(\cdot)$ functions introduce aversion to intergenerational risk if $\varphi(\cdot)$ has curvature.

One arrives at the parallel conclusions for a strongly ergodic (X, A, K, u, Q) by noting that, in this case, $\mathbf{lra}(\boldsymbol{u})$ and $\mathbf{lra}(\varphi \circ \boldsymbol{u})$ are $p_\alpha(\cdot|x_0)$-almost everywhere constant for stationary α's.

3.3.2. A climate change/species extinction model

Two state Markovian decision problems with strong ergodic properties provide a useful class of test cases for our social welfare functions. To focus on the role of different social welfare functions, we work with a drastically simplified example in which the richness of the biosphere is a crucial ingredient to human welfare.[18]

Example 3.4. The world's ecosystem can be in one of two states, damaged or undamaged: in the damaged state, the seas, forests and the biota that survive are unable to produce oxygen and resources in the amounts humans have become accustomed to; in the undamaged state, the seas and forests are able to produce oxygen concentrations and resources that can support life as we currently know it. Payoffs and actions capture the following tradeoffs: a generation in a good state

[18] The ideas in this model have three sources: Ehrlich and Ehrlich's (1981) analogy between species in an ecosystem and rivets in an airplane, some rivets are redundant and one would not miss them if they are lost, but a large enough cumulative loss of rivets leads to a crash; Costanza et al. (1997) (under)estimate the value of the non-marketed services that humanity receives, yearly, from the world's ecosystem as 1.8 times the yearly world GDP; and Bacinni et al. who show, more precisely than before, that degradation of tropical forests have made them into a large net carbon source.

can sacrifice some present utility in order to lower the future probability of disastrous change; a generation in a bad state must sacrifice some of their present utility in order to raise the future probability of a return to a better world.

(a) In the undamaged state, $x = G$, society chooses the transition probability, $r \in [\underline{r}, \overline{r}]$ to the damaged state, $x = B$, $0 < \underline{r}$ and $\overline{r} < 1$. The expected utility of choosing r is $u(G, r)$, and higher choices of r lead to a higher expected utility for a generation in the good state, $\partial u(G, r)/\partial r > 0$.
(b) In a parallel fashion, in the damaged state, $x = B$, society chooses the transition probability, $s \in [\underline{s}, \overline{s}]$ to the good state, $x = G$, $0 < \underline{s}$ and $\overline{s} < 1$. The expected utility of choosing s is $u(B, s)$, and higher choices of s lead to lower expected utility for a generation in the bad state, $\partial u(B, s)/\partial s < 0$.

We make the blanket assumption that for all r, s, $u(G, r) > u(B, s)$.

3.3.3. The "no sacrifice" social welfare functions

For the social welfare functions $S_\Delta(\boldsymbol{u}) = \inf_t u_t$, $S_{\Delta pfa}(\boldsymbol{u}) = \liminf_t u_t$ and $S_{\mathbb{BM}}(\boldsymbol{u}) = \liminf_{T\uparrow\infty} \inf_{j\geq 0} \frac{1}{T+1}\sum_{t=0}^{T} u_{j+t}$, the optimal policies for the given model call for no sacrifice in the bad state, no matter how beneficial for future generations, and any level of exploitation is acceptable in the good state, no matter how detrimental to future generations. The argument runs as follows.

- For any policy, stationary or not, the world's ecosystem spends infinitely many periods in the damaged state. Ergo, to maximize either $S_\Delta(\boldsymbol{u}) = \inf_t u_t$ or $S_{\Delta pfa}(\boldsymbol{u}) = \liminf_t u_t$, it is necessary and sufficient to maximize the welfare of the generations living in the damaged state.
- For any policy, stationary or not, for every T, it is a probability 1 event that, in the long run, there are infinitely many length-$(T+1)$ periods spent in the bad state. Ergo, to maximize $S_{\mathbb{BM}}(\boldsymbol{u}) = \liminf_{T\uparrow\infty} \inf_{j\geq 0} \frac{1}{T+1}\sum_{t=0}^{T} u_{j+t}$, we have the same necessary and sufficient condition.

3.3.4. Balancing the weighted welfare of generations

For a stationary policy (r, s), let $\rho_{r,s}$ denote the occupation measure for the good state, that is, the long-run proportion of generations living in the good state. For the social welfare functions $S_{\mathfrak{U}}(\cdot)$ and $S_{\mathcal{P}}(\cdot)$ from Theorem C, as well as for $S_\varphi(\cdot)$ with $\varphi(r) = r$, the optimal stationary policy for the two state model of Example 3.4 is the one that maximizes long-run average payoff,

$$\rho_{r,s} u(G, r) + (1 - \rho_{r,s}) u(B, s). \tag{9}$$

The Poisson equation (the generalization of the Bellman equation appropriate for maximizing long-run average payoffs) that describes the optimal r and s explicitly gives a tradeoff between the utility of the present generation and future generations. One sees the same tradeoffs here — $\rho_{r,s}$ decreases in r and increases in s.

For the $S_\varphi(\cdot)$ preferences with a strictly concave $\varphi(\cdot)$, the optimal policy shifts more weight to the welfare of worst off generations — it is the one that maximizes

$$\rho_{r,s}\varphi(u(G, r)) + (1 - \rho_{r,s})\varphi(u(B, s)). \tag{10}$$

In the neighborhood of an optimal policy, the relevant weights given to the generations are the derivatives $\varphi'(u(G,r))$ and $\varphi'(u(B,s))$. The concavity of $\varphi(\cdot)$, combined with $u(G,r) > u(B,s)$, implies that there is more weight on generations in the bad state, but not infinitely more weight as in the "no sacrifice" social welfare functions.

3.4. Pareto responsiveness and sets of tangents

Sets of tangents, $\mathcal{C}$, determine the class of null sets and substantial sets, and these in turn determine the associated Pareto properties. For any given $\mathcal{C} \subset \boldsymbol{ba}(\mathbb{N}_0)$, define $\boldsymbol{N}_{\mathcal{C}}$ as the class of $\mathcal{C}$-null coalitions, that is,

$$\boldsymbol{N}_{\mathcal{C}} = \{N \subset \mathbb{N}_0 : (\forall \eta \in \mathcal{C})[\langle 1_N, \eta\rangle = 0]\}, \tag{11}$$

and define $\boldsymbol{B}_{\mathcal{C}}$ as the class of $\mathcal{C}$-substantial coalitions, that is,

$$\boldsymbol{B}_{\mathcal{C}} = \{B \subset \mathbb{N}_0 : \inf_{\eta\in\mathcal{C}} \langle 1_B, \eta\rangle > 0\}. \tag{12}$$

Comments.

- The closed span of the indicators of the $\mathcal{C}$-null coalitions is a subset of the orthogonal complement of $\mathcal{C}$.
- The only Δ-null coalition is the empty set, and the only Δ-substantial coalition is $\mathbb{N}_0$.
- The only Δ^{pfa}-null coalitions are the finite sets, and every set with finite complement is Δ^{pfa}-substantial.
- The $\mathfrak{U}$-null and substantial coalitions are those identified as null and substantial in Definition 2.1 (see Lemma 7 below).
- The $\mathbb{BM}$-null coalitions are the ones with 0 Banach density, i.e. those for which $\lim_T \sup_j \frac{1}{T+1}\sum_{t=0}^{T} 1_N(j+t) = 0$, and the $\mathbb{BM}$-substantial coalitions are those for which $\lim_T \inf_j \frac{1}{T+1}\sum_{t=0}^{T} 1_B(j+t) > 0$.

The set of tangents to a social welfare function determine the form that its Pareto responsiveness takes.

Lemma 6. *For each closed, convex $\mathcal{C} \subset \boldsymbol{ba}(\mathbb{N}_0)$, if $S : \mathbf{W} \to [0,\infty)$ is concave and all of its tangents belong to $\mathcal{C}$, then*

(a) *for all interior $\boldsymbol{u} \in \mathbf{W}$, all $r > 0$ and $N \in \boldsymbol{N}_{\mathcal{C}}$, $S(\boldsymbol{u} + r1_N) = S(\boldsymbol{u})$, and*
(b) *for all interior $\boldsymbol{u} \in \mathbf{W}$, all $r > 0$ and $B \in \boldsymbol{B}_{\mathcal{C}}$, $S(\boldsymbol{u} + r1_B) > S(\boldsymbol{u})$.*

3.5. Medial limits and dependence on long run averages

For $\mathcal{C} \subset \Delta$, the concave, Hd1 function $S_{\mathcal{C}}(\cdot) := \inf\{\langle \boldsymbol{u}, \eta\rangle : \eta \in \mathcal{C}\}$ is everywhere weakly below the convex, Hd1 function $T_{\mathcal{C}}(\boldsymbol{u}) := \sup\{\langle \boldsymbol{u}, \eta\rangle : \eta \in \mathcal{C}\}$. Following Meyer (1973), a **$\mathcal{C}$-medial limit** is a strongly affine map L satisfying

$$S_{\mathcal{C}}(\boldsymbol{u}) \le L(\boldsymbol{u}) \le T_{\mathcal{C}}(\boldsymbol{u}) \text{ for all } \boldsymbol{u}. \tag{13}$$

Meyer shows that, despite the deep involvement of the Axiom of Choice in the construction of the set Δ^{pfa}, the almost everywhere Δ^{pfa}-medial limits of sequences of bounded measurable functions are universally measurable.

3.5.1. Infinite dimensional indifference sets

For preferences satisfying our postulates, Example 3.2 shows that streams of utilities with the same occupation measure are indifferent. Such sets are closed, convex, and infinite dimensional. The indifference sets for the $S_{\mathcal{C}}(\cdot)$ are yet larger — they ignore all moments of the occupation measures except the first.

There is a linear subset of $\boldsymbol{u}$ in ℓ_∞ for which the inequalities in (13) are equalities, namely the set $\{\boldsymbol{u} : (\forall \eta, \eta' \in \mathcal{C})[\langle \boldsymbol{u}, \eta - \eta' \rangle = 0]\}$. From these, one finds useful structure in the indifference sets for $S_{\mathcal{C}}(\cdot)$ — they always contain sets of the form $I_{\mathcal{C}}(r) := \{\boldsymbol{u} \in \mathbf{W} : (\forall \eta \in \mathcal{C})[\langle \boldsymbol{u}, \eta \rangle = r]\}$. For example, we have the following.

Lemma 7. *$\boldsymbol{u} \in I_{\mathfrak{U}}(r)$ if and only if $\boldsymbol{u}$ is Cesaro summable with long run average equal to r.*

In particular, the social welfare functions based on $\mathfrak{U}$-medial limits must ignore distributional concerns. To put it another way, if $S(\cdot)$ is a medial limit or has only medial limit tangents, then $\liminf_T \frac{1}{T+1} \sum_{t=0}^{T} (u_t - v_t) > 0$ implies that $S(\boldsymbol{u}) > S(\boldsymbol{v})$, but $S(\cdot)$ cannot discriminate between ergodic $\boldsymbol{u}$ and $\boldsymbol{v}$ *except* on the basis of long-run averages.

3.5.2. The implications of patient tangents

Lauwers (1998) shows that the $\mathfrak{U}$-medial limits are invariant for the one-to-one onto permutations satisfying $\lim_T \frac{|\pi(T)-T|}{T} = 0$. Closely related results hold for the asymptotic permutations used here. The tangents to the social welfare functions $S_{\mathcal{C}}(\cdot)$ belong to $\mathcal{C}$, and the patience of tangents implies an indifference across convex combinations of permutations.

Lemma 8. *If the tangents to a concave $S : \mathbf{W} \to [0, \infty)$ at all interior $\boldsymbol{u}$ are invariant with respect to a permutation π, then for all $\alpha \in [0, 1]$ and all interior $\boldsymbol{u}$, $S(\alpha \boldsymbol{u} + (1 - \alpha)\boldsymbol{u}^\pi) = S(\boldsymbol{u})$.*

For example, if the tangents are invariant with respect to finite permutations, then Lemma 8 implies indifference between

$$\boldsymbol{u} = (1, 3, 1, 3, 1, 3, \ldots),\ \boldsymbol{v} = (3, 1, 3, 1, 3, 1, \ldots),\ \text{and}$$
$$(\tfrac{1}{2}\boldsymbol{u} + \tfrac{1}{2}\boldsymbol{v}) = (2, 2, 2, 2, 2, 2, \ldots)$$

The occupation measures for $\boldsymbol{u}$ and $\boldsymbol{v}$ are identical, half of the generations receive utility 1 and half receive utility 3. The indifference between the equally risky $\boldsymbol{u}$ and $\boldsymbol{v}$ and the generationally riskless outcome $(\frac{1}{2}\boldsymbol{u} + \frac{1}{2}\boldsymbol{v})$ is a consequence of having patient tangents.

4. Two applications

In this section we present two applications. The first application examines the desirability of informative experimentation before making an irreversible decision. Patient preferences deliver a precautionary principle with the interpretation that regulatory approval or disapproval does not require complete knowledge. The second application verifies that the standard general equilibrium results can be recovered in our setting with no discounting. If dynasties in a general equilibrium model have patient preferences, then we have, under mild assumptions, equilibrium existence, as well as the first and second welfare theorems.

4.1. A precautionary principle

Even if it is not definitive, information that can change the optimal action is, in general, worth waiting for. Patience magnifies this effect. Species extinction represents an irreversible negative shock to the well-being of future generations, and patient optima stipulates going for probability 0 of species loss if possible. However, some irreversible decisions deliver shocks to well-being that could turn out both ways, positive or negative. We analyze how much research should be done before making such a decision. To highlight the role of patience in characterizing the optima, we again work in a starkly simplified model.

Example 4.1. Assume that the present path of utilities, $\boldsymbol{u}$, is ergodic with occupation measure μ. There is a hidden state X with $Prob(X < 0) > 0$ and $Prob(X > 0) > 0$. When the action $a = 1$ is taken, X will be added to generational well-being forever thereafter and no further actions are available. When the action $a = 0$ is taken, the possibility of taking the decision is closed off, and utility will be $\boldsymbol{u}$, no further actions are available. Until either $a = 0$ or $a = 1$ is chosen, the action s is available. When $a = s$ is chosen, a signal that is stochastically related to X will be observed. The action s corresponds to researching into the consequences of the irreversible decision. We assume that there is a random, unknown number of informative signals available. When they are exhausted, this is observed and any further signals are stochastically independent of X. Further, every attempt to observe an informative signal costs c.

Let p_t denote the posterior distribution of X after t signals have been observed and let $\mu * p_t$ be the convolution of μ with p_t. If at t, it is known that there are no more informative signals available, the action $a = 1$ (or $a = 0$) is optimal for the utility function $S_\varphi(\cdot)$ as

$$\int \varphi(r)\, d(\mu * p_t)(r) > (\text{or } <) \int \varphi(r)\, d\mu(r). \tag{14}$$

Again, the curvature of $S_\varphi(\cdot)$ encoded in $\varphi(\cdot)$ determines the attitude toward risk. One might expect more risk tolerance for a richer society (i.e. one with a $\boldsymbol{u}$ having a better occupation measure).

More generally, let B^- denote the set of beliefs $\mathbb{b}$ for which not taking the decision, $a = 0$, is optimal and B^+ the set for which $a = 1$ is optimal. The optimal policy takes the following form.

> Research until posterior beliefs either converges to the set B^+ or to the set B^-. If the beliefs have converged to B^+, take the irreversible decision, if they have converged to B^-, abandon the project.

In contrast with Sunstein's (2002), Deutsch's (2011), or other 'straw man' versions of the precautionary principle, it asks that irreversible decisions be delayed until uncertainty is reduced, not to nothing, but until society is as sure as possible that the expected utility of the benefits outweigh the expected utility of the costs. In practice, one way this convergence may be reached is that research may reveal ways to mitigate negative consequences, thereby lowering $Prob(X < 0)$.[19]

To have the costs enter the analysis in a more substantive way, one should consider a version of this model in which each generation is faced with their own irreversible decision(s). In such an

[19] Deutsch's *Principle of Optimism* is, in this context, the position that the probability of loss being positive is a consequence of a lack of knowledge (Deutsch, 2011, Ch. 9).

analysis, costs become a permanent component of generational well-being, and society optimally trades off between this component and the future additions/subtractions to the utility path.

4.2. *Patience and myopia in general equilibrium theory*

Bewley (1972) studies general equilibrium models with commodity spaces that are uniformly bounded sequences of non-negative consumption vectors in $\mathbb{R}^k$. Bewley's Theorem 1 gives sufficient conditions for the existence of a competitive equilibrium with prices that positive, finitely additive $\mathbb{R}^k$-valued measures, while his Theorems 2 and 3 study conditions for the existence of equilibria with prices in the set of summable $\mathbb{R}^k$-valued sequences. Brown and Lewis (1981) and Araujo (1985) study the same class of general equilibrium models and give results on the need for myopia in preferences for the existence of Pareto optima.

4.2.1. *General equilibrium*

Let ℓ_∞^k denote the set of $\boldsymbol{x} \in (\mathbb{R}^k)^{\mathbb{N}_0}$ with $\sup_{t\in\mathbb{N}_0} \|x_t\| < \infty$. Feasible consumptions and endowments belong to $\mathbf{W}^k := \{\boldsymbol{x} \in \ell_\infty^k : \boldsymbol{x} \geq 0\}$. Each agent/dynasty i in a finite set I, has an endowment $\boldsymbol{\omega}_i \in \mathbf{W}^k$. We give $\mathbf{W}^k$ the sup norm.

Assumption A. Each $\boldsymbol{\omega}_i$ is an interior point of $\mathbf{W}^k$.

This implies that $\boldsymbol{\omega} := \sum_i \boldsymbol{\omega}_i$ is also an interior point of $\mathbf{W}^k$. Feasible consumption streams are vectors $(\boldsymbol{x}_i)_{i\in I}$ with each $\boldsymbol{x}_i \in \mathbf{W}^k$, and $\sum_i \boldsymbol{x}_i \leq \boldsymbol{\omega}$. The preferences of i are given by a utility function $\boldsymbol{x} \mapsto U_i(\boldsymbol{x})$.

Assumption B. Preferences are represented by $U_i(\boldsymbol{x}_i) = S_i((u_{i,0}(x_{i,0}), u_{i,1}(x_{i,1}), \ldots)$; $S_i(\cdot)$ is patient and has Pareto responsive tangents at the boundary of $\mathbf{W}^k$, the $u_{i,t}(\cdot)$, $t \in \mathbb{N}_0$ are uniformly bounded, continuous, concave, strictly increasing period utility functions on $[0, \boldsymbol{\omega}_t] \subset \mathbb{R}_+^k$ with $u_{i,t}(0) = 0$ for all i, t, and $\liminf_t \{\min_i u_{i,t}(\boldsymbol{\omega}_{i,t})\} > 0$ for all i.

The assumption that the range of the utilities $u_{i,t}$ does not disappear guarantees that all substantial coalitions command a non-vanishing part of the economy. The assumption that the dynasties have preferences represented by one of our social welfare functions implies that they internalize the intergenerational externalities within the dynasty, but not across dynasties.

Definition 4.1. An **equilibrium** for an exchange economy model $\mathcal{E} = (\boldsymbol{\omega}_i, U_i)_{i\in I}$ is a feasible consumption stream, $(\boldsymbol{x}_i)_{i\in I}$ and a price $\boldsymbol{p}$ in the topological dual of ℓ_∞^k such that for all $i \in I$ and all $\boldsymbol{y} \in W^k$,

$$[U_i(\boldsymbol{y}) > U_i(\boldsymbol{x}_i)] \Rightarrow [\langle \boldsymbol{y}, \boldsymbol{p}\rangle > \langle \boldsymbol{\omega}, \boldsymbol{p}\rangle]. \tag{15}$$

4.2.2. *Equilibrium existence and the welfare theorems*

The following equilibrium existence result follows directly from Bewley (Bewley, 1972, Theorem 1).

Proposition 2. *Under Assumptions A and B, an equilibrium exists.*

For finite dimensional production and exchange economies without intragenerational externalities,[20] the First Welfare Theorem states that competitive equilibria are Pareto optimal, and the Second Welfare Theorem states that all Pareto optima are competitive equilibria after appropriate re-arrangement of the initial endowments. Because the tangents of $\mathbb{V}$-concave functions have representations as integrals against purely finitely measures, the equilibrium price in Proposition 2 must be purely finitely additive. Our next result shows that, for such preferences, we have the First and Second Welfare Theorems.

Proposition 3. *Under Assumptions A and B, every equilibrium is Pareto optimal, and every Pareto optimal allocation is an equilibrium for an appropriate re-arrangement of the initial endowments.*

4.2.3. The necessity of patience

Using preferences with tangents that are integrals against positive linear combinations of countably and purely finitely additive measures (Araujo, 1985, Theorem 3) shows that Pareto optimal allocations may not exist. Such preferences are not in the class of patient preferences we are using here, and this difference is key.

Example 4.2 *(Araujo).* For $I = \{1, 2\}$, let $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_2 = \mathbf{1}$ be the constant endowment of one unit of the single good. For an allocation $(\boldsymbol{x}, \boldsymbol{y})$ to 1 and 2, let $U_1(\boldsymbol{x}_1) = x_{1,0} + \langle \boldsymbol{x}, \eta \rangle$ where η is a non-negative, purely finitely additive measure (not a probability) that satisfies $\langle \boldsymbol{\omega}_1, \eta \rangle > 1$, and let $U_2(\boldsymbol{y}) = \langle \boldsymbol{y}, \gamma \rangle + \langle \boldsymbol{y}, \eta \rangle$ where γ is countably additive and strictly positive, say $\gamma_t = (1-\beta)\beta^t$ so that $\langle \boldsymbol{y}, \gamma \rangle = (1-\beta) \sum_{t=0}^{\infty} y_t \beta^t$.

Suppose now that $(\boldsymbol{x}, \boldsymbol{y})$ is an individually rational Pareto optimal allocation. Because $\gamma_t > 0$ for all t, Pareto optimality implies that $x_{1,t} = 0$ and $y_{2,t} = 2$ for all $t \geq 1$. Feasibility implies that $x_{1,0} \leq 2$. This means that 1's utility is bounded above by $2 + 0$. However, $U_1(\boldsymbol{\omega}_1) = 1 + \langle \boldsymbol{\omega}_1, \eta \rangle > 2$, a contradiction.

Comment. Continuous linear preferences on $\mathbf{W}^k$ can be decomposed into a countably additive part and a purely finitely additive part. To within any $\epsilon > 0$, the value of the countably additive parts are determined on $\{0, 1, \ldots, T\}$ for sufficiently large T, while the value of the purely finitely additive part is entirely determined on $\{T+1, T+2, \ldots\}$. For linear preferences, this implies that ϵ-individually rational and ϵ-Pareto optimal allocations exist. We conjecture that the same is true for concave utility functions.

4.2.4. Pareto optimal allocations with extreme properties

The next example demonstrates how the Pareto optimal allocations for dynasties with different patient and Pareto responsive utility functions can be extreme. The essential intuition is the same as that of optimal risk-sharing between two people, one of whom assigns mass 0 to an event E while the other assigns mass 1 to it.

Example 4.3. Let the agents and their endowments be as in the previous example. Suppose that $U_i(\boldsymbol{y}) = \langle \boldsymbol{y}, \gamma_i \rangle$ where (γ_1, γ_2) is an accumulation point of the set $\{(\eta_{0,T!}, \eta_{0,(T!)^2}) : T \in \mathbb{N}_0\}$

[20] Here we refer to externalities across agents within a generation, which is distinct from the intergenerational externalities across time that is the central concern of this paper.

where $\eta_{0,T}$ is the uniform distribution on $\{0, \ldots, T\}$. There exist (many) coalitions C with the property that $\langle 1_C, \gamma_1 \rangle = 1$ and $\langle 1_C, \gamma_2 \rangle = 0$. All Pareto optimal allocations must assign dynasty 1 the entire economy's endowment in C and must assign dynasty 2 the entire economy's endowment in C^c.

5. Summary and future work

We have studied the lengthening of the horizon for societal optimization and how this interacts with intertemporal externalities, up to and including their strongest form, irreversibilities. Viewing society as an aggregate of present and future generations, intergenerational equity captures societal patience. We propose a resolution of the conflict between intergenerationally equitable preferences and the Pareto criterion by specifying a class of social welfare functions that are patient, ignore null coalitions and pay attention to substantial coalitions.

We have investigated the properties of our social welfare functions from a pragmatic point of view, through the study of their implications in well-understood models.[21] For irreversible decisions, our social welfare functions allow for aversion to both inequality and risk in a way that previous ones could not. An easily characterized subclass of our social welfare functions pay attention to long-run distributional concerns and can be easily applied to illustrative examples of weakly ergodic problems. In dynastic general equilibrium models, our patient preferences restore both equilibrium existence and the classical welfare theorems. In situations where delay allows the accumulation of decision-relevant knowledge, our patient preferences deliver a version of the precautionary principle. There remain many other applications.

We now turn to the question of the relevance of our results within the bigger picture. We are talking about optimizing over very long time horizons, but we can barely imagine how the world will look, say, seven generations from now. This brings into question the practical relevance of polices that arise out of such an exercise based on our current knowledge and perceptions. The key idea that we want to stress here is that choices made today that contribute to an expansion of the choice set for future generations are desirable, even if we do not know the contingencies they will face. A particularly clear example is support for the social processes that create and accumulate reliable and publically available knowledge.[22]

In practice, planning for even a 200 year horizon with an uncertain future is not possible at a detailed level. However, at a less granular level, we can operate from the conviction that deeper understanding of causal structures of our natural and human environment increases the potential choices for our species. As a society, we facilitate enhancement in such knowledge in a variety of ways, e.g. by funding portfolios of science and technology projects at various levels of understanding and completion. Many of those projects at their early stages are too speculative to be undertaken by any private firm acting in the interests of its shareholders, who would want a rate of return commensurate with investment covering only a single lifetime.[23]

Viewing society as consisting of present and future people, the need for work on new knowledge that benefits the future suggests that over the appropriate time horizon for such consider-

[21] From the founder of pragmatism, Charles Sanders Peirce, "… the practical effects of the objects of your conception … is the whole of your conception of the object."

[22] For the history of emergence of the 'open science' model of knowledge, see David (2004).

[23] See Mazzucato (2015) for analyses of the central role that relatively small governmental initiatives like SBIR, ARPA, NIH, NASA, SEMATECH and others have played in facilitating both basic research and commercialization of that research long before venture capitalists were willing to provide funding.

ations, discounting yields the wrong answers. This is another area in which we are following Pigou (1920, pp. 29–30), "(long-run planning) is the clear duty of the Government, which is the trustee for unborn generations as well as for its present citizens ... ." This is Pigou's response to the observation that "our (individual) telescopic faculty is defective" (Pigou 1920, p. 25).

A related question then arises about the planning horizon of the actual decision making bodies in any given society at any given time. Even if we theoretically establish the desirability of a policy that takes a very long view, what are the incentives for the current lot of decision makers to implement such a policy? In the context of the US, for example, it can be argued that there is a pressing need for institutions able to withstand the vicissitudes of a political system that operates on a 2-year electoral cycle with constant fund-raising pressures.[24]

Another theoretical issue that plagues the discussion of the kind of long run optima that we propose is the problem of "underselectiveness," namely, the observation that the class of optimal policies are too large as they ignore what happens to null sets of generations. The long-run average is immune to finite shift permutations (and many more besides). "Present profligacy" arises if a finite number of the early generations squander resources, leaving the future generations barely able to recover. Following Chichilnisky (1996), a "dictatorship of the future" arises if a finite number of early generations sacrifice nearly all of their own consumption for the purposes of arriving at a richer future more quickly. The profligate and dictatorial policies can arise as optima for any criterion that is immune to the welfare of a finite subset of an infinite population.

We have not addressed these issues in this paper. In separate work, we show that perfection-like criteria — that look at limits of solutions to full support perturbations of the problem — deliver strategies satisfying a "conditional equal treatment property." The essence of the conditional equal treatment property is most easily seen in a Markov environment, where optima can be represented by stationary policies, and, as such, prescribe similar actions for generations in similar situations. Conditional on being in a particular state, each generation is treated equally, without there being a dictatorship of the present or the future.

There are a number of possible extensions to this work, and paths leading to future research, including direct incorporation of intergenerational altruism and topics related to sustainability. Our model does not address intergenerational altruism directly. There are no conceptual problems with supposing that the utility and well-being of a generation at time t depends not only on its own actions and state but also on those of temporally nearby generations. Suppose that u_t is a function of (x_{t-1}, a_{t-1}), (x_t, a_t), and (x_{t+1}, a_{t+1}). In this setting, one again picks a policy to optimize the relevant $S(\cdot)$. The difficulty with this approach is that the optimality conditions become more complicated. However, an assumption such as supermodularity in the utilities of the present and the next generation can be expected to push the optima toward sacrificing to avoid intergenerational declines.

One future strand of research that we plan to pursue is to explore the implications of our results in the context of viability and sustainability. In this work, we suggest optimization criteria that have already pushed the expected time till the random end of society, τ, to the limit. This makes it difficult to evaluate policies on the basis of their effect on increases in τ. In general, sustainability analyses ask for policies that will not drive the system out of a certain stable and desirable range, even in the long run. We would like to turn this into an analysis of the "principles" encoded in the general characteristics of optimal policies satisfying such constraints.

[24] Block (2008) discusses the structural flaws and tensions arising from current political economic forces.

Appendix A. Proofs of lemmas

For the proofs using nonstandard analysis, we work in a κ-saturated, nonstandard enlargement of a superstructure $V(Z)$ where the base set, Z, contains $\mathbb{R}$ and ℓ_∞, and κ is a cardinal greater than the cardinality of $V(Z)$. For $E \in V(Z)$, ${}^*E$ denotes its nonstandard version in $V({}^*Z)$. A nearstandard $r \in {}^*\mathbb{R}^k$, ${}^\circ r \in \mathbb{R}^k$ denotes the standard part of r (see Corbae et al., 2009, Ch. 11, Hurd and Loeb, 1985, §II.1 and II.8, or Lindstrøm, 1988, Ch. 3 for this material). For $T \in {}^*\mathbb{N}_0$, we write $T \simeq \infty$ for $T \in {}^*\mathbb{N} \setminus \mathbb{N}$.

For a bounded sequence s_t in $\mathbb{R}$: $\lim_t s_t = s$ if and only if for all $T \simeq \infty$, $s_T \simeq s$; $\liminf_t s_t = s$ if and only if there exists $T' \simeq \infty$ with $s_{T'} \simeq s$ and for all $T \simeq \infty$, ${}^\circ s_T \geq s$; and $\limsup_t s_t = s$ if and only if there exists $T' \simeq \infty$ with $s_{T'} \simeq s$ and for all $T \simeq \infty$, ${}^\circ s_T \leq s$ (see Hurd and Loeb, 1985, Ch. 1, Prop. 8.1, 8.7). For $T < T' \in {}^*\mathbb{N}_0$, let $\eta_{T,T'}$ denote the uniform distribution on the interval $\{T, T+1, \ldots, T'\}$ so that for $\boldsymbol{u} \in \ell_\infty$, $\langle {}^*\boldsymbol{u}, \eta_{T,T'} \rangle = \frac{1}{(T'-T)+1} \sum_{t=T}^{T'} u_t$.

Lemma 9. *We have the following.*

(a) *$\boldsymbol{u}$ is Cesaro summable with long run average* $\mathbf{lra}(\boldsymbol{u})$ *if and only if for all $T \simeq \infty$, $\langle {}^*\boldsymbol{u}, \eta_{0,T} \rangle \simeq$* $\mathbf{lra}(\boldsymbol{u})$.
(b) *$\boldsymbol{u} \in$* **Erg** *has occupation measure $\mu(\cdot|\boldsymbol{u})$ if and only if for all $T' \simeq \infty$ and all continuous bounded $f : \mathbb{R} \to \mathbb{R}$, $\frac{1}{T'+1} \sum_{t=0}^{T'} f(u_t) \simeq \int f(r)\, d\mu(r|\boldsymbol{u})$.*
(c) *N is a null coalition if and only if for all $T \simeq \infty$, $\eta_{0,T}(N) \simeq 0$.*
(d) *B is a substantial coalition if and only if there exists a strictly positive $\underline{b}$ in $\mathbb{R}$ such that for all $T \simeq \infty$, ${}^\circ\eta_{0,T}(B) \geq \underline{b}$.*
(e) *A permutation π is asymptotic if and only if $\frac{|{}^*\pi(T)-T|}{T} \simeq 0$ for all infinite T.*
(f) *For any $\boldsymbol{u} \gg_{fo} \boldsymbol{v}$ in* **W**, *any null N and any $r > 0$, $\boldsymbol{u} \gg_{fo} (\boldsymbol{v} + r1_N)$.*
(g) *For any $\boldsymbol{u} \in$* **W**, *any substantial B and any $r > 0$, $(\boldsymbol{u} + r1_B) \gg_{fo} \boldsymbol{u}$.*
(h) *For every $\epsilon \in \mathbb{R}_{++}$, any $\boldsymbol{u} \in$* **W**, *and any asymptotic π, $\boldsymbol{u} + 2\epsilon 1_{\mathbb{N}_0} \gg_{fo} \boldsymbol{u}^\pi + \epsilon 1_{\mathbb{N}_0} \gg_{fo} \boldsymbol{u}$.*

Proof. Statements (a)–(e) follow immediately from the paragraph preceding the statement of the lemma.

Recall that $\mathbb{F}$ denotes the set of continuously differentiable, strictly increasing functions from $\mathbb{R}$ to $\mathbb{R}$ with derivative uniformly bounded above 0.

For (f), pick $f \in \mathbb{F}$, $\boldsymbol{u} \gg_{fo} \boldsymbol{v}$ in **W**, and null coalition N. For all $T \simeq \infty$,

$$\frac{1}{T+1} \left| \sum_{t=0}^{T} (f(v_t + r1_N(t)) - f(v_t)) \right| \leq r\overline{\delta}_f \frac{1}{T+1} \sum_{t=0}^{T} 1_N(t) \simeq 0 \tag{16}$$

where $\overline{\delta}_f > 0$ is the maximum slope of f over the compact interval $[0, \max(\|\boldsymbol{u}\|, \|\boldsymbol{v}\|)]$.

For (g), pick $f \in \mathbb{F}$, $\boldsymbol{u} \in$ **W**, and substantial coalition B. For any $T \simeq \infty$,

$$\frac{1}{T+1} \sum_{t=0}^{T} (f(u_t + r1_B(t) - f(u_t)) \geq r\underline{\delta}_f \frac{1}{T+1} \sum_{t=0}^{T} 1_B(t) \geq r\underline{\delta}_f \underline{b} \tag{17}$$

where $\underline{\delta}_f > 0$ is the infimum of $f'(x)$ and $\underline{b}$ is from the characterization of substantial coalitions in (d).

For (h), it is sufficient to show that for any $f \in \mathbb{F}$, any $\boldsymbol{u} \in$ **W** and any asymptotic π, and any $T \simeq \infty$, $\frac{1}{T+1} \sum_{t=0}^{T} f(u_t) \simeq \frac{1}{T+1} \sum_{t=0}^{T} f(u_t^\pi)$. To this end, we first show that for any $\boldsymbol{v} \in \ell_\infty$, any asymptotic π, and any $T \simeq \infty$,

$$\left| \frac{1}{T+1} \sum_{t=0}^{T} v_t - \frac{1}{T+1} \sum_{t=0}^{T} v_t^\pi \right| \simeq 0. \tag{18}$$

Let $\mathbb{T} = \{t \in \{0, \ldots, T\} : \pi^{-1}(t) \in \{0, \ldots, T\}\}$. Because π is asymptotic, $T \simeq \infty$ and (the integer part of) $\sqrt{T} \simeq \infty$, $\mathbb{T}$ includes at least the interval $\{(1+\delta)\sqrt{T}, (1-\delta)T\}$ for some infinitesimal $\delta \geq 0$. Using the triangle inequality,

$$\left|\tfrac{1}{T+1}\sum_{t=0}^{T}(v_t - v_t^\pi)\right| \leq \tfrac{1}{T+1}\sum_{t=0}^{T}\left|v_t - v_t^\pi\right| \tag{19}$$

$$< \tfrac{1}{T+1}\sum_{t\in\mathbb{T}}\left|v_t - v_t^\pi\right| + 3\delta\|{}^*\boldsymbol{v}\| \tag{20}$$

$$\simeq 0 + 3\delta\|\boldsymbol{v}\| \simeq 0. \tag{21}$$

To complete the proof, set $v_t = f(u_t)$, and note that $f(u_t + \epsilon) \geq f(u_t) + \underline{\delta}_f \epsilon$ where $\underline{\delta}_f > 0$ denotes $\inf_{x\in\mathbb{R}} f'(x)$. □

Proof of Lemma 1. This follows from Lemma 9 (c) and (d). □

Proof of Lemma 2. For any $x \geq 0$, $S_{\mathcal{C}}(x\boldsymbol{u}) = x \cdot S_{\mathcal{C}}(\boldsymbol{u})$. Let P_0 denote the distribution of X_0, $\int S_{\mathcal{C}}(x\boldsymbol{u})\,dP_0(x) = \int x \cdot S_{\mathcal{C}}(x\boldsymbol{u})\,dP_0(x) = S_{\mathcal{C}}(\boldsymbol{u}) \cdot \int x\,dP_0(x)$. □

Proof of Lemma 3. By the assumption on $\boldsymbol{u}$, for any $\epsilon > 0$, u_t is infinitely often in the interval $(\underline{u}, \underline{u} + \epsilon)$ where $\underline{u}$ is the lower end of the support of $\boldsymbol{u}$'s occupation measure. Because $\boldsymbol{X}$ is i.i.d., on the infinitely many t such that $u_t \in (\underline{u}, \underline{u} + \epsilon)$, X_t is infinitely often in the interval $(a, a + \epsilon)$. Since ϵ was arbitrary, by the representation results for $S_\Delta(\cdot)$ and $S_{\Delta^{pfa}}(\cdot)$ in Theorem C, $E\,S_\Delta(\boldsymbol{u} \odot \boldsymbol{X}) = E\,S_{\Delta^{pfa}}(\boldsymbol{u} \odot \boldsymbol{X}) = a \cdot S_{\mathcal{C}}(\boldsymbol{u})$. For $S_{\mathbb{BM}}(\cdot)$, note that for every T, it is a probability 1 event that, infinitely often, we have $(T+1)$-strings with $u_t \in (\underline{u}, \underline{u} + \epsilon)$ and $X_t \in (a, a + \epsilon)$, so that we also have $S_{\mathbb{BM}}(\boldsymbol{u} \odot \boldsymbol{X}) = a \cdot S_{\mathbb{BM}}(\boldsymbol{u})$.

For $\eta \in \mathfrak{U} \cup \mathcal{P}$, it is a probability 1 event that the realization $\boldsymbol{x} = (x_0, x_1, x_2, \ldots)$ satisfies $\langle \boldsymbol{x}, \eta\rangle = E\,X_0$, and $\boldsymbol{X}$ is independent of $\boldsymbol{u}$. □

Proof of Lemma 4. The arguments are small variants of the previous case. □

Proof of Lemma 5. For every $\boldsymbol{u} \in \mathbf{Erg}$ and every $\eta \in \mathcal{G}$, $\langle \boldsymbol{u}, \eta\rangle = \mathbf{lra}(\boldsymbol{u})$. Therefore, if $p_\alpha(\mathbf{Erg}) = 1$, then for $\mathcal{C} \subset \mathcal{G}$, $\int \min_{\eta\in\mathcal{C}}\langle \boldsymbol{u}, \eta\rangle\,dp_\alpha(\boldsymbol{u}) = \int \mathbf{lra}(\boldsymbol{u})\,dp_\alpha(\boldsymbol{u})$, establishing (a). For (b), $\varphi \circ \boldsymbol{u} = (\varphi(u_0), \varphi(u_1), \ldots) \in \mathbf{Erg}$ so that $\langle \varphi(\boldsymbol{u}), \eta\rangle = \mathbf{lra}(\varphi \circ \boldsymbol{u})$ for every $\eta \in \mathcal{G}$. □

Proof of Lemma 7. This is a restatement of Lemma 9(c). □

Proof of Lemma 8. Let $\boldsymbol{u}' = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{u}^\pi$ for some $0 < \alpha < 1$. Let $L \in \boldsymbol{\partial} S(\boldsymbol{u})$, $L' \in \boldsymbol{\partial} S(\boldsymbol{u}')$ and $L'' \in \boldsymbol{\partial} S(\boldsymbol{u}^\pi)$. We have

$$S(\boldsymbol{u}) + L(\boldsymbol{u}' - \boldsymbol{u}) \geq S(\boldsymbol{u}'), \tag{22}$$

$$S(\boldsymbol{u}') + L'(\boldsymbol{u}^\pi - \boldsymbol{u}') \geq S(\boldsymbol{u}^\pi), \text{ and} \tag{23}$$

$$S(\boldsymbol{u}^\pi) + L''(\boldsymbol{u} - \boldsymbol{u}^\pi) \geq S(\boldsymbol{u}). \tag{24}$$

Since L, L', and L'' are invariant under π and linear, $L(\boldsymbol{u}' - \boldsymbol{u}) = L'(\boldsymbol{u}^\pi - \boldsymbol{u}') = L''(\boldsymbol{u} - \boldsymbol{u}^\pi) = 0$. This yields $S(\boldsymbol{u}) \geq S(\boldsymbol{u}') \geq S(\boldsymbol{u}^\pi) \geq S(\boldsymbol{u})$. □

Proof of Lemma 6. Follow the proof of Theorem B (below) with $\mathcal{C}$ in the place of $\mathbb{V}$. □

Appendix B. Proofs of theorems and propositions

Proof of Theorem A. Suppose that $\succ$ satisfies Postulates I–V.

Fishburn shows that if $\succ$ satisfies his A1, A2, A3, A4*, A5* and A0.2, then there exists an integrable $S : \mathbf{W} \to \mathbb{R}$ such that $p \succ q$ if and only if $\int S(\boldsymbol{v})\,dp(\boldsymbol{v}) > \int S(\boldsymbol{v})\,dq(\boldsymbol{v})$ (Fishburn, 1982, Theorem 4, Ch. 3). Our Postulates I and II are Fishburn's A1 and A2, our Postulate III is a strong form of his A3 and it directly implies his A4* and A5*. Our mixture set, $\mathcal{M}$, the set of measures with bounded support is closed under finite convex combinations. As we work with the Borel σ-field on $\mathbf{W}$, our Postulate III implies that the domain for our probabilities contains all preference intervals, which then implies that $\mathcal{M}$ is closed under taking conditional measures on preference intervals, completing the verification of Fishburn's A0.2.

Restricted to the closed set of point masses, $\{\delta_{\boldsymbol{u}} : \boldsymbol{u} \in \mathbf{W}\}$, Postulate III implies that $S(\cdot)$ is continuous. By considering measures with two point supports, Postulate IV implies that $S(\cdot)$ is concave. For the normalization that $S : \mathbf{W} \to [0, \infty)$, note that there is no loss in setting $S(\mathbf{0}) = 0$ — for any $\boldsymbol{u} \in \mathbf{W}$, $(\boldsymbol{u} + \epsilon 1_{\mathbb{N}_0}) \ggcurly_{fo} \mathbf{0}$, sending $\epsilon \downarrow 0$ and appealing to continuity, $S(\boldsymbol{u}) \geq S(\mathbf{0})$.

We now show that $S(\cdot)$ is Pareto responsive. By Lemma 9(f), for any null coalition N and any $\epsilon > 0$,

$$(\boldsymbol{u} + 2\epsilon 1_{\mathbb{N}_0}) \ggcurly_{fo} (\boldsymbol{u} + r 1_N + \epsilon 1_{\mathbb{N}_0}) \ggcurly_{fo} \boldsymbol{u}. \tag{25}$$

By Postulate V,

$$S(\boldsymbol{u} + 2\epsilon 1_{\mathbb{N}_0}) > S(\boldsymbol{u} + r 1_N + \epsilon 1_{\mathbb{N}_0}) > S(\boldsymbol{u}). \tag{26}$$

By continuity, as $\epsilon \downarrow 0$, we find that $S(\boldsymbol{u}) \geq S(\boldsymbol{u} + r 1_N) \geq S(\boldsymbol{u})$.

Suppose now that B is a substantial coalition. By Lemma 9(g), for any $r > 0$, $(\boldsymbol{u} + r 1_B) \ggcurly_{fo} \boldsymbol{u}$. By Postulate V, $S(\boldsymbol{u} + r 1_B) > S(\boldsymbol{u})$.

We now show that $S(\cdot)$ is patient. By Lemma 9(h), for every $\epsilon \in \mathbb{R}_{++}$, any $\boldsymbol{u} \in \mathbf{W}$, and any asymptotic π, $\boldsymbol{u} + 2\epsilon 1_{\mathbb{N}_0} \ggcurly_{fo} \boldsymbol{u}^\pi + \epsilon 1_{\mathbb{N}_0} \ggcurly_{fo} \boldsymbol{u}$. Since $\mathbb{N}_0$ is a substantial coalition, the previous step implies that $S(\boldsymbol{u} + 2\epsilon 1_{\mathbb{N}_0}) > S(\boldsymbol{u}^\pi + \epsilon 1_{\mathbb{N}_0}) > S(\boldsymbol{u})$. By continuity, as $\epsilon \downarrow 0$, we find that $S(\boldsymbol{u}) \geq S(\boldsymbol{u}^\pi) \geq S(\boldsymbol{u})$. □

Proof of Theorem B. Suppose first that $S : \mathbf{W} \to [0, \infty)$ is continuous, concave, and Pareto responsive, that $\boldsymbol{u}$ is interior, and that $L \in \boldsymbol{\partial} S(\boldsymbol{u})$. We must show that L is Pareto responsive.

Suppose first that N is a null set. Because $\boldsymbol{u}$ is interior, there exists an $r' > 0$ such that $\boldsymbol{u} - r' 1_N$ is interior. Because $S(\cdot)$ is Pareto responsive, for all $r > 0$, $S((\boldsymbol{u} - r' 1_N) + r 1_N) = S(\boldsymbol{u} - r' 1_N)$. By considering the $r \in (0, 2r')$, we know that $S(\cdot)$ is constant on the line joining $(\boldsymbol{u} - r' 1_N)$ and $(\boldsymbol{u} + r' 1_N)$. Therefore any tangent to $S(\cdot)$ at $\boldsymbol{u}$, the mid-point of the line, must be constant on the line, and this implies that $L(1_N) = 0$ for any $L \in \boldsymbol{\partial} S(\boldsymbol{u})$.

Now suppose that B is substantial. Picking $r' > 0$ so that $\boldsymbol{u} - r' 1_B$ is interior and using the Pareto responsiveness of $S(\cdot)$, we know that

$$S(\boldsymbol{u} - r' 1_B) < S(\boldsymbol{u}) < S(\boldsymbol{u} + r' 1_B). \tag{27}$$

If $L(1_B) \leq 0$, we contradict $S(\boldsymbol{u}) < S(\boldsymbol{u} + r' 1_B)$.

Now suppose that $S : \mathbf{W} \to [0, \infty)$ is $\mathbb{V}$-concave. We must show that $S(\cdot)$ is Pareto responsive.

Pick arbitrary $\boldsymbol{u} \in \mathbf{W}$, arbitrary null coalition N, and arbitrary $r > 0$. By the definition of the supgradient, for any $L \in \boldsymbol{\partial} S(\boldsymbol{u} + r 1_N)$,

$$S(\boldsymbol{u} + r 1_N) + L(\boldsymbol{u} - (\boldsymbol{u} + r 1_N)) \geq S(\boldsymbol{u}). \tag{28}$$

In a similar fashion, for any $L' \in \boldsymbol{\partial} S(\boldsymbol{u})$,

$$S(\boldsymbol{u}) + L'((\boldsymbol{u} + r1_N) - \boldsymbol{u}) \geq S(\boldsymbol{u} + r1_N). \tag{29}$$

Since $L, L' \in \mathbb{V}$, $L(-r1_N) = 0$ and $L'(r1_N) = 0$. Combining (28) and (29), $S(\boldsymbol{u} + r1_N) \geq S(\boldsymbol{u}) \geq S(\boldsymbol{u} + r1_N)$.

Pick arbitrary $\boldsymbol{u} \in \mathbf{W}$, arbitrary substantial coalition B, and arbitrary $r > 0$. By the definition of the supgradient, for any $L \in \boldsymbol{\partial} S(\boldsymbol{u} + r1_B)$,

$$S(\boldsymbol{u} + r1_B) + L(\boldsymbol{u} - (\boldsymbol{u} + r1_B)) \geq S(\boldsymbol{u}). \tag{30}$$

Since B is substantial and $L \in \mathbb{V}$, $L(-r1_B) < 0$. Therefore, $S(\boldsymbol{u} + r1_B) > S(\boldsymbol{u})$. □

We now give an example of a social welfare function satisfying our Postulates with tangents that are not invariant even to bounded shifts, hence fail to be patient.

Example B.1. Let η be the weak* standard part of the uniform distributions $\eta_{0,T}$ for $T = 2T' + 1$, $T' \simeq \infty$; let $\varphi(r) = 2\sqrt{r}$; let $S(\boldsymbol{u}) = \langle \varphi \circ \boldsymbol{u}, \eta \rangle = \int \varphi(u_t)\, d\eta(t)$; at $\boldsymbol{u} = (1, 9, 1, 9, 1, 9, \ldots)$, the sequence $(\varphi'(1), \varphi'(9), \varphi'(1), \varphi'(9), \ldots)$ is equal to $(1, \frac{1}{3}, 1, \frac{1}{3}, \ldots)$; $L \in \boldsymbol{\partial} S(\boldsymbol{u})$ if and only if $L(\boldsymbol{x}) = \langle \boldsymbol{x}, \alpha \rangle$ where $\alpha = \mathrm{st}\,(\alpha')$ and α' is the vector in $\frac{2}{3} \cdot {}^*P$ given by

$$\alpha_t = \begin{cases} \frac{1}{(T+1)} & \text{for } t = 0, 2, 4, \ldots, T-1, \\ \frac{1}{3(T+1)}) & \text{for } t = 1, 3, 4, \ldots, T, \text{ and} \\ \alpha_t = 0 & \text{otherwise.} \end{cases} \tag{31}$$

$L(\cdot)$ is not invariant to finite shifts because $L(1_{Ev}) = \frac{1}{2} \neq \frac{1}{6} = L(1_{Odd})$ where Ev is the set of even integers and Odd the set of odd integers. Alternatively, from Robinson (1964, Theorem 3.6) $L(\cdot)$ is not invariant to finite shifts because $\sum_{t=0}^{T} |\alpha_{t+1} - \alpha_t| = (T+1)\frac{2}{3(T+1)} = \frac{2}{3} \not\simeq 0$.

Proof of Theorem C. Let $P = \{\boldsymbol{\alpha} \in \ell_1 : (\forall t \in \mathbb{N}_0)[\alpha_t \geq 0 \text{ and } \sum_t \alpha_t = 1\}$. For $\eta' \in {}^*P$, the weak*-star standard part of η' is the probability $\eta = \mathrm{st}\,(\eta')$ defined by $\langle \boldsymbol{u}, \eta \rangle = {}^\circ\langle {}^*\boldsymbol{u}, \eta' \rangle$. The probability $\mathrm{st}\,(\eta')$ is purely finitely additive if and only if for all (necessarily finite) $T \in \mathbb{N}_0$, $\eta'(\{T+1, T+2, \ldots\}) \simeq 1$. From Robinson (1964, Theorem 4.1) $\mathrm{st}\,(\cdot)$ is onto Δ.

For (a), that $\inf_t u_t = S_\Delta(\boldsymbol{u}) = \min\{\langle \boldsymbol{u}, \eta \rangle : \eta \in \Delta\}$, pick arbitrary $\boldsymbol{u}$ and let $\underline{u} = \inf_t u_t$. For any $\eta \in \Delta$, $\langle \boldsymbol{u}, \eta \rangle \geq \underline{u}$. To show that equality is achieved, let $n \mapsto t_n$ be a (perhaps constant) sequence in $\mathbb{N}_0$ such that $u_{t_n} \to \inf_t u_t$. For any $n \simeq \infty$, $u_{t_n} \simeq \underline{u}$. If η' is point mass on t_n and $\eta = \mathrm{st}\,(\eta')$, then $\langle \boldsymbol{u}, \eta \rangle \simeq \langle {}^*\boldsymbol{u}, \eta' \rangle \simeq \underline{u}$.

For (b), that $\liminf_t u_t = S_{\Delta^{pfa}}(\boldsymbol{u}) = \min\{\langle \boldsymbol{u}, \eta \rangle : \eta \in \Delta^{pfa}\}$, pick arbitrary $\boldsymbol{u}$ and let $\underline{u} = \liminf_t u_t$. For $T \in \mathbb{N}_0$, let $U_T = \inf\{u_t : t > T\}$. If $\eta'(\{T+1, T+2, \ldots\}) \simeq 1$ for $\eta' \in {}^*P$, then ${}^\circ\langle \boldsymbol{u}, \mathrm{st}\,(\eta') \rangle \geq U_T$ and $U_T \uparrow \underline{u}$ so that $\langle \boldsymbol{u}, \eta \rangle \geq \underline{u}$ for all purely finitely additive η. To show that equality is achieved, let $n \mapsto t_n$ be an unbounded sequence with $u_{t_n} \to \underline{u}$. If η' is point mass on t_n for some $n \simeq \infty$, then $\langle \boldsymbol{u}, \eta \rangle = \underline{u}$ where $\eta = \mathrm{st}\,(\eta')$.

For (c), pick arbitrary $\boldsymbol{u}$, let $\underline{u} = \liminf_{T \uparrow \infty} \inf_{j \geq 0} \frac{1}{T+1} \sum_{t=0}^{T} u_{j+t}$. We must show that $\underline{u} = S_{\mathbb{BM}}(\boldsymbol{u}) = \min\{\langle \boldsymbol{u}, \eta \rangle : \eta \in \mathbb{BM}\}$. For each N, let $U_N = \inf_{(T'-T) \geq N} \langle \boldsymbol{u}, \eta_{T,T'} \rangle$ so that $U_N \to \underline{u}$. For each N, there exists a pair T'_N, T_N with $(T'_N - T_N) \geq N$ and $|U_N - \langle \boldsymbol{u}, \eta_{T_N, T'_N} \rangle| < 1/N$. Taking $N \simeq \infty$, $\eta := \mathrm{st}\,(\eta_{T_N, T'_N})$ is a Banach–Mazur limit and $\langle \boldsymbol{u}, \eta \rangle \simeq \langle {}^*\boldsymbol{u}, \eta_{T_N, T'_N} \rangle \simeq \underline{u}$ so the equality is achieved for at least one $\eta \in \mathbb{BM}$. To show that $\langle \boldsymbol{u}, \eta \rangle \geq \underline{u}$ for each $\eta \in \mathbb{BM}$, note that $\mathbb{BM}$ is a weak* compact, convex subset of Δ, hence the minimum of the continuous linear function $\eta \mapsto \langle \boldsymbol{u}, \eta \rangle$ achieves its minimum on an extreme point of $\mathbb{BM}$. From Keller and Moore (1992, Theorem 3.1), any extreme point of $\mathbb{BM}$ has an expression as $\mathrm{st}(\eta')$ where η' is the uniform distribution on some interval $\{T, \ldots, T'\} \subset {}^*\mathbb{N}_0$ with $(T' - T) \simeq \infty$.

For (d), repeat the previous using $U_\epsilon := \liminf_T \langle \boldsymbol{u}, \eta_{(1-\epsilon)T,T} \rangle$.

For (e), repeat the previous using $U_T := \inf_{T'>T} \langle \boldsymbol{u}, \eta_{0,T'} \rangle$. □

Proof of Proposition 1. After setting $S_\varphi(\boldsymbol{u}) = {}^\circ \frac{1}{T+1} \sum_{t=0}^{T} \varphi(u_t)$ for some $T \simeq \infty$, this follows from Lemma 9(b). □

Proof of Proposition 2. As this is an exchange economy model, we need only verify (i)–(iv) in Bewley's Theorem 1. (i) is the assumption that the consumption sets are convex and Mackey closed, which is immediate. (ii) is the assumption that the preference relations are transitive, reflexive and complete, which is satisfied because the preferences are given by utility functions. (iii) is the assumption that for all $i \in I$ and all consumption vectors, $\boldsymbol{x}$, the set $\{z \in W^k : U_i(z) \geq U_i(\boldsymbol{x})\}$ is convex and Mackey closed. Convexity follows from the concavity of $U_i(\cdot)$, and a convex subset of the dual of a Banach space is closed for all the topologies between the weak*-topology and the norm topology if and only if it is norm closed (Dunford and Schwartz, 1988, Cor. V.2.14). Therefore, norm continuity of the $U_i(\cdot)$ delivers the necessary closure. (iv) is the assumption that for all $i \in I$ and all consumption vectors, $\boldsymbol{x}$, the set $\{z \in W^k : U_i(z) \leq U_i(\boldsymbol{x})\}$ is norm closed, which follows directly from the norm continuity of $U_i(\cdot)$. □

Proof of Proposition 3. This is a standard general equilibrium argument based on the equivalence of the weak* and norm closure of convex sets e.g. (Dunford and Schwartz, 1988, Cor. V.2.14). □

References

Arapostathis, Aristotle, Borkar, Vivek S., Fernández-Gaucherand, Emmanuel, Ghosh, Mrinal K., Marcus, Steven I., 1993. Discrete-time controlled Markov processes with average cost criterion: a survey. SIAM J. Control Optim. 31 (2), 282–344.

Araujo, Aloisio, 1985. Lack of Pareto optimal allocations in economies with infinitely many commodities: the need for impatience. Econometrica 53 (2), 455–461.

Asheim, Geir B., 1988. Rawlsian intergenerational justice as a Markov-perfect equilibrium in a resource technology. Rev. Econ. Stud. 55 (3), 469–483.

Asheim, Geir B., 1991. Unjust intergenerational allocations. J. Econ. Theory 54 (2), 350–371.

Asheim, Geir B., 2010. Intergenerational equity. Annu. Rev. Econ. 2 (1), 197–222.

Atkinson, Anthony B., 2001. The strange disappearance of welfare economics. Kyklos 54 (2–3), 193–206.

Basu, Kaushik, Mitra, Tapan, 2003. Aggregating infinite utility streams with intergenerational equity: the impossibility of being Paretian. Econometrica 71 (5), 1557–1563.

Basu, Kaushik, Mitra, Tapan, 2007. Possibility theorems for equitably aggregating infinite utility streams. In: Intergenerational Equity and Sustainability. Springer, pp. 69–84.

Bewley, Truman F., 1972. Existence of equilibria in economies with infinitely many commodities. J. Econ. Theory 4 (3), 514–540.

Blackorby, Charles, Bossert, Walter, Donaldson, David, 1995. Intertemporal population ethics: critical-level utilitarian principles. Econometrica 63 (6), 1303–1320.

Block, Fred, 2008. Swimming against the current: the rise of a hidden developmental state in the United States. Politics Soc. 36 (2), 169–206.

Bossert, Walter, Sprumont, Yves, Suzumura, Kotaro, 2007. Ordering infinite utility streams. J. Econ. Theory 135 (1), 579–589.

Brock, William A., 1970. An axiomatic basis for the Ramsey–Weizsäcker overtaking criterion. Econometrica 38 (6), 927–929.

Brock, William A., Haurie, Alain B., 1976. On existence of overtaking optimal trajectories over an infinite time horizon. Math. Oper. Res. 1 (4), 337–346.

Brown, Donald J., Lewis, Lucinda M., 1981. Myopic economic agents. Econometrica 49 (2), 359–368.

Chakravarty, Sukhamoy, 1962. The existence of an optimum savings program. Econometrica 30 (1), 178–187.

Chambers, Christopher P., 2012. Inequality aversion and risk aversion. J. Econ. Theory 147 (4), 1642–1651.

Chambers, Christopher P., Echenique, Federico, 2012. When does aggregation reduce risk aversion? Games Econ. Behav. 76 (2), 582–595.

Chichilnisky, Graciela, 1996. An axiomatic approach to sustainable development. Soc. Choice Welf. 13 (2), 231–257.

Choquet, Gustave, 1969. In: Marsden, J., Lance, T., Gelbart, S. (Eds.), Lectures on Analysis. Vol. II: Representation Theory. W. A. Benjamin, Inc., New York–Amsterdam.

Collard, David, 1996. Pigou and future generations: a Cambridge tradition. Camb. J. Econ. 20 (5), 585–597.

Corbae, Dean, Stinchcombe, Maxwell B., Zeman, Juraj, 2009. An Introduction to Mathematical Analysis for Economic Theory and Econometrics. Princeton University Press, Princeton, N.J.

Costanza, Robert, D'Arge, Ralph, De Groot, Rudolf, Farber, Stephen, Grasso, Monica, Hannon, Bruce, Limburg, Karin, Naeem, Shahid, Robert, V.O'Neill, Paruelo, Jose, Raskin, Robert G., Sutton, Paul, Van, Den B., 1997. The value of the world's ecosystem services and natural capital. Nature 387 (6630), 253–260.

David, Paul A., 2004. Understanding the emergence of 'open science' institutions: functionalist economics in historical context. Ind. Corp. Change 13 (4), 571–589.

Denardo, Eric V., Miller, Bruce L., 1968. An optimality condition for discrete dynamic programming with no discounting. Ann. Math. Stat. 39, 1220–1227.

Deutsch, David, 2011. The Beginning of Infinity: Explanations that Transform the World. Viking, New York, NY.

Diamond, Peter A., 1965. The evaluation of infinite utility streams. Econometrica 33 (1), 170–177.

Dunford, Nelson, Schwartz, Jacob T., 1988. Linear Operators. Part I: General Theory. Wiley Classics Library. John Wiley & Sons, Inc., New York. With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

Ehrlich, Paul R., Ehrlich, Anne H., 1981. Extinction: The Causes and Consequence of the Disappearance of Species. Random House, New York.

Epstein, Larry G., Zin, Stanley E., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica, 937–969.

Feinberg, Eugene A., Kasyanov, Pavlo O., Zadoianchuk, Nina V., 2012. Average cost Markov decision processes with weakly continuous transition probabilities. Math. Oper. Res. 37 (4), 591–607.

Fishburn, Peter C., 1982. The Foundations of Expected Utility. Theory and Decision Library, vol. 31. D. Reidel Publishing Co., Dordrecht.

Fleurbaey, Marc, Michel, Philippe, 2003. Intertemporal equity and the extension of the Ramsey criterion. J. Math. Econ. 39 (7), 777–802.

Pope Francis, 2015. Encyclical Letter Laudato Si' of the Holy Father Franci on Care for our Common Home. Vatican Press, The Holy See.

Gale, David, 1967. On optimal development in a multi-sector economy. Rev. Econ. Stud. 34 (1), 1–18.

Gollier, Christian, Hammitt, James K., 2014. The long-run discount rate controversy. Annu. Rev. Resour. Econ. 6 (1), 273–295.

Hildenbrand, Werner, 1969. Pareto optimality for a measure space of economic agents. Int. Econ. Rev. 10 (3), 363–372.

Hurd, Albert E., Loeb, Peter A., 1985. An Introduction to Nonstandard Real Analysis. Pure and Applied Mathematics, vol. 118. Academic Press Inc., Orlando, FL.

Jaśkiewicz, Anna, Nowak, Andrzej S., 2006. On the optimality equation for average cost Markov control processes with Feller transition probabilities. J. Math. Anal. Appl. 316 (2), 495–509.

Keller, Gordon, Moore Jr., Lawrence C., 1992. Invariant means on the group of integers. In: Analysis and Geometry. Bibliographisches Inst., Mannheim, pp. 1–18.

Kreps, David M., Porteus, Evan L., 1978. Temporal resolution of uncertainty and dynamic choice theory. Econometrica, 185–200.

Lauwers, Luc, 1998. Intertemporal objective functions: strong Pareto versus anonymity. Math. Soc. Sci. 35 (1), 37–55.

Lindstrøm, Tom, 1988. An invitation to nonstandard analysis. In: Nonstandard Analysis and Its Applications. In: London Math. Soc. Stud. Texts, vol. 10. Cambridge Univ. Press, Cambridge, pp. 1–105.

Marinacci, Massimo, 1998. An axiomatic approach to complete patience and time invariance. J. Econ. Theory 83 (1), 105–144.

Mazzucato, Mariana, 2015. The Entrepreneurial State: Debunking Public vs. Private Sector Myths. Anthem Press, London.

McKenzie, Lionel W., 1976. Turnpike theory. Econometrica 44 (5), 841–865.

Meyer, Paul-André, 1973. Limites médiales d'après Mokobodzki. In: Séminaire de Probabilités, VII. Univ. Strasbourg, Année Universitaire 1971–1972. In: Lecture Notes in Math., vol. 321. Springer, Berlin, pp. 198–204.

Meyn, Sean P., 1997. The policy iteration algorithm for average reward Markov decision processes with general state space. IEEE Trans. Autom. Control 42 (12), 1663–1680.

Meyn, Sean P., Tweedie, Richard L., 2012. Markov Chains and Stochastic Stability. Springer Science and Business Media.

Pigou, Arthur C., 1912. Wealth and Welfare. Macmillan, London.

Pigou, Arthur C., 1920. The Economics of Welfare. Macmillan, London.

Ramsey, Frank Plumpton, 1928. A mathematical theory of saving. Econ. J. 38, 543–559.

Robinson, Abraham, 1964. On generalized limits and linear functionals. Pac. J. Math. 14, 269–283.

Sidgwick, Henry, 1874. The Methods of Ethics. Macmillan & Co., London.

Sidgwick, Henry, 1883. The Principles of Political Economy. Macmillan, London, New York.

Sunstein, Cass R., 2002. The paralyzing principle. Regul. 25, 32–37.

de Vauban, Sèbastien Le Prestre, 1910. Traité de la culture des forêts. In: Vauban, sa famille et ses écrits, ses oisivetés et sa correspondance: analyse et extraits, vol. 2. Berger-Levrault, Paris.

de Vauban, Sèbastien Le Prestre, 2007. Traité de la culture des forêts. In: Vérin, Hélène (Ed.), Édition intégrale. Les Oisivetés de Monsieur de Vauban. CDHTE-Cnam, SeaCDHTE, Seyssel, Champ Vallon.

Veinott, Arthur F., 1966. On finding optimal policies in discrete dynamic programming with no discounting. Ann. Math. Stat. 37 (5), 1284–1294.

Von Weizsäcker, Carl Christian, 1965. Existence of optimal programs of accumulation for an infinite time horizon. Rev. Econ. Stud. 32 (2), 85–104.

Zame, William R., 2007. Can intergenerational equity be operationalized? Theor. Econ. 2, 187–202.