THE PARETO COMPARISONS OF A GROUP OF EXPONENTIAL DISCOUNTERS

CHRISTOPHER P. CHAMBERS AND FEDERICO ECHENIQUE

Abstract. Agents with different discount factors disagree about some intertemporal tradeoffs, but they will also agree sometimes. We seek to understand precisely the nature of their agreements and disagreements.

A group of agents is identified with a set of discount factors. We characterize the comparisons that a given interval of discount factors will agree on, including what all discount factors in the interval $[0, 1]$ will agree on. Our result is analogous to how all risk-averse and monotone agents agree on mean-preserving spreads. We also characterize the comparisons that are consistent with some set of discount factors, when the set is not known or exogenously given. In other words, we describe the Pareto comparisons that are consistent with a society, or group, of exponentially discounting agents.

(Chambers) Department of Economics, Georgetown University
(Echenique) Division of the Humanities and Social Sciences, California Institute of Technology
Most of the results in this paper appeared first in the working paper version of "On multiple discount rates" (see Caltech SSWP 1418; first posted May 2016). We are grateful to Luke Boosey, Simone Cerreia-Vioglio, Vijay Krishna, Efe Ok, Phil Reny, Itai Sher, Tomasz Strzalecki, and participants of numerous seminars and conferences where we have presented the paper, for comments. Echenique thanks the National Science Foundation for its support through the grants SES-1558757 and CNS-151894.

1. Introduction

A group of agents with different discount factors will disagree sometimes, and agree some times. Some intertemporal tradeoffs will be desirable to all agents in the group, while some tradeoffs will only be desirable to a strict subset of agents. The point of this paper is to characterize these areas of agreement and disagreement.

We assume that all agents rank streams of utils in discrete time. We assume there is no disagreement in what constitutes a util, and that a stream of utils should be evaluated by discounting exponentially. Rather, all disagreements arise from disagreements in the appropriate discount rate to be used. In this environment, we seek to understand all possible rankings of streams on which all agents will agree. Our paper provides a characterization of such objects in an infinite-horizon framework.

There are two natural approaches. A first approach envisions that the discount factors held by members of the group are given and known. For a given set of discount factors, we seek to understand the comparisons that all members of the group would agree on. The second approach is an investigation of the Pareto relations themselves: Which orderings of utility streams constitute the Pareto relations for some group of agents. We proceed to discuss our findings as they relate to these two approaches.

As said, our first approach to the question takes the set of discount factors as given. In this exercise, we would like a simple condition on pairs of streams which would allow us to conclude directly whether one stream is at least as good as the other for all discount factors in the set. For simplicity, we assume that the discount factors of the members of the group correspond to a closed interval in $[0, 1]$.[1] One method would be to simply check, for each discount factor in the set, whether one stream is at least as good as the other. By contrast, we provide a "dual" method, whereby one stream is at least as good as another if and only if the first can be (approximately) arrived from the second by a sequence of transformations. In many ways, the exercise here is analogous

[1] This can be significantly generalized, but the statement is already analytically cumbersome.

to the classical results on risk aversion and mean preserving spreads.[2] A lottery is preferred to another by all risk averse expected utility maximizers if and only if the first can be (approximately) arrived at from the second by a sequence of elementary mean-preserving spreads.[3]

We are not the first to investigate this type of question, and there is a substantive literature axiomatizing Pareto relations for intertemporal choice. For example, Pratt and Hammond (1979); Bøhren and Hansen (1980); Ekern (1981); Trannoy (1999); Foster and Mitra (2003); Bastianello and Chateauneuf (2016) describe and axiomatize these objects for different classes of discounters, and different domains of consumptions streams.[4] In particular, Bøhren and Hansen (1980); Trannoy (1999); Foster and Mitra (2003); Bastianello and Chateauneuf (2016) consider related problems, but come up with a "primal" axiomatization, whereas ours is "dual" in a formal sense. Our characterization is closer in spirit to the papers on mean-preserving spreads, Rothschild and Stiglitz (1970) and Blackwell (1953), than to the work on intertemporal choice.

Our decomposition results from a natural recursive application of three basic properties of discounting. For any discounter, shifting a util from tomorrow to today is better than doing nothing. This tells us that a stream in which tomorrow's utility is -1 and today's is 1, and all other generations have 0 utility is better than an environment in which all generations have 0 utility. But now, by discounting, we also know that the stream in which tomorrow's utility is -1 and today's is 1 is at least as good as the stream where tomorrow's utility is 1 and the day after tomorrow's is -1. Using the linearity (in streams) of the discounting structure, this allows us to claim that the stream in which -11 unit is consumed today and -1 unit the day after tomorrow, with 2 units tomorrow is better than a utility of 0 throughout. By applying this operation recursively, we get many streams which should dominate the null stream. Our

[2]The literature initiated in economics by Rothschild and Stiglitz (1970). The mathematical results go back at least to Blackwell (1953), where an experiment in that context is a lottery over lotteries.

[3]In fact, there is a clear technical connection with these works as well, which we will explain below.

[4]Our results rely on an application of the Hausdorff moment problem. Other applications in economics include Hara (2008) and Minardi and Savochkin (2016), who use the continuous version.

contribution is to show that any stream at least as good as the null stream can be arrived at, arbitrarily closely, by applying such operations a finite number of times.

Our second approach assumes that discount factors are not exogenously specified, but rather identifies the collective conditions satisfied by *all* Pareto relations generated by exponential discounters. That is, we elicit a list of properties which are satisfied by a relation if and only if that relation could be the Pareto relation for some collection of exponential discounters. The characterization is related to a companion paper of ours, Chambers and Echenique (2016), in which we focus on complete preferences over utility streams. Here, we establish that a certain weakening of a stationarity axiom of Koopmans (1960) is the driving force behind Pareto relations. Our axiom imagines a constant stream of payoffs; constancy of a stream reflects a sequence of payoffs which is "time-invariant." This constant stream is to be understood as a kind of baseline alternative. Our property roughly states that a stream is at least as good as the constant stream if and only if a delayed version of the stream (where the initial segment is replaced by the baseline outcome) is also at least as good as the constant stream.[5]

Together with a standard additivity axiom (all exponential discounters have additive preferences over util streams, hence so does a Pareto relation) and some other mild technical conditions, this property effectively characterizes the implications of the Pareto model for some closed and nonempty set of discount factors. In line with the exogenous discount factor analysis, we also seek to understand when this set of discount factors is an interval. A characterization of such relations is available, via a tradeoff axiom that we introduce.

The tradeoff property considers a certain type of "bad" stream, or at least one which is not good, in the sense that one would not choose to add it to a status quo. Now, this stream is one which can be "shifted forward," bringing the bad outcome earlier in time. In principle, an agent would be willing to do this in the case that the bad stream is also simultaneously deflated by a small enough amount. The tradeoff axiom states that a certain deflator cannot be

[5]Technically, the stationarity property also requires mixtures of the delayed stream and the constant stream to be considered.

considered small for this particular stream. This certain deflator is one which would not be considered small enough to deflate for a loss of one deflated util today to be replaced by a gain of one full util tomorrow.

Finally, we also investigate a class of more general binary relations over util streams. It is of interest to ask whether there is a "maximal" subrelation of the Pareto type of the given relation. For example, suppose we were given a complete ranking over util streams. We would want to know whether this ranking over streams "could be" a social welfare ranking for some collection of exponential discounters, and if so, what the set of discount factors is. The maximal subrelation serves as such a collection of discount factors. One stream which dominates another for every discounter in the (endogenously derived set) would be deemed better for the preference relation, by the standard pareto property. Hence, this maximal set constitutes the entire class of potential exponential discounters whose opinions might be reflected in a deliberation on util streams.

The paper proceeds linearly, discussing each of the preceding results sequentially. Proofs are in an appendix.

2. The Model

We study the problem of choosing among intertemporal streams of utils. The objects of choice are sequences of real numbers $x = (x_t)_{t=0}^{\infty}$. These are restricted to lie in a set of bounded sequences $X \subseteq \ell_\infty$. Interpret a sequence x as a *stream of utils*, meaning that x_t is the utility received at time t. For some of the results in our paper, we shall take $X = \ell_1$, the space of all absolutely summable sequences. For other results we shall assume that $X = \ell_\infty$.

A collection of agents is modeled through a set $D \subseteq (0,1)$. Each of the agents discounts utility exponentially. So $\delta \in D$ evaluates a stream x as $\sum_{t=0}^{\infty} \delta^t x_t$. Given a set D, the *Pareto* ordering on X is defined by $x \succeq^D y$ iff $\sum_{t=0}^{\infty} \delta^t x_t \geq \sum_{t=0}^{\infty} \delta^t y_t$ for all $\delta \in D$.

We consider two types of questions in the paper. First, when D is given exogenously, we want to characterize, or describe, the ordering $\succeq^D$. Second, given an ordering $\succeq$ over X, we want to understand when there exists a $D \subseteq (0,1)$ such that $\succeq = \succeq^D$.

2.1. **Notational conventions.** The sequence $(1, 1, \ldots)$, which is identically 1, is denoted by $\mathbf{1}$. When $\theta \in \mathbf{R}$ is a scalar we often abuse notation and use θ to denote the constant sequence $\theta\mathbf{1}$. If x is a sequence, we denote by (θ, x) the concatenation of θ and x: the sequence (θ, x) takes the value θ for $t = 0$, and then x_{t-1} for each $t \geq 1$. Similarly, the sequence

$$(\underbrace{\theta, \ldots, \theta}_{T \text{ times}}, x)$$

takes the value θ for $t = 0, \ldots, T-1$ and x_{t-T} for $t \geq T$.

The notation for inequalities of sequences is: $x \geq y$ if $x_t \geq y_t$ for all $t \in \mathbf{N}$, $x > y$ if $x \geq y$ and $x \neq y$, and $x \gg y$ if $x_t > y_t$ for all $t \in \mathbf{N}$.

Finally, for $\delta \in (0, 1)$, let $m(\delta)$ denote the sequence in ℓ_1 where $m(\delta)_t = \delta^t$.

3. The Pareto relation with exogenous D.

First we seek to understand the comparisons of streams that *all* discount factors must agree on: the Pareto relation when the set of discount factors is $D = [0, 1]$.

Given that we allow for $\delta = 1$, we work with $X = \ell_1$. Our set of choice objects is the set of absolutely summable sequences. Observe that $\succeq^{[0,1]}$ is well-defined as $\sum_t \delta^t x_t \in \mathbf{R}$ for all $\delta \in [0, 1]$ and $x \in \ell_1$.

We can gain some insight as to the structure of $\succeq^{[0,1]}$ from four seemingly trivial observations:

(1) $(1, 0, 0, \ldots) \succeq^{[0,1]} 0$
(2) If $x \succeq^{[0,1]} 0$, then $x \succeq^{[0,1]} (0, x) \succeq^{[0,1]} 0$
(3) If $x \succeq^{[0,1]} y$, then $(x - y) \succeq^{[0,1]} 0$.

Statement 1 is simply a very weak implication of the claim that all exponential discounters like more consumption to less. Statement 2 is the essence of discounting: if a stream is "good," in the sense that it is at least as good as 0, then shifting its start date back a period cannot improve on the stream, but also cannot render the stream a "bad." Finally, statement 3 reflects that discounting is linear in consumption streams.

Let us work out some recursive implications of these statements. Statements 1 and 2 imply that $(1, 0, 0, \ldots) \succeq^{[0,1]} (0, 1, 0, 0, \ldots)$. Then statement 3 implies that $(1,-1,0,\ldots) \succeq^{[0,1]} 0$. This is a first-order implication of impatience; let us work out a second-order implication: using 2, $(1,-1,0,\ldots) \succeq^{[0,1]} (0,1,-1,0,0,\ldots)$, from which 3 implies $(1,-2,1,0,0,\ldots) \succeq^{[0,1]} 0$. Observe that $(1,-2,1,0,0,\ldots) \succeq^{[0,1]} 0$ reflects "convexity" of the discount function, or the idea that mean preserving spreads (in time) are desirable. One can go further and work out a third-order expression, and a fourth-order expression, and so forth. All such statements are implications of an idea we refer to as *recursive impatience*.

So far we have not yet used that $x \succeq^{[0,1]} 0$ implies $(0,x) \succeq^{[0,1]} 0$, but it is easy to see what happens when we do: the fact that $(1,-2,1,0,\ldots) \succeq^{[0,1]} 0$ implies that $(0,1,-2,1,0,\ldots) \succeq^{[0,1]} 0$.

By pursuing all the implications of recursive impatience, we shall (essentially) exhaust all the situations in which $x \succeq^{[0,1]} y$. To this end, define a class of vectors, which we call *alternating binomial coefficients:* For $s,t \in \mathbf{N}$, let $\eta(s,t) \in l_\infty$ be defined as $\eta(s,t)_i = (-1)^{(i-s)}\binom{t}{i-s}$ for all $i \in \{s,\ldots,s+t\}$ and $\eta(s,t)_i = 0$ otherwise. For example, $\eta(0,1) = (1,-1,0,\ldots)$ is a transfer of one util from time $t = 1$ to $t = 0$. Our previous discussion of recursive impatience implies that $\eta(0,1) \succeq^{[0,1]} 0$. We shift the transformation $\eta(0,t)$ by s units of time to obtain $\eta(s,t)$: for example, $\eta(5,1)$ is a transfer of consumption on date $t = 6$ to $t = 5$. For a few examples, observe that $\eta(0,0) = (1,0,\ldots)$, $\eta(2,0) = (0,0,1,0,\ldots)$, $\eta(1,1) = (0,1,-1,0,\ldots)$, and $\eta(2,3) = (0,0,1,-3,3,-1,0,\ldots)$.

If we continue, reasoning by induction, our discussion of recursive impatience, we obtain that for all $s,t \in \mathbf{N}$, $\eta(s,t) \succeq^{[0,1]} 0$. In other words, the implications of the four basic statements about discounting is that for all $s,t \in \mathbf{N}$, $\eta(s,t) \succeq^{[0,1]} 0$. Except for the case in which $t = 0$, each $\eta(s,t)$ can be identified with shifting an unambiguously good stream backward one unit in time. For example, $\eta(0,2) = (1,-2,1,0,\ldots) \succeq^{[0,1]} 0$ reflects the fact $\eta(0,1) \succeq^{[0,1]} (0,\eta(0,1))$. Equivalently, $(1,-1,0,\ldots) \succeq^{[0,1]} (0,1,-1,0,\ldots)$. More generally, for all $t > 0$, $\eta(s,t) \succeq^{[0,1]} 0$ reflects that $\eta(s,t-1) \succeq^{[0,1]} (0,\eta(s,t-1))$.

The main result of this section is that the statements derived inductively, using recursive impatience, from statements (1)-(3), essentially exhaust all of the ways in which we may have $x \succeq^{[0,1]} y$, When $x \succeq^{[0,1]} y$, then $(x-y)$ can be expressed as a (limit of) nonnegative linear combination of streams of the form $\eta(s,t)$. Hence, y must arise from x by adding subtracting a lump sum to period zero, and then constructing a sequence of shifts of unambiguously good streams backwards in time.

Define an *elementary transformation of order* s (for $s \in \{0,\ldots\}$) to be a vector of the form $\lambda\eta(s,t)$ for some t and $\lambda > 0$.

Theorem 1. *$y \succeq^{[0,1]} x$ if and only if for each $\epsilon > 0$, there is a finite collection of elementary transformations $\{\lambda_i \eta(s_i,t_i)\}$ for which*

$$\|(y-x) - \sum_i \lambda_i \eta(s_i,t_i)\|_1 \leq \epsilon.$$

Remark 2. If each of y and x are eventually constant (and hence eventually 0), then $(y-x)$ can be expressed as a finite weighted sum of elementary transformations. In other words, the approximation in the preceding is not needed.

The ordering $\succeq^{[0,1]}$ and Theorem 1 presume that one allows for all $\delta \in [0,1]$, but it is possible to extend the theorem.[6] Namely, suppose that it is agreed that the discount factor must lie in a compact interval $[a,b] \subseteq [0,1]$. This would be the case, for example, if there were a lower bound on discounting future generations.

In the three statements discussed above, properties 1 and 3 would remain unchanged. However, property 2 would be replaced. Consider what happens when x dominates 0 for all $\delta \in [a,b]$. Instead of $(0,x) \succeq^{[0,1]} 0$, we can actually say more: we can say that $(0,x) \succeq^{[a,b]} ax$. Further, instead of $x \succeq^{[0,1]} (0,x)$, we can say more: we can say that $bx \succeq^{[a,b]} (0,x)$. So, we would replace 2 with the statement that $x \succeq^{[a,b]} 0$ implies

$$bx \succeq^{[a,b]} (0,x) \succeq^{[a,b]} ax.$$

Otherwise, the induction argument remains the same. We investigate this further in Section 3.1, in the context of bounded sequences (rather than absolutely summable sequences).

The following example illustrates Theorem 1.

[6]We thank Itai Sher for suggesting this question. Observe that Foster and Mitra (2003) perform a similar exercise.

Example 3. Consider the stream $x = (1, 4, 2, -7, 6, -2, 0, 0, \ldots)$. We claim that $x \succeq^{[0,1]} 0$. To see this, observe that shifting back the consumption bundle $(1, 0, 0, \ldots)$ back two units in time results in $x - (1, 0, 0, -1, 0, \ldots) = (0, 4, 2, -6, 6, -2, 0, \ldots) = x_2$. Impatience implies that $x \succeq^{[0,1]} x_2$. Shifting the sequence $(0, 0, 2, -4, 2, 0, \ldots)$ back one unit in time results in $x_2 - (0, 0, 2, -6, 6, -2, \ldots) = (0, 4, 0, 0, \ldots) = x_3$. So $x_2 \succeq^{[0,1]} x_3$. Finally, subtracting 4 units of consumption from period 1 results in $x_3 - (0, 4, 0, 0, \ldots) = 0$. Thus $x \succeq^{[0,1]} x_2 \succeq^{[0,1]} x_3 \succeq^{[0,1]} 0$.

In term of the transformations in Theorem 1,

$$(x - 0) = 4\eta(1, 0) + \eta(0, 1) + \eta(2, 1) + \eta(3, 1) + 2\eta(2, 3).$$

3.1. **The Pareto ordering $\succeq^{[a,b]}$ when $X = \ell_\infty$.** The previous discussion assumed that $X = \ell_1$, and that D was any closed interval in $[0, 1]$. We now turn to $X = \ell_\infty$; a common choice set in applications of intertemporal choice. There is obviously a difficulty here in dealing with the case of $\delta = 1$. We focus on understanding $\succeq^{[a,b]}$ for any $0 \leq a < b < 1$.

The starting point is the observation that for any $\delta \in [a, b]$ and any $s, t \in \mathbf{N}$:

$$(\delta - a)^s (b - \delta)^t \geq 0.$$

The general formula for this expression is a bit messy, but it works out to:

$$(\delta - a)^s (b - \delta)^t \equiv \sum_{m=0}^{s} \sum_{n=0}^{t} \binom{s}{m} \binom{t}{n} (-1)^{t+s-m-n} a^{s-m} b^n \delta^{m+t-n}.$$

Such a polynomial is of degree $s + t$. For any $i \in \{0, \ldots, s + t\}$, it is possible to determine the coefficient on δ^i. The explicit formula for this object is:

$$\eta(i; s, t, a, b) \equiv \sum_{\{(m,n) \in \mathbf{N}^2 : m-n=t+i, 0 \leq m \leq s, 0 \leq n \leq t\}} \binom{s}{m} \binom{t}{n} (-1)^{t+s-m-n} a^{s-m} b^n.$$

The explicit functional form of this object is not important. What is important is that it determines an element of $\eta(s, t, a, b) \in \ell^\infty(\mathbf{N})$ via $[\eta(s, t, a, b)]_i \equiv \eta(i; s, t, a, b)$ for $i \in \{0, \ldots, s + t\}$ and $[\eta(s, t, a, b)]_i = 0$ otherwise. These coefficients generalize the transformations in Theorem 1.

Let $m(\delta) \in \ell^1$ (a summable sequence) be given by $m(\delta) = (1, \delta, \delta^2, \ldots)$. What we have just shown is that for any $s, t \geq 0$ and any $\delta \in [a, b]$, $\eta(s, t, a, b) \cdot$

$m(\delta) \geq 0$. Consequently, for any *finite* list of pairs $(s^1, t^1), \dots, (s^K, t^K)$, $\lambda_k \geq 0$, and $\delta \in [a, b]$, we have

$$\sum_{k=1}^{K} \lambda_k \eta(s^k, t^k, a, b) \cdot m(\delta) \geq 0.$$

The set of such vectors will be denoted $\text{cone}(a, b)$, and is the smallest convex cone containing each $\eta(s, t, a, b)$ as $(s, t) \in \mathbf{N}^2$.

It turns out that a kind of converse is true.

Theorem 4. *$x \succeq^{[a,b]} y$ iff for every $\epsilon > 0$, and every $\{m_1, \dots, m_K\} \subseteq \ell^1$, there is $z \in$ cone(a, b) for which for all $k \in \{1, \dots, K\}$, we have $|m_k \cdot (x - (y + z))| < \epsilon$.*

Corollary 5. *If $x \succeq^{[a,b]} y$ then there is a sequence $z_n \in$ cone(a, b) such that z_n converges pointwise to $x - y$.*

4. The Pareto relation $\succeq^D$ with endogenous D

We now turn to the problem of understanding when an ordering $\succeq$ equals the Pareto ordering $\succeq^D$, for some D. The exercise is axiomatic. We establish the set of axioms that must be satisfied by $\succeq$ for this to be true.

4.1. Axioms.

We proceed to introduce a collection of axioms relevant to the analysis. Recall that a binary relation is an *ordering* if it is reflexive and transitive.

4.1.1. *Standard axioms.*

We state some basic axioms that are either commonly used in the literature, or variations on commonly-used axioms. Then we say a few words about what they mean in our context, and why they might be considered reasonable impositions.

The letters x, y and z refer to streams in X; θ is a constant stream. Unbound variables are universally quantified.

- *Monotonicity*: $x \geq y$ implies $x \succeq y$, and for any constant θ,θ', $\theta \succeq \theta'$ iff $\theta \geq \theta'$.
- *Convexity*: For all $\lambda \in [0, 1]$, if $x \succeq z$ and $y \succeq z$, then $\lambda x + (1 - \lambda) y \succeq z$.
- *Translation invariance*: $x \succeq y$ implies $x + z \succeq y + z$.
- *Homotheticity*: For all $x, y \in X$ and all $\alpha \geq 0$, if $x \succeq y$, then $\alpha x \succeq \alpha y$.
- *Continuity*: $\{y \in X : y \succeq x\}$ and $\{y \in X : x \succeq y\}$ are closed.

The convexity axiom imposes a preference for "smoothing" utility across time. In an intergenerational context, such a preference would naturally result from equity considerations. Note that, in the standard intertemporal choice model with discounted utility, smoothing is a consequence of the concavity of the utility function. There is no such concavity in our model. The streams under consideration are already measured in "utils" per period of time, and the standard intertemporal choice model is linear in utils. Our convexity axiom says that smoothing may be intrinsically desirable. This interpretation appears already in Marinacci (1998).

Translation invariance is usually understood as the requirement that there are no utility comparisons made across periods. It allows for the possibility that the "scale" of utility across periods matters. Note that Translation Invariance imposes separability across time (in the sense that if $x_t = y_t$ and $x'_t = y'_t$ for all $t \in E \subseteq \mathbf{N}$, while $x_t = x'_t$ and $y_t = y'_t$ for all $t \in E^c = \mathbf{N} \setminus E$, then $x \succeq y$ implies $x' \succeq y'$).

We do not have much to say about Continuity, Non-degeneracy or Homotheticity. These axioms are very well known, and have no special meaning in our context.

4.1.2. *Novel axioms.* Our first novel axioms are versions of the Koopmans (1960) stationarity property. Koopmans requires that a stream x is at least as good as y if and only if this preference holds when an identical payoff is appended to the first period of each stream. Our axiom weakens Koopmans', in that they apply only when y is a constant stream (i.e. smooth) and when the payoff appended is equal to the constant in y.

Stationarity: *For all $t \in \mathbf{N}$ and all $\lambda \in [0,1]$,*

$$x \succeq \theta \text{ iff } \lambda x + (1-\lambda)(\underbrace{\theta, \ldots, \theta}_{t \text{ times}}, x) \succeq \theta.$$

Generally speaking, stationarity requires certain choices to be time-invariant. It requires that the comparison between two streams remains the same whether it is made today or in the future. We impose a form of stationarity that requires time-invariance of comparisons with constant, or smooth, streams. The reason is that postponing the decision has a natural interpretation in the case of smooth streams.

Suppose that a policy maker has to choose between two streams, x and a constant stream θ. Think of θ as a baseline, or status quo. The baseline θ is constant, and delivers θ in every period, so (θ, x) is the same as staying with the θ policy for one period and then switching to x. A postponed version of this decision problem would be to choose between (θ, x) and θ. The idea behind stationarity is that the two decision problems are equivalent: one should choose x over θ if and only if one would choose (θ, x) over θ.

A stronger version of stationarity (such as Koopmans') would demand that any decision is preserved if postponed. If our policy maker chooses x over y, then she would be required to choose (θ, x) over (θ, y) for any θ; that is, independently of history. But it is easy to imagine reasons for the decision to be reversed, and (θ, y) chosen over (θ, x).[7] Since (θ, y) is different from y we can imagine situations where θ in period 0 may "enhance" the value of y, for example if θ is a large positive value, and the stream y starts out poorly. The difference with our axiom, in which y is required to be the constant stream θ, is that (θ, y) is different from y. So in our case, we can justify the axiom by saying that if a policy maker is willing to switch from θ to x today, then she must be willing to switch tomorrow.

Finally, our stationarity axiom says more. Not only must the comparison of x and θ be the same as that between (θ, x) and θ, but this must also be true of the comparison of any lottery $\lambda x + (1-\lambda)(\theta, x)$ and θ. In particular, the only basis for choosing between $\lambda x + (1-\lambda)(\theta, x)$ and θ must be the comparison of x with θ, because the only basis for comparing (θ, x) and θ is the comparison between x and θ. The meaning is that there is no additional smoothing (or "hedging") motive in the comparisons of x with θ, now or in the future.

The following axiom, *compensation*, is a technical non-triviality axiom. Its purpose is to ensure that the future is never irrelevant. It is similar in spirit to Koopmans' sensitivity axiom (Postulate 2 of Koopmans (1960)).

[7]See also Hayashi (2016).

Compensation: *For all* t *there are scalars* $\bar{\theta}^t$, θ^t, *and* $\underline{\theta}^t$, *with* $\bar{\theta}^t > \theta^t > \underline{\theta}^t$, *such that*

$$(\underbrace{\underline{\theta}^t, \ldots, \underline{\theta}^t}_{t \text{ times}}, \bar{\theta}^t, \ldots) \succeq \theta^t.$$

Compensation says that for any t there must exists three numbers: $\bar{\theta}^t > \theta^t > \underline{\theta}^t$, such that the worse outcome $\underline{\theta}^t$ for t periods is compensated by a better outcome $\bar{\theta}^t$ for all periods $t+1, \ldots$, relative to the smooth stream that gives the intermediate value θ^t in every period. The axiom ensures that no future period is irrelevant for the purpose of comparing utility streams.

Our last axiom is a weak expression of discounting. Roughly, it states that whenever a stream x is at least as good as a smooth stream θ, then the preference is always willing to wait "long enough" so that changes in x do not matter. Axioms along these lines were introduced by Villegas (1964) and Arrow (1974).

Continuity at infinity: *If, for all* T, $(\underbrace{1, ..., 1}_{T \text{ times}}, 0, \ldots) \not\succeq \theta$, *then* $\theta \succeq \mathbf{1}$.

4.2. **A characterization of $\succeq^D$.** We wish to understand the common properties of all orderings that are the Pareto relation for some society of individuals who are exponential discounters. Put differently, we want to understand the assumptions behind the use of the Pareto criterion in general, abstracting away from particular properties of the Pareto relation for some given group of exponential discounters. The following result says that the Pareto criterion is characterized by a subset of the properties we have discussed above.

Theorem 6. *An ordering* $\succeq$ *satisfies continuity, monotonicity, convexity, translation invariance, stationarity, compensation and continuity at infinity iff there is a nonempty closed*[8] *set* $D \subseteq (0,1)$ *such that* $\succeq = \succeq^D$. *Furthermore,* D *is unique.*

The substantive axioms are convexity, translation invariance, and stationarity. Convexity and translation invariance can be interpreted, respectively,

[8]Closed means with respect to the standard Euclidean topology, and not with respect to the relative topology on $(0,1)$. So any closed set must exclude 0 and 1.

as fairness and intergenerational utility comparability. See Chambers and Echenique (2016) for a detailed discussion. Stationarity was discussed above.

Continuity, compensation, and continuity at infinity are technical axioms. We do not have much to say about them.

4.3. **Interval D.** The set of discount factors obtained in Theorem 6 has no structure other than being closed. It is natural to ask for the conditions under which D will be an interval, as in Section 3. The condition turns out to be a statement of the tradeoff between intertemporal comparisons and utility magnitudes.

Tradeoff: *Let* $0 < a < b < 1$. *If* $(-a, 1, 0, \dots) \not\succeq 0$ *and* $(b, -1, 0, \dots) \not\succeq 0$ *then* $a(b, -1, 0, \dots) \not\succeq (0, b, -1, 0, \dots)$.

The "tradeoff" axiom expresses how making an outcome occur earlier trades off with its magnitude. An agent who discounts the future would not want to shift a bad outcome from a later period to an earlier period, unless the earlier outcome is sufficiently "deflated," or smaller in magnitude.

More specifically, the meaning of $(b, -1, 0, \dots) \not\succeq 0$ is that the tradeoff of receiving b today at the loss of 1 tomorrow is undesirable. It is a "bad," not a "good." To shift the bad forward in time is not desirable: $(b, -1, 0, \dots) \not\succeq (0, b, -1, 0, \dots)$ because we are discounting future outcomes. In principle, we could have $a(b, -1, 0, \dots) \succeq (0, b, -1, 0, \dots)$ if a were small enough. The "bad" $(b, 1, 0, \dots)$ would be deflated, diminished, when multiplied by a small a; a the smaller is a, the more likely it is that shifting $(b, -1)$ forward in time while deflating by a would be desirable. However, since we have $(-a, 1, 0, \dots) \not\succeq 0$, then a is not small enough.

Theorem 7. *An ordering* $\succeq$ *satisfies tradeoff, continuity, monotonicity, convexity, translation invariance, stationarity, compensation and continuity at infinity iff there is a nonempty closed interval* $[a, b] \subseteq (0, 1)$ *such that* $\succeq = \succeq^{[a,b]}$.

4.4. **Maximal subrelations.** We now focus on the following question. Theorem 6 axiomatizes a class of incomplete relations. However, many preference relations need not satisfy the axioms stated there. Motivated by (Cerreia-Vioglio, Ghirardato, Maccheroni, Marinacci, and Siniscalchi, 2011), who work in a framework of uncertainty, we study whether, for a given relation, there exists a maximal subrelation of the type axiomatized in Theorem 6.

We show that whenever there exists a subrelation satisfying the axioms of Theorem 6, there is a maximal such subrelation.

Theorem 8. *Let $\succeq$ be a continuous and convex weak order satisfying that there exists $D^* \subset (0,1)$ closed such that $\forall \delta \in D^*$, $\sum_{t=0}^{\infty} \delta^t x_t \geq \sum_{t=0}^{\infty} \delta^t y_t \Longrightarrow (x-y)+z \succeq z$. Then there is a maximal ordering $\succeq^*$ with the properties that:*

(1) $\succeq^* \subseteq \succeq$;
(2) *there is $D \subseteq (0,1)$, closed, such that $x \succeq^* y$ iff for all $\delta \in D$*

$$\sum_{t=0}^{\infty} \delta^t x_t \geq \sum_{t=0}^{\infty} \delta^t y_t.$$

5. Discussion

5.1. **On the proof of Theorem 6.** Theorem 6 is obtained by first treating $\mathbf{N}$ as a state space, and establishing that the at least as good as set at the origin is supported by a set of probabilities (multiple priors), as in the literature of decisions under uncertainty. We then use the stationarity axiom to update some of the priors, and use updating to show that they must be geometric distributions. The proof of Theorem 6 relies on first obtaining a multiple prior representation as in Bewley (2002): there is a set of probability distributions M over $\mathbf{N}$ such that $x \succeq y$ iff the expected value of x is larger than the expected value of y for all probability distributions in M. We use the continuity at infinity axiom, and ideas from Villegas (1964), Arrow (1974), and Chateauneuf, Maccheroni, Marinacci, and Tallon (2005), to show that the measures in M are countably additive.

The main contribution in our paper is to use stationarity to show that M is the convex hull of *geometric* probability distributions. This is carried out in Lemma 12, which contains the core of the proofs of Theorem 6. The idea is to choose a subset of the extreme points of M (the exposed points of M; these are the extreme points that are the unique minimizers in M of some supporting linear functional), and show that when these priors are updated then they have the *memoryless* property that characterizes the geometric distribution.

Think of each $m \in M$ as representing the beliefs over when the world will end, and choose a particular extreme point m of M. We show that the stationarity axiom implies that for any time period $t \geq 0$, if m' is the belief $m \in M$ conditional (Bayesian updated) on the event $\{t, t+1, \dots\}$ (that is, conditional on the event that the world does not end before time t), then $m' = m$. This means that m is the geometric distribution.

5.2. **On Koopmans' axiomatization.** Koopmans (1960) is the first axiomatization of discounted utility. He relies on two crucial ideas: one is separability and the other is stationarity. Separability means two things. First that $(\theta, x) \succeq (\theta', x)$ iff $(\theta, y) \succeq (\theta', y)$ for all y. Second, that $(\theta, x) \succeq (\theta, y)$ iff $(\theta', x) \succeq (\theta', y)$ for all θ'. It is easy to see that translation invariance implies separability.

The second of Koopman's main axioms is stationarity. It says that $x \succeq y$ iff $(\theta, x) \succeq (\theta, y)$. It is probably obvious how his axiom differs from ours, but let us stress two aspects. In our stationarity axiom, stationarity is only imposed for comparisons with a smooth stream. As we explained in 4.1.2, our idea is that the smooth stream is a status quo, and that the comparison in the stationarity axiom can be phrased as postponing the decision to move away from the status quo.

The other way in which we depart from Koopmans is that our stationarity axiom requires that $\lambda x + (1-\lambda)(\theta, x) \succeq \theta$ implies $x \succeq \theta$ (recall the discussion on page 12). The idea is again that the comparison between $\lambda x + (1-\lambda)(\theta, x)$ and θ is based on the comparison between x and θ.

6. Proof of Theorem 1

To establish the theorem, we need a preliminary definition.

Given $\gamma \in l_\infty$, define the *difference function* $\Delta_\gamma : \mathbf{N}^2 \to \mathbf{R}$ inductively as follows:

(1) $\Delta_\gamma(0, t) = \gamma(t)$
(2) $\Delta_\gamma(m, t) = (-1)^m[\Delta(m-1, t+1) - \Delta(m-1, t)]$.

Say that γ is *totally monotone* if for all $m, t \in \mathbf{N}$, $\Delta_\gamma(m, t) \geq 0$. Total monotonicity is basically the concept of infinite-order stochastic dominance, applied to a discrete environment. The class of totally monotone functions is a subset of l_∞ which we denote by $\mathcal{T}$.

Total monotonicity means for all t:

- $\gamma(t) \geq 0$
- $-\gamma(t+1) + \gamma(t) \geq 0$
- $\gamma(t+2) - 2\gamma(t+1) + \gamma(t) \geq 0$
- $-\gamma(t+3) + 3\gamma(t+2) - 3\gamma(t+1) + \gamma(t) \geq 0$
- $\gamma(t+4) - 4\gamma(t+3) + 6\gamma(t+2) - 4\gamma(t+1) - \gamma(t) \geq 0$

The inequalities are the same as $\eta(m,t) \cdot \gamma \geq 0$ for all $m, t \in \mathbf{N}$.

The following result is due to (Hausdorff, 1921), and is referred to as the *Hausdorff Moment Problem*.[9]

Proposition 9. *Let $\gamma(1) = 1$. Then γ is totally monotone if and only if there is a Borel measure (*i.e. *nonnegative measure on the Borel sets) μ on $[0,1]$ for which $\gamma(t) = \int_0^1 \delta^t \mu(\delta)$.*

Proof. (of Theorem 1) First, we establish that $x \succeq^{[0,1]} y$ if and only if for all $\gamma \in \mathcal{T}$, $\gamma \cdot x \geq \gamma \cdot y$.[10] For $\delta \in [0,1]$, $\gamma(t) = \delta^t$ is totally monotone by Proposition 9. So, if $\gamma \cdot x \geq \gamma \cdot y$ for all $\gamma \in \mathcal{T}$, then $x \succeq^{[0,1]} y$. Conversely, suppose that $x \succeq^{[0,1]} y$. Let $\gamma \in \mathcal{T}$. Then let μ be the Borel over $[0,1]$ associated with γ. Since $x \succeq^{[0,1]} y$, we know that $\sum_t \delta^t x_t \geq \sum_t \delta^t y_t$ for all $\delta \in [0,1]$; integrating with respect to μ obtains $\int_0^1 \sum_t \delta^t x_t d\mu(\delta) \geq \int_0^1 \sum_t \delta^t y_t d\mu(\delta)$. Now, $|\delta^t x_t| \leq |x_t|$ for all t, so $\int_0^1 \sum_t |x_t| d\mu(t) \leq \mu([0,1]) \sum_t |x_t|$. So by Fubini's Theorem (see Theorem 11.26 of Aliprantis and Border (1999), $\int_0^1 \sum_t \delta^t x_t d\mu(t) = \sum_t \int_0^1 \delta^t x_t d\mu(\delta) = \gamma \cdot x$. Similarly, $\int_0^1 \sum_t \delta^t y_t d\mu(\delta) = \gamma \cdot y$, so that $\gamma \cdot x \geq \gamma \cdot y$.

Therefore, if $x \succeq^{[0,1]} y$ is false, there is a totally monotone γ for which $\gamma \cdot (x - y) < 0$. By renormalizing, we can choose γ so that $\gamma \cdot (y - x) \geq 1$. Now, it is simple to verify that γ is totally monotone if and only if $\gamma \cdot \eta(m,t) \geq 0$ for all $m, t \in \mathbf{N}$.[11] So $x \succeq^{[0,1]} y$ being false is equivalent to the consistency of the set of linear inequalities:

[9]Observe that this result is closely related to the characterization of belief functions as those capacities which are totally monotone, *e.g.* Shafer (1976).

[10]We use the notation $\gamma \cdot x = \sum_t \gamma(t) x_t$.

[11]The proof uses Pascal's identity: $\binom{m-1}{i-(t+1)} + \binom{m-1}{i-t} = \binom{m}{i-t}$ to show (by induction on m) that $\gamma \cdot \eta(m,w) = \Delta_\gamma(m,t)$. See, *e.g.* Aigner (2007), p. 5.

- $\gamma \cdot (y - x) \geq 1$
- $\gamma \cdot \eta(m, t) \geq 0$ for all $m, t \in \mathbf{N}$.

for some $\gamma \in l_\infty$.

Consider the set of vectors $(y - x, 1) \in \ell_1 \times \mathbf{R}$ and $(\eta(m,t), 0) \in \ell_1 \times \mathbf{R}$ for all (m,t); we can call this set $\mathcal{V}$. By the Corollary of p. 97 on Holmes (1975), we may conclude that our inequality system is inconsistent if and only if $(0,1)$ is in the closed convex cone spanned by $\mathcal{V}$.

Therefore, we can conclude that for any $\epsilon > 0$, there is $(z, a) \in \ell_1 \times \mathbf{R}$, where (z, a) is in the convex cone spanned by $\mathcal{V}$ and for which $\|z\|_1 + |1 - a| < \epsilon$; which implies that each of $\|z\|_1 < \epsilon$ and $|1 - a| < \epsilon$. In particular, by taking a sufficiently close to 1, we can also guarantee that $\|\frac{1}{a}z\|_1 < \epsilon$.[12] The vector $\left(\frac{1}{a}z, 1\right)$ is in the convex cone spanned by $\mathcal{V}$.

To simplify notation, write $w = \frac{1}{a}z$. Now, $(w, 1)$ is a finite combination of vectors of the form $(\lambda_i \eta(m_i, t_i), 0)$ and $(b(y - x), b)$. Clearly, it must be that $b = 1$, so we have $w = (y - x) + \sum_{i=1}^{N} \lambda_i \eta(m_i, t_i)$, which is what we wanted to show. $\square$

The extension mentioned after the statement of Theorem 1 follows from a generalization of Proposition 9. Specifically, it is known that for $\gamma : \mathbf{N} \to \mathbf{R}$, there is a Borel probability measure μ on $[a, b]$ for which $\gamma(t) = \int_0^1 \delta^t \mu(\delta)$ if and only if for every polynomial $P : \mathbf{R} \to \mathbf{R}$, given by $P(x) = \sum_{i=0}^{n} a_i x^i$ for which for all $x \in [a, b]$, we have $P(x) \geq 0$, it follows that $\sum_{i=0}^{n} a_i \gamma(i) \geq 0$ (see, *e.g.* Theorem 1.1 of Shohat and Tamarkin (1943)). Further, it is known that if P is a nonnegative polynomial on $[a, b]$, then it can be written as $P(x) = \sum_{(s,t) \in S} \lambda_{(s,t)} (x - a)^s (b - x)^t$ for some set of indices $S \subseteq \mathbf{N}^2$ and $\lambda_{(s,t)} \geq 0$. A variant of this fact is due to Bernstein (1915), for the case $[-1, 1]$; see again Shohat and Tamarkin (1943), p. 8 who consider the case $[0, 1]$. The result then follows from renormalizing. Finally this leads to the result, as it implies that we only need to check nonnegativity of the polynomials $(x - a)^s (b - x)^t$ for each s, t.

[12] For example, let $\nu > 0$ so that $\nu^2 + \nu < \epsilon$, and take (z, a) so that $|\frac{1}{a}| < 1 + \nu$ and $\|z\|_1 < \nu$. Then $\|\frac{1}{a}z\|_1 \leq |\frac{1}{a}| \|z\|_1 < \nu^2 + \nu < \epsilon$.

6.1. **Proof of Theorem 4.** Define

$$P(a,b) \equiv \bigcap_{\delta\in[a,b]} \{x \in \ell^\infty : x \cdot m(\delta) \geq 0\}.$$

The discussion preceding the statement of the theorem shows that $\text{cone}(a,b) \subseteq P(a,b)$.

The theorem can then be stated as:

Theorem 10. *$x \in P(a,b)$ iff for every $\epsilon > 0$ and every $\{m_1,\ldots,m_K\} \subseteq \ell^1$, there is $y \in cone(a,b)$ for which for all $k \in \{1,\ldots,K\}$, we have $|m_k\cdot(x-y)| < \epsilon$.*

Proof. Suppose that the second hypothesis is satisfied, and observe that we have shown that $\text{cone}(a,b) \subseteq P(a,b)$.

The second hypothesis establishes that for each $\delta \in [a,b]$ and each $\epsilon > 0$, there is $y \in \text{cone}(a,b)$ such that we have $|m(\delta)\cdot(x-y)| < \epsilon$. But $y \in \text{cone}(a,b)$ implies that $m(\delta)\cdot y \geq 0$. Thus $m(\delta)\cdot x \geq -\epsilon$. Since the inequality holds for any $\epsilon > 0$, we have $m(\delta)\cdot x \geq 0$.

For the other direction, let τ_p denote the topology on ℓ^∞ such that ℓ^1 constitutes the set of continuous linear functionals. That is, the weak topology with respect to the pairing $\langle\ell^\infty,\ell^1\rangle$.

We will show that if $x \notin \overline{\text{cone}(a,b)}$, where $\overline{\text{cone}(a,b)}$ refers to the closure of $\text{cone}(a,b)$ in (ℓ^∞,τ_p), then $x \notin P(a,b)$, which will be enough to establish the claim.

If $x \notin \overline{\text{cone}(a,b)}$, then there is $m^* \in \ell^1$ such that $x\cdot m^* < 0$ and for all $\eta \in \text{cone}(a,b)$, $\eta\cdot m^* \geq 0$ (Theorem 5.58 of Aliprantis and Border (1999)).

By Haviland's Theorem (Shohat and Tamarkin (1943) Theorem 1.1) together with the discussion of the Hausdorff moment problem, we know that for any $m \in \ell^1$, $\eta(s,t,a,b)\cdot m \geq 0$ for all s,t iff there is a nonnegative Borel measure μ on $[a,b]$ for which $m\cdot x \equiv \int x\cdot m(\delta)d\mu(\delta)$.

Consequently, $m^* \in \ell^1$ corresponds to some Borel measure μ^*, so that it follows that there is $\delta \in [a,b]$ for which $x\cdot m(\delta) < 0$. □

7. Proof of Theorem 6

The following lemma is useful.

Lemma 11. *The function* $m : [0,1) \to \ell_1$ *given by* $m(\delta) = (1-\delta)(1, \delta, \delta^2, \ldots)$ *is norm-continuous.*

Proof. First, we show that the map $d : [0,1) \to \ell_1$ given by $d(\delta) = (1, \delta, \delta^2, \ldots)$ is continuous. The result will then follow as $m(\delta) = (1-\delta)d(\delta)$.[13]

So, let $\delta_n \to \delta^*$. Then $\|d(\delta_n) - d(\delta^*)\|_1 = \sum_t |\delta_n^t - (\delta^*)^t|$. Observe that for each t, $|\delta_n^t - (\delta^*)^t| \to 0$. By letting $\hat{\delta} = \sup_n(\delta_n) < 1$, we have that for each t, $|\delta_n^t - (\delta^*)^t| \leq \max\{|(\delta^*)^t|, |\hat{\delta}^t - (\delta^*)^t|\}$, since the expression $|\delta^t - (\delta^*)^t|$ increases monotonically when δ moves away from δ^*. And observe that $\sum_t \max\{|(\delta^*)^t|, |\hat{\delta}^t - (\delta^*)^t|\} < +\infty$. Conclude by the Lebesgue Dominated Convergence Theorem (Theorem 11.20 of Aliprantis and Border (1999)) that $\|d(\delta_n) - d(\delta^*)\|_1 \to 0$. □

Lemma 12, following, characterizes cones in ℓ_∞ which are the set of streams which have nonnegative discounted payoff for every discount factor in some (endogenously determined) closed set of discount factors. The lemma is the main building block in the Bewley style representation. In each environment, the cone of vectors deemed at least as good as 0 must be a cone of this type. From there, it is a matter of translating the properties of the cone into the properties of the preference $\succeq$.

The lemma uses similar ideas to those of Villegas (1964), Arrow (1974), and Chateauneuf, Maccheroni, Marinacci, and Tallon (2005) to obtain countably additive measures. Villegas and Arrow show the existence of countably additive priors in Savage's subjective expected utility model. Chateauneuf et. al show that the set of priors in the α-maximin model is countably additive.

The main novelty in the lemma lies in using the boundary property 4 to show that the measures supporting the cone take the exponential form. This is achieved essentially by updating the supporting measures and by showing the "memoryless" property of the exponential distribution.

Lemma 12. *Let* $P \subseteq \ell_\infty$*. Suppose* P *satisfies the following properties.*

(1) P *is a* ℓ_∞*-closed, convex cone.*
(2) *There is* $p \notin P$*.*

[13] The latter is easily deemed continuous. By a simple application of the triangle inequality, if $\delta_n \to \delta^*$, we have $\|(1-\delta_n)d(\delta_n) - (1-\delta)d(\delta)\|_1 \leq |(\delta - \delta_n)| \|d(\delta_n)\|_1 + (1-\delta)\|d(\delta_n) - d(\delta)\|_1$.

(3) $\ell_\infty^+ \subseteq P$.
(4) $p \in bd(P)$ *implies* $(0,0,\ldots,0,p) \in P$ *and* $p + (0,0,\ldots,0,p) \in bd(P)$.
(5) *For all* $\theta \in [0,1)$, *there is* T *so that*

$$(\underbrace{1-\theta,\ldots,1-\theta}_{T\ times}, -\theta, -\theta, \ldots) \in P.$$

(6) *For all* T, $(\underbrace{0,\ldots,0}_{T\ times}, 1, \ldots) \in int(P)$.

Then there is a nonempty closed $D \subseteq (0,1)$ *so that* $P = \bigcap_{\delta \in D}\{x : \sum_t (1-\delta)\delta^t x_t \geq 0\}$. *Conversely, if there is such a set* D, *all of the properties are satisfied.*

Proof. Establishing that if there is such a D, then the properties are satisfied is mostly simple: Let $M = \{m(\delta) : \delta \in D\}$, so that $P = \bigcap_{\delta \in D}\{x : m(\delta) \cdot x \geq 0\}$. Each set $\{x : m(\delta) \cdot x \geq 0\}$ is closed, and contains ℓ_∞^+, so (1) and (3) are satisfied. Property (2) is immediate as P contains no negative sequences.

For the other properties, note that Lemma 11 and the compactness of D imply that M is norm-compact. Observe that $x \in P$ iff $\inf_{\delta \in D}(1-\delta)\sum_t \delta^t x_t \geq 0$, and that moreover this infimum is achieved (by norm-compactness of M). Then, to see that (4) is satisfied, observe that if $x \in \text{bd}(P)$, then there is $\delta \in D$ for which $m(\delta) \cdot x = 0$, and in particular then, $m(\delta) \cdot (\underbrace{0,\ldots,0}_{T \text{ times}}, x) = 0$, and hence $m(\delta) \cdot (x + (\underbrace{0,\ldots,0}_{T \text{ times}}, x)) = 0$. This means that $x + (0,\ldots,0,x) \in \text{bd}(P)$.

Properties (5) and (6) obtain as $0 < \inf D \leq \sup D < 1$. First, $m(\delta) \cdot (1-\theta,\ldots,1-\theta,-\theta,\ldots) = (1-\delta^T) - \theta$. So $\theta < 1$ means that there is T such that $(1-\delta^T) - \theta \geq 0$ for all $\delta \in D$. Second, for any T, let $\varepsilon > 0$ be such that $\inf\{\delta^T : \delta \in D\} > \varepsilon$. Then if $m(\delta) \cdot (-\varepsilon,\ldots,-\varepsilon,1-\varepsilon,\ldots) = \delta^T - \varepsilon \geq 0$ for all $\delta \in D$. This means that if $\|x - (0,\ldots,0,1,\ldots)\| < \varepsilon$ then $x \in P$.

We now turn to proving that properties (1)-(6) imply the existence of D as in the statement of the lemma.

Step 1: Constructing a set M of finitely additive probabilities on $\mathbf{N}$ as the polar cone of P.

Let $\text{ba}(\mathbf{N})$ denote the bounded, additive set functions on $\mathbf{N}$, and observe that $(\ell_\infty, (ba)(\mathbf{N}))$ is a dual pair. Consider the cone $M^* \subseteq \text{ba}(\mathbf{N})$ given by

$M^* = \bigcap_{p\in P}\{x : x\cdot p \geq 0\}$. By Aliprantis and Border (1999) Theorems 5.86 and 5.91, $P = \bigcap_{x\in M^*}\{p : x\cdot p \geq 0\}$.[14] Since $\ell_\infty^+ \subseteq P$ (property (3)), we can conclude that $M^* \subseteq \text{ba}(\mathbf{N})^+$. Moreover, there is nonzero $m \in M^*$ (by the existence of $p \notin P$, property 2.) For any such nonzero m, observe that since $m \geq 0$, it follows that $m(\mathbf{1}) > 0$.[15] Let $M = \{m \in M^* : m(\mathbf{1}) = 1\}$ and conclude that $P = \bigcap_{m\in M}\{p : x\cdot p \geq 0\}$.

Step 2: Verifying that all elements of M are countably additive, and that $m(\{T,\ldots\}) > 0$ for all $m \in M$.

We show now that each $m \in M$ is countably additive. Since for all $\theta \in [0,1)$, there is T so that $(\underbrace{1-\theta,\ldots,1-\theta}_{T\text{ times}}, -\theta, -\theta, \ldots) \in P$ (property (5)), it follows that for all $m \in M$, $m(\{0,\ldots,T-1\}) \geq \theta$. Conclude that $\lim_{t\to\infty} m(\{0,\ldots,t\}) = m(\mathbf{N})$, so that countable additivity is satisfied.[16] So we write $m(z) = m\cdot z$.

Since $(\underbrace{0,\ldots,0}_{T\text{ times}}, 1, \ldots) \in \text{int}(P)$ (property (6)), we can conclude that $m(\{T,\ldots\}) > 0$ for all $m \in M$.

Step 3: Establishing that M is weakly compact Countably additive and nonnegative set functions can be identified with elements of ℓ_1, so we can view M as a subset of ℓ_1. We show that M is weakly compact, under the pairing (ℓ_1, ℓ_∞).

We first show that M is *tight* as a collection of measures: for all $\varepsilon > 0$ there is a compact (finite) set $E \subseteq \mathbf{N}$ such that $m(E) > 1-\varepsilon$ for all $m \in M$. So let $\varepsilon > 0$ and $\theta' \in (1-\varepsilon, 1)$. Then we know that there is T such that

$$(\underbrace{1-\theta',\ldots,1-\theta'}_{T\text{ times}}, -\theta', \ldots) \in P.$$

The set $E = \{0,\ldots,T-1\}$ works in the definition of tightness because for every $m \in M$, we have $m(\{0,\ldots,T-1\}) \geq \theta' > 1-\varepsilon$.

[14] One needs to verify that P is weakly closed with respect to the pairing $(\ell_\infty, \text{ba}(\mathbf{N}))$, but it is by Theorem 5.86 since $(ba)(\mathbf{N})$ are the ℓ_∞ continuous linear functionals by Aliprantis and Border (1999), Theorem 12.28.

[15] Otherwise, we would have $m(x) = 0$ for all $x \in [0,\mathbf{1}]$, which would imply $m = 0$.

[16] For example, see Aliprantis and Border (1999), Lemma 9.9. Suppose $E_k \subset \mathbf{N}$ is a sequence of sets for which $\bigcap_k E_k = \varnothing$ and $E_{k+1} \subseteq E_k$. Then for each k, there is $t(k) \in \mathbf{N}$ such that $E_k \subseteq \{t(k), t(k)+1, \ldots\}$ and for which $t(k) \to \infty$. Without loss, take t to be nondecreasing. The result then follows as $m(E^k) \leq m(\{t(k), t(k)+1, \ldots,\}) \to 0$.

The weak compactness of M then follows from a few simple identifications. Denote the set of countably additive probability measures on $\mathbf{N}$ by $\mathcal{P}(\mathbf{N})$, and the set of nonnegative summable sequences which sum to 1 by $\mathbf{1}(\mathbf{N})$. Observe that the weak* topology on $P(\mathbf{N})$ induced by the pairing $(\ell_\infty, P(\mathbf{N}))$ coincides with the weak topology on $\mathbf{1}(\mathbf{N})$ induced by the pairing $(\mathbf{1}(\mathbf{N}), \ell_\infty)$, when in the second instance we identify each $m \in P(\mathbf{N})$ with an element of $\mathbf{1}(\mathbf{N})$. By Lemma 14.21 of Aliprantis and Border (1999), since M is tight, its closure is compact in the first topology (and hence the second). But M is already closed, as the intersection of a collection of closed sets.[17] Therefore, we know that every net in M has a subnet which converges in the weak topology on $\mathbf{1}(\mathbf{N})$. Viewing now M as a subset of ℓ_1, we know that every net in M has a convergent subnet in the weak topology induced by the pairing (ℓ_1, ℓ_∞), which is what we wanted to show.

Step 4: Characterizing exposed points of M. A point of M is *exposed* if there is a linear functional f with $f(m) < f(m')$ for all $m' \in M \setminus \{m\}$. We now show that any exposed point of M has the form $(1-\delta)(1, \delta, \delta^2, \ldots)$ for some $\delta \in [0, 1]$. So, suppose that $m \in M$ is an exposed point. Then there exists $x \in \ell_\infty$ such that $x \cdot m < x \cdot m'$ for all $m' \in M \backslash \{m\}$. Clearly it is without loss to suppose that $x \cdot m = 0$.[18] Since $x \cdot m = 0$, it follows that x is on the boundary of P. Therefore, for any T, $x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x)$ is also on the boundary of P (property 4). Since $x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x)$ is on the boundary, it has a supporting hyperplane $m^x \in M^*$ passing through the origin, for which for all $y \in P$,

$$0 = m^x \cdot (x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x)) \leq m^x \cdot y.^{19}$$

We can choose m^x to be non-constant; so we can take $m^x \in M$. So there is $m^x \in M$ such that $0 = m^x \cdot ((\underbrace{0, \ldots, 0}_{T \text{ times}}, x) + x)$. But observe that, since

[17]Namely, the sets $\{m : p \cdot m \geq 0\}$ for $p \in P$ and $\{m : \mathbf{1} \cdot p = 1\}$.

[18]If $x \cdot m > 0$, observe that $x - (x \cdot m)\mathbf{1}$ satisfies $0 = (x - x \cdot m\mathbf{1}) \cdot m < (x - x \cdot m\mathbf{1}) \cdot m'$.

[19]That it has a supporting hyperplane follows from Aliprantis and Border (1999), Lemma 5.78. That the supporting hyperplane passes through zero follows as P is a cone. That m^x is in the polar cone to P follows by definition.

$x \in P$ and $(\underbrace{0,\ldots,0}_{T \text{ times}},x) \in P$, $m^x \cdot x \geq 0$ and $m^x \cdot (0,\ldots,0,x) \geq 0$. Then $0 = m^x \cdot (\underbrace{0,\ldots,0}_{T \text{ times}},x) + m^x \cdot x$ means that $m^x \cdot x = 0$ and $m^x \cdot (0,\ldots,0,x) = 0$. But $m^x \cdot x = 0$ implies that $m^x = m$, as x was chosen to expose m. In turn, $m^x = m$ implies that $m \cdot (\underbrace{0,\ldots,0}_{T \text{ times}},x) = 0$ as well.

Let

$$m^T = \frac{(m(T-1), m(T), m(T+1), \ldots)}{m(\{T-1,\ldots\})} \in \ell_1.$$

(recall that we established that $m(\{T-1,\ldots\}) > 0$.) We shall first show that $m^T \in M$. Let $p \in P$. It is enough to show that $(\underbrace{0,\ldots,0}_{T \text{ times}},p) \in P$, as $m^T \cdot p = m \cdot (0,\ldots,0,p) \geq 0$ and $p \in P$ is arbitrary. So let $0 \leq c = \inf\{p \cdot m' : m' \in M\}$, and note that $0 = \inf\{\cdot(p - c\mathbf{1}) : m' \in M\}$, the infimum being achieved at some $m' \in M$ by compactness of M. Then $p - c\mathbf{1} \in \text{bd}(P)$. Property (4) implies that $(0,\ldots,0,p-c\mathbf{1}) \in P$. Property (3) implies that $(0,\ldots,0,p) \in P$.

Now, $m^T \cdot x = 0$ and x exposes m, so $m^T \in M$ implies that $m = m^T$. This equation ($m^T = m$ for all T) characterizes the geometric distribution: Let $h(s) = m(\{s, s+1,\ldots\})$. Then we have

$$\begin{aligned}\frac{h(s+t)}{h(t)} &= \frac{m(\{t+s, t+s+1,\ldots\})}{m(\{t,t+1,\ldots\})} \\ &= m(\{s,s+1,\ldots\}) = h(s).\end{aligned}$$

Then we obtain $h(t) = h((t-1)+1) = h(t-1)h(1)$. Continuing by induction $h(t) = h(1)^t$. If we let $\delta = h(1) = m^*(\{1,2,\ldots\})$, we have $m^*(\{t,\ldots\}) = \delta^t$ for all $t \geq 1$, and $m^*(\{0\}) = 1 - m^*(\{1,\ldots\}) = 1 - \delta$. Finally, observe $\delta > 0$ as $m(\{T,\ldots\}) > 0$ for all T.

So, conclude that each exposed point of M takes the form $(1-\delta)(1,\delta,\delta^2,\ldots)$ for some $\delta > 0$ (and clearly $\delta < 1$).

Step 5: Finalizing the characterization

Since we have established that M is weakly compact, a theorem of Lindenstrauss and Troyanski ensures that it is the weakly closed convex hull of its strongly exposed points (see Corollary 5.18 of Benyamini and Lindenstrauss (1998)); and, in particular then, of its exposed points. This then allows us to conclude that P has the desired form; let D denote the set of associated discount factors. By Lemma 11, we may take D to be closed. Moreover, $0 \notin \delta$, since for any $m \in M$ and any T, $m(\{T, \ldots, \}) > 0$. □

7.1. **Proof of Theorem 6.** We establish the sufficiency of the axioms first. Let $P = \{x \in \ell_\infty : x \succeq 0\}$. Translation invariance implies that $x \succeq y$ iff $x - y \succeq 0$. So $x \succeq y$ iff $x - y \in P$. If we can show that P satisfies the conditions of Lemma 12 then we are done, because if $D \subseteq (0,1)$ is as delivered by the lemma, then $x \succeq y$ iff $x - y \in P$ iff $\forall \delta \in D$ $\sum_{t=0}^{\infty}(1-\delta)\delta^t(x_t - y_t) \geq 0$.

Lemma 13. *The set P satisfies all of the properties listed in Lemma 12.*

Proof. First, we show that P is closed under positive scalar multiplication. If $x \in P$, then for any $\lambda \in [0,1]$, we have $\lambda x \in P$ by convexity. On the other hand, if $x \in P$, then for any $n \in \mathbf{N}$, we have $nx \in P$ by translation invariance, transitivity, and a simple induction argument.[20] Conclude that if $x \in P$ and $\lambda > 0$, then $\lambda x \in P$.

Hence P is a cone. P is closed since $\succeq$ is continuous. That P is convex follows from the convexity of $\succeq$.

Monotonicity of $\succeq$ implies that the set of positive vectors is contained in P (property 3) and that $-\mathbf{1} \notin P$, so property 2 is satisfied.

Let $x \in \text{bd}(P)$ and $T > 0$. Strong stationarity of $\succeq$ implies that $(\underbrace{0, \ldots, 0}_{T \text{ times}}, x) \in P$. So $x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x) \in P$ because P is a convex cone. To show that $x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x) \in \text{bd}(P)$, let $\varepsilon > 0$ and x' be such that $\|x - x'\|_\infty < \varepsilon/2$ and $x' \notin P$. Note that

$$\|x + (0, \ldots, 0, x) - x' + (0, \ldots, 0, x')\|_\infty < \varepsilon.$$

We claim that $x' + (0, \ldots, 0, x') \notin P$. So suppose that $x' + (0, \ldots, 0, x') \in P$. Then $(1/2)x' + (1/2)(0, \ldots, 0, x') \in P$ as P is a cone. Thus $(1/2)x' + (1/2)(0, \ldots, 0, x') \succeq 0$, which by stationarity implies that $x' \succeq 0$, contradicting that $x' \notin P$.

[20]Namely, since $x \in P$, if $(n-1)x \in P$, then $x + (n-1)x \succeq 0 + (n-1)x$, by translation invariance. Thus by transitivity, $nx \succeq 0$.

Now turn to property 5. Suppose that the property does not hold. Then there is some $\theta \in [0, 1)$ such that for all T, $(\underbrace{1-\theta, \dots, 1-\theta}_{T \text{ times}}, -\theta, -\theta, \dots) \notin P$. Using translation invariance, for all T,

$$(\underbrace{1, \dots, 1}_{T \text{ times}}, 0, 0, \dots) \not\succeq \theta.$$

Then continuity at infinity implies that $\theta \succeq \mathbf{1}$, contradicting monotonicity of $\succeq$ (as $\theta < 1$).

Finally, property 6 follows from compensation. For all T,

$$(\underbrace{\underline{\theta}^t - \theta^t, \dots, \underline{\theta}^t - \theta^t}_{t \text{ times}}, \bar{\theta}^t - \theta^t, \dots) \succeq 0$$

(using c-translation invariance). So monotonicity of $\succeq$ and $\underline{\theta}^t < \theta^t$ implies that $(0, \dots, 0, \bar{\theta}^t - \underline{\theta}^t, \dots) \succ 0$. Homotheticity of $\succeq$ then implies that $(\underbrace{0, \dots, 0}_{T \text{ times}}, 1, \dots) \succ 0$. Property 6 then follows from the continuity of $\succeq$. □

Now we turn to the necessity of the axioms. Continuity at infinity is necessary: Suppose that for all T, $(\underbrace{1, \dots, 1}_{T \text{ times}}, 0, \dots) \not\succeq \theta$. Then for every T, there exists $\delta_T \in D$ for which $\theta > (1-\delta_T)\sum_{t=0}^{T} \delta_T^t 1 = (1-\delta_T^{T+1})$. Without loss we can take $\delta_T = \delta_* = \max\{\delta : \delta \in D\}$. Since D is closed, $\delta_* \in D$. Now, $\theta > 1 - \delta_*^{T+1}$ for all T implies that $\theta \geq 1$. Then monotonicity of $\succeq$ implies that $\theta \succeq 1$.

Compensation is also a simple consequence of D being closed and therefore bounded away from 1.

Lemma 14. *Stationarity is necessary.*

Proof. Let $t > 0$ and $\lambda \in [0, 1]$. Let $z = \lambda x + (1-\lambda)(\underbrace{\theta, \dots, \theta}_{t \text{ times}}, x) - \theta\mathbf{1}$. Then for any $\delta \in (0, 1)$

$$\begin{aligned}\sum_{\tau=0}^{\infty} \delta^\tau z_\tau &= \lambda \sum_{\tau=0}^{\infty} \delta^\tau (x_\tau - \theta) + (1-\lambda) \sum_{\tau=t}^{\infty} \delta^\tau (x_{\tau-t} - \theta) \\ &= [\lambda + (1-\lambda)\delta^t] \sum_{\tau=0}^{\infty} \delta^\tau (x_\tau - \theta)\end{aligned}$$

Note that $[\lambda + (1-\lambda)\delta^t] > 0$. So $(1-\delta)\sum_{\tau=0}^{\infty}\delta^\tau z_\tau$ for all $\delta \in D$ iff $(1-\delta)\sum_{\tau=0}^{\infty}\delta^\tau(x_\tau - \theta) \geq 0$ for all $\delta \in D$. □

7.2. **Uniqueness.**

Proof. By Lemma 11, $m(D)$ and $m(D')$ are closed, as the continuous image of compact sets. Let M and M' be the closed convex hulls of $m(D)$ and $m(D')$, respectively. If $\delta \in D' \setminus D$ then $m(\delta) \notin M$ (because no $m(\delta)$ can be written as a convex combination of some finite $m(\delta_1), \ldots, m(\delta_n)$).

Topologize $\Delta(\mathbf{N})$ with the weak*-topology on $\sigma(C_b(\mathbf{N}), \Delta(\mathbf{N}))$; that is, the weakest topology for which the map $\mu \mapsto x \cdot \mu$ is continuous for every $x \in C_b(\mathbf{N})$ (observe also that any such $x \in l_\infty$). By Lemma 14.21 of Aliprantis and Border (1999), each of M and M' is compact.

Since $m(\delta) \notin M$, there is a continuous linear functional x separating $m(\delta)$ from M (Theorem 5.58 of Aliprantis and Border (1999)). By Lemma 14.4 and Theorem 5.83 of Aliprantis and Border (1999), there is $x \in l_\infty$ for which $x \cdot m(\delta) < \inf_{m' \in M} x \cdot m'$. Let $y \in \mathbf{R}$ be given by $y = \frac{x \cdot m(\delta) + \inf_{m' \in M} x \cdot m'}{2}$ and observe that $(x-y) \cdot m(\delta) < 0 < \inf_{m' \in M}(x-y) \cdot m'$. Conclude that $0 \succ (x-y)$ and $(x-y) \succ' 0$. □

8. Proof of Theorem 7

Let D be as in Theorem 6 and suppose, towards a contradiction, that D is not an interval. The set D is closed, so D can only fail to be a closed interval if if there exists $\delta \in (0,1) \setminus D$ and $\delta_0, \delta_1 \in D$ with $\delta_0 < \delta < \delta_1$. In fact, since D is closed we can find $x, y \in (0,1)$ with $\delta_0 < x < \delta < y < \delta_1$ and $[x, y] \cap D = \emptyset$.

Now $\delta_0 \in D$ and $\delta_0 < x$ means that $(-x, 1, 0, \ldots) \not\succeq 0$, while $\delta_1 \in D$ and $y < \delta_1$ means that $(y, -1, 0, \ldots) \not\succeq 0$. Then the tradeoff axiom implies that $x(y, -1, 0, \ldots) \not\succeq (0, y, -1, 0, \ldots)$; or (using translation invariance) that $(xy, -x-y, 1, 0, \ldots) \not\succeq 0$.

So there is $\eta \in D$ with

$$0 > xy - \eta(x+y) + \eta^2 = (\eta - x)(\eta - y).$$

This means that $\eta \in (x, y)$, a contradiction.

9. Proof of Theorem 8

Let $\mathcal{P}$ be the set of all cones P in ℓ_∞ that satisfy the properties listed in Lemma 12, and for which, if $z \in P$, then $x + z \succeq x$ for all x. The set $\mathcal{P}$ is nonempty because it contains $\{z \in \ell_\infty : \forall \delta \in D^*, \sum \delta^t z_t \geq 0\}$.

Let K be the closure of the convex hull of $\bigcup \mathcal{P}$. We show that if $(x-y) \in K$, then $x \succeq y$. First, if $x - y = \sum_i \lambda_i z_i$, for $\lambda \geq 0$, where $\sum_i \lambda_i = 1$ and for each i, $z_i \in \bigcup \mathcal{P}$, then $x \succeq y$ follows from convexity of $\succeq$. Otherwise, for any $\epsilon > 0$, there are λ_i^ϵ, z_i^ϵ where $\|(x - y) - \sum_i \lambda_i^\epsilon z_i^\epsilon\|_\infty < \epsilon$, and $z_i \in \mathcal{P}$. In this case, since $y + \sum_i \lambda_i^\epsilon z_i^\epsilon \succeq y$ for each ϵ, the result follows by continuity of $\succeq$.

Now note that if $K = \ell_\infty$ then we are done because the theorem is true trivially when $\succeq = \ell_\infty \times \ell_\infty$. So suppose that $\ell_\infty \setminus K \neq \emptyset$. We show that $K \in \mathcal{P}$, which proves the theorem. By Lemma 15 below, K satisfies the properties listed in Lemma 12. So Lemma 12 implies that $K \in \mathcal{P}$, and we are therefore done.

In the following, $\overline{\text{co}}$ refers to the closed, convex hull.

Lemma 15. *Let $\mathcal{P}$ be a nonempty collection of cones satisfying the properties listed in Lemma 12. Then there is a nonempty closed $D \subseteq (0,1)$ so that*

$$\overline{co}\left(\bigcup \mathcal{P}\right) = \bigcap_{\delta \in D} \{x : \sum_t (1-\delta)\delta^t x_t \geq 0\}.$$

Proof. Let $\tilde{m}$ denote the function defined in Lemma 11.

Let $\mathcal{P}$ be a collection of closed convex cones with the property that for each $P \in \mathcal{P}$ there is $D_P \subseteq (0,1)$, closed, such that

$$P = \bigcap_{\delta \in D_p} \{z : \tilde{m}(\delta) \cdot z \geq 0\}.$$

Denote by M_P the ℓ_1-closed convex hull of $\{m(\delta) : \delta \in D_P\}$. Note that by basic properties of polars and duals (see Aliprantis and Border (1999), Theorem 5.91), $z \in \overline{\text{co}}\,(\bigcup \mathcal{P})$ iff $m \cdot z \geq 0$ for all $m \in \bigcap_{P \in \mathcal{P}} M_P$.

Let m be an extreme point of $\bigcap_{P \in \mathcal{P}} M_P$. For each $P \in \mathcal{P}$, $m \in M_P$. We claim that there exists a probability measure μ_P on D_P such that for all t, $m_t = \mathbf{E}_{\mu_P} m(\delta)_t$. To see this, let m^n be a sequence, where each $m^n \in \text{co}\{m(\delta) : \delta \in D_P\}$, such that $m^n \to_1 m$. For each n, $m^n = \sum \lambda_i^n m(\delta_i^n)$ for some λ_i^n, δ_i^n. The set of probability measures on D_P is weak*-compact (Theorem 6.25 of

Aliprantis and Border (1999)), so there is a probability measure μ_P on D_P so that (taking a subsequence if necessary), $\lambda^n \to_{w*} \mu_P$. This implies that for each t,

$$m_t^n \to \mathbf{E}_{\mu_P} m_t(\delta) = \mathbf{E}_{\mu_P}(1-\delta^t)\delta^t.$$

Thus $m_t = \mathbf{E}_{\mu_P}(1-\delta^t)\delta^t$.

The cone P was arbitrary, so the uniqueness of the moment curve implies that μ_P is independent of P; and can be identified with a probability on $\bigcap D_P$, say $\mu = \mu_P$. Thus m is an expectation of $\{m(\delta) : \delta \in \bigcap D_P\}$. We assumed that m is an extreme point of M, so μ must be degenerate and there must exist $\delta \in \bigcap D_P$ with $m = m(\delta)$. □

References

Aigner, M. (2007): *A Course in Enumeration*. Springer.

Aliprantis, C. D., and K. C. Border (1999): *Infinite dimensional analysis: a hitchhiker's guide*. Springer, second edition.

Arrow, K. J. (1974): *Essays in the theory of risk-bearing*. North-Holland Amsterdam.

Bastianello, L., and A. Chateauneuf (2016): "About delay aversion," *Journal of Mathematical Economics*, 63, 62–77.

Benyamini, Y., and J. Lindenstrauss (1998): *Geometric Nonlinear Functional Analysis*, no. pt. 1 in American Mathematical Society colloquium publications. American Mathematical Soc.

Bernstein, S. N. (1915): "Sur la représentation des polynômes positifs," *Har'k. Mat. Obshch*, 14(5), 227–228.

Bewley, T. F. (2002): "Knightian decision theory. Part I," *Decisions in Economics and Finance*, 25(2), 79–110.

Blackwell, D. (1953): "Equivalent comparisons of experiments," *The Annals of Mathematical Statistics*, 24(2), 265–272.

Bøhren, Ø., and T. Hansen (1980): "Capital budgeting with unspecified discount rates," *The Scandinavian Journal of Economics*, pp. 45–58.

Cerreia-Vioglio, S., P. Ghirardato, F. Maccheroni, M. Marinacci, and M. Siniscalchi (2011): "Rational preferences under ambiguity," *Economic Theory*, 48(2-3), 341–375.

Chambers, C. P., and F. Echenique (2016): "On multiple discount rates," Caltech Social Science Working Paper 1418.

Chateauneuf, A., F. Maccheroni, M. Marinacci, and J.-M. Tallon (2005): "Monotone continuous multiple priors," *Economic Theory*, 26(4), 973–982.

Ekern, S. (1981): "Time dominance efficiency analysis," *The Journal of Finance*, 36(5), 1023–1033.

Foster, J. E., and T. Mitra (2003): "Ranking investment projects," *Economic Theory*, 22(3), 469–494.

Hara, C. (2008): "Complete monotonicity of the representative consumer's discount factor," *Journal of Mathematical Economics*, 44(12), 1321–1331.

Hausdorff, F. (1921): "Summationsmethoden und momentfolgen. I," *Mathematische Zeitschrift*, 9, 74–109.

Hayashi, T. (2016): "Consistent updating of social welfare functions," *Social Choice and Welfare*, 46, 569–608.

Holmes, R. B. (1975): *Geometric functional analysis and its applications*. Springer-Verlag New York.

Koopmans, T. C. (1960): "Stationary ordinal utility and impatience," *Econometrica*, 28(2), 287–309.

Marinacci, M. (1998): "An axiomatic approach to complete patience and time invariance," *Journal of Economic Theory*, 83(1), 105–144.

Minardi, S., and A. Savochkin (2016): "Characterizations of Smooth Ambiguity Based on Continuous and Discrete Data," forthcoming, Mathematics of Operations Research.

Pratt, J. W., and J. S. Hammond (1979): "Evaluating and comparing projects: Simple detection of false alarms," *The Journal of Finance*, 34(5), 1231–1242.

Rothschild, M., and J. E. Stiglitz (1970): "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2, 225–243.

Shafer, G. (1976): *A mathematical theory of evidence*, vol. 1. Princeton University Press, Princeton.

Shohat, J. A., and J. D. Tamarkin (1943): *The problem of moments*. American Mathematical Soc.

Trannoy, Alain and Karcher, T. (1999): "Critres de dominance temporelle. Le manuel de l'utilisateur," *conomie & prvision*, 2-3(138-139), 147–162.

Villegas, C. (1964): "On Qualitative Probability σ-Algebras," *The Annals of Mathematical Statistics*, pp. 1787–1796.