

ON MULTIPLE DISCOUNT RATES

CHRISTOPHER P. CHAMBERS AND FEDERICO ECHENIQUE

ABSTRACT. Disagreements over long-term projects can often be traced to assumptions about the discount rate. The debate in economics over climate change is a case in point. We propose a theory of intertemporal choice that is robust to specific assumptions on the discount rate. Our discussion is centered around three models: The PARETO model requires that one utility stream be chosen over another if and only if its discounted value is higher for all discount factors in a set of possible factors. The UTILITARIAN model focuses on an average discount factor. The MAXMIN model evaluates a flow by the lowest available discounted value. We propose these models as robust decision criteria for intertemporal choice, investigate their properties, and break them down axiomatically.

(Chambers) DEPARTMENT OF ECONOMICS, UNIVERSITY OF CALIFORNIA AT SAN DIEGO
(Echenique) DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES, CALIFORNIA INSTITUTE OF TECHNOLOGY

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.

1. INTRODUCTION

When making long-term decisions, economists calculate the present value of the stream of future consequences of a plan, or a project. Such calculations involve compound interest, and are naturally highly sensitive to the assumed discount rate. But it is very difficult to say precisely which discount rate should be used: The discount rate depends on ethical and empirical considerations, on which economists and other experts disagree. As a result, important long-term decisions hinge on specific assumptions about a parameter that is very hard to pin down, and that many people disagree over. *In response to this problem, we develop a theory of decisions over long-term streams that is robust to specific assumptions about the discount rate.*

A case in point is the debate on climate change. The well-known Stern review of climate change (Stern (2007); commissioned by the British government) calculates the effects of progressive climate change, recommending drastic policy measures. Economists such as Robert Barro, Partha Dasgupta, William Nordhaus, and Martin Weitzman take issue with Stern's calculations. The debate centers, in fact, around Stern's assumed discount rates. Hal Varian aptly summarizes the debate (Varian, 2006):

“So, should the social discount rate be 0.1 percent, as Sir Nicholas Stern, who led the study, would have it, or 3 percent as Mr. Nordhaus prefers? There is no definitive answer to this question because it is inherently an ethical judgment that requires comparing the well-being of different people: those alive today and those alive in 50 or 100 years.”

Varian points to a fundamental problem with how economists evaluate long-term projects. Present-value calculations are very sensitive to the assumed discount rate, but people naturally disagree over the specific discount rate to be used. Weitzman (2001) makes a similar point. Weitzman reports the results of a survey of over 2,000 economists, in which he asks them for the discount rate that they would use to evaluate long-term projects, such as proposals to abate climate change. The mean of the answers is 3.96 % with a standard deviation of 2.94 %, reflecting a substantial disagreement over the discount rate. Weitzman considers the possibility that the most prominent economists

do agree over discounting. So he runs the survey on a subsample of 50 very distinguished economists (including many who had won, or have since won, a Nobel prize). The results are very similar, with a mean of 4.09 % and a standard deviation of 3.07 %. It is therefore clear that there is substantial disagreement among economists about the proper discount rate for discounting long-term streams. In fact, Weitzman concludes that:

“The most critical single problem with discounting future benefits and costs is that no consensus now exists, or for that matter has ever existed, about what actual rate of interest to use.¹”

The problem of multiple discount rates goes beyond climate change. It shows up in any kind of long term project evaluation. For example, the US Office of Management and Budget recommends a wide range, between 1% and 7%, for the discount rate when evaluating “intergenerational benefits and costs.” Of course many present-value calculations are going to depend a lot on which number between 1 and 7 is chosen for the discount rate.

The point of our paper is to propose some solutions to the problem highlighted by Varian, Weitzman, and others. *We propose a decision theory that uses discounting, and present-value calculations, yet is robust to specific assumptions about the discount rate.* We explore the consequences of remaining agnostic about the specific discount rate to use (including the possibility of remaining fully agnostic, and allowing for every possible discount rate).

We operationalize robustness in three different ways. Think of choosing among sequences of real numbers: these could be consumption streams, utils, or monetary quantities computed from the costs and benefits of an economic project. In the sequel we often refer to utility streams for concreteness. First we take robustness to mean that the only valid comparisons are those that hold for any discount factor in some set of possible factors. These are the comparisons of a *Pareto criterion* because they capture what a group of agents with different discount factors would agree on. Then we consider average discounting, where we assign probability weights to the different discount factors. Average discounting corresponds to a *utilitarian criterion*, where the average

¹There is no shortage of economists making the point that analysis of climate change is almost hopeless given disagreements over the discount rate; see also Pindyck (2013).

sums up the utilities of a population of agents with different discount factors. Third, we propose to evaluate each stream according to its worst-case present value calculation. The worst-case criterion corresponds to a *max-min* rule that uses the utility of the agent with the lowest present-value evaluation of the stream. The following table summarizes the models and points to the relevant results in the paper. The rest of the introduction discusses each of these robustness proposals.

Pareto	Utilitarian	Max-min
$x \succeq y$ iff $\forall \delta \in D$	$x \succeq y$ iff $U(x) \geq U(y)$	$x \succeq y$ iff $U(x) \geq U(y)$
$\sum_{t=0}^{\infty} \delta^t x_t \geq \sum_{t=0}^{\infty} \delta^t y_t$	$U(x) = \sum_{t=0}^{\infty} (\int_0^1 \delta^t d\mu(\delta)) x_t$	$U(x) = \min\{(1 - \delta) \sum_{t=0}^{\infty} \delta^t x_t : \delta \in D\}$
Theorems 1 and 5	Theorem 4	Theorem 6

First we describe the dominance criterion, a special case of the Pareto criterion. The dominance criterion is fully agnostic about the discount rate, it ranks one stream over another when its present value is higher regardless of the assumed discount rate. We say that a stream x *discounting dominates* a stream y if $\sum_{t=0}^{\infty} \delta^t x_t \geq \sum_{t=0}^{\infty} \delta^t y_t$ for all $\delta \in [0, 1]$.² If there is a society of agents, deciding how to rank intergenerational streams, and each possible discount factor in $[0, 1]$ is held by some agent, then discounting dominance would coincide with *Pareto* dominance. The dominance criterion has been previously studied by Foster and Mitra (2003). We provide (Theorem 1) a “dual” characterization to theirs, focusing on economic primitives. Below we explain in some detail how our results differ from theirs.

The dominance relation is useful for at least three reasons. a) A social planner evaluating multiple streams can use dominance to filter out dominated streams. Everyone who uses discounting would agree that dominated streams should not be pursued. b) If we think of different time periods as “generations,” then discounting dominance refines the Pareto relation.³ Dominance would then be used in guiding the choice of a social welfare function. In fact we illustrate how to do that by means of our utilitarian and maxmin rules. Each of these rules is a refinement of the dominance criterion. c) Finally, as we

²As we explain in Section 2.2, the results can be generalized to the case $\delta \in [a, b] \subseteq [0, 1]$.

³The Pareto relation according to the individuals deciding how to rank the streams, but not the different generations themselves.

shall see, discounting dominance distills the basic properties of discounting common to all discount factors, and broadens our theoretical understanding of the concept of discounting.

We expand on the ideas behind discounting dominance by endogenizing the set of discount factors. Concretely, we characterize the orderings \succeq for which there exists $D \subseteq (0, 1)$ such that $x \succeq y$ iff for all $\delta \in D$

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

Note that D is not given, but \succeq is. We seek to understand (Theorem 5) the orderings \succeq that capture robust comparisons according to some set of discount factors. Such orderings can be interpreted as the *Pareto relation* of a society of exponentially discounting agents.⁴

Our second operationalization of robustness is motivated by the ideas of Weitzman (2001) and Jackson and Yariv (2015). We consider a ranking of streams based on an average of discount factors. Concretely, we characterize the weak orders \succeq for which there exists a probability measure μ on $[0, 1]$ such that $x \succeq y$ iff $\sum_{t=0}^{\infty} (\int_0^1 \delta^t d\mu(\delta)) x_t \geq \sum_{t=0}^{\infty} (\int_0^1 \delta^t d\mu(\delta)) y_t$. If μ is the distribution of discount factors in a population, then average discounting is the *utilitarian criterion*. The average avoids marriage to a specific discount factor; and we show (Theorem 4) that it is the unique ranking that respects discounting dominance and an independence-type assumption (as well as a continuity and nontriviality axiom).

To formalize robustness through an average discount factor is Weitzman's idea, not ours. Weitzman (2001) proposed that an average discount factor in every period best reflects economists' disagreements over which discount rate to use (Weitzman proposes an average taken from a specific distribution, the Gamma distribution). Jackson and Yariv think of averaging as a utilitarian aggregation of a population of agents with different discount factors. They show that utilitarianism results from a Pareto property and a utility which is additively separable across time. Their work leaves open the possibility

⁴In an exercise inspired by Cerreia-Vioglio, Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2011), we also establish (Theorem 7) that if a relation is monotonic with respect to such an ordering, there is a maximal such ordering ensuring monotonicity.

that there are other behavioral predictions implied by the model. We expand on Weitzman’s idea by fully characterizing average discounting, and by going beyond the assumption of a specific distribution. We expand on Jackson-Yariv by characterizing all behaviors consistent with their utilitarian model, for an arbitrary set of individuals. Moreover, our characterization of the utilitarian model relies on similar ideas to that of the dominance criterion, the Pareto criterion for $D = [0, 1]$.

Finally, our last result imagines that streams are evaluated according to a “worst-case” analysis. A policy maker that listens to advise from economists who, like Stern and Nordhaus, disagree on the discount rate, may want to conservatively use the worst-case scenario provided by each economist for each possible policy. A set of discount factors is given, each normalized so that a constant stream is treated identically by each factor in the set. Any stream is judged by the *minimum* discounted present value across all discount factors in the set. Concretely, we characterize (Theorem 6) the weak orders \succeq with the property that $x \succeq y$ iff $U(x) \geq U(y)$, with

$$U(x) = \inf\left\{(1 - \delta) \sum_{t=0}^{\infty} \delta^t x_t : \delta \in D\right\}.$$

Our robustness ideas are inspired by the literature on decisions under uncertainty. The literature on uncertainty obtains sets of priors over a state space: What is new in our paper is to show that the set of priors is described by certain *geometric* distributions, when the state space is the time line. Our Pareto criterion follows the ideas in Bewley (2002). The worst-case, or maxmin representation is analogous to Huber (1981) or Gilboa and Schmeidler (1989). Bewley’s model is the Pareto relation of a society of agents with different priors. The maxmin relation evaluates each stream according to the worst present value considered possible, analogous to how Huber and Gilboa-Schmeidler evaluate uncertain acts according to its worst expected value. Our results for these models proceed by treating time as a state space, obtaining a Bewley or Gilboa-Schmeidler representation, and then showing that the set of possible priors results from geometric distributions (strictly speaking that the set of possible priors is the closed convex hull of priors with the memoryless

property that characterizes the geometric distribution; we explain more in in Section 3.1).

The substantive properties, or axioms, behind the Pareto and max-min models (Theorems 5 and 6) relate to utility smoothing. The first is a quasiconcavity property, stating, roughly, that smoother streams are better. As smoother streams reflect more “fair” streams in an intergenerational context; this property seems entirely natural. The second property is novel, and modifies the stationarity property of Koopmans (1960). Koopmans imagined that if two streams are ranked, that ranking would not change were a common utility appended to the initial period of each stream. The property is usually understood as a stationarity property: the agent anticipates his preference tomorrow to coincide with his preference today.

We argue that the standard stationarity conclusion may not necessarily hold in a context in which there may be an innate preference for smoothing. For example, appending the common utility to each stream may reverse the already stated preference if one of the new streams becomes more smooth. We rectify this issue in the following way. We do not know what “more smooth” means, but we can at least say that a constant utility stream is smoother than anything else. Thus, if a stream x is at least as good as a smooth stream θ , this cannot be due to a preference for smoothing. In such a case, we would ask stationarity to hold; but we want to ensure that appending a new initial consumption cannot lead to new smoothing opportunities. We do so by requiring that the appended consumption is θ itself. Only in this case is the ranking preserved. We refer to this property as *stationarity*.

The characterizations of Pareto and maxmin obviously differ in other aspects. For example, the Pareto representation requires an incomplete preference in general, and will give rise to status quo bias (as discussed, for example, by Bewley (2002)). It is also separable. The maxmin representation involves a complete preference, but may violate separability. These aspects are discussed in Section 3.2.

Related literature. Our first two results (Theorems 1 and 4) use a version of Hausdorff’s moment problem. We are not the first to note the relation between that problem and discounting in economics: Foster and Mitra (2003) made

the connection earlier. Foster and Mitra (2003) provide a characterization of discounting dominance, the ordering we denote by \succeq^d below. They focus on “finite” streams, and their characterization is of the following nature. There is a collection of vectors, $\pi_1, \pi_2, \dots, \pi_n$ such that $x \succeq^d 0$ iff $\pi_i \cdot x \geq 0$ for all i . The contribution is in showing that instead of verifying that $(1, \delta, \delta^2, \dots) \cdot x \geq 0$ for all $\delta \in [0, 1]$, one can search over a much smaller set; indeed, the set one needs to search over becomes finite when streams are finite. Our result is distinguished from theirs in that we do not reference extraneous vectors such as π_i . Our result is the “dual” (in a linear programming sense) version of theirs. We should also say that we use these ideas in our characterization of the utilitarian rule (Theorem 4); there is no analogous exercise in Foster and Mitra.

The Hausdorff moment problem has been used elsewhere in economics. Hara (2008) uses the continuous version of the Hausdorff moment problem; *i.e.* Bernstein’s Theorem. See also Minardi and Savochkin (2016). Bertsimas, Popescu, and Sethuraman (2000) use the Hausdorff moment problem in the context of pricing an asset whose moments are known.

An important motivation for our paper is the literature on multiple discount rates and the evaluation of long-term projects, see Weitzman (2001) and Jackson and Yariv (2015). In particular, the result on expected discount rates presented in Theorem 4 is motivated by these two papers. Jackson and Yariv consider utilitarian aggregation of discounted utilities, and Weitzman argues for the use of an expected discount rate (that he obtains through a survey of economists) like we obtain in Theorem 4.

Our results on the Pareto and the maxmin models are related to the literature on multiple priors by interpreting the set of time periods as a state space. The Pareto and maxmin models are related to Bewley (2002), and Gilboa and Schmeidler (1989). We explain how we depart from these papers in Section 3.1. The same approach of identifying time with states is taken by Marinacci (1998) and Gilboa (1989). Marinacci suggests interpreting convexity, or “uncertainty aversion,” as a preference for intertemporal smoothing, as we have done here. Bastianello and Chateauneuf (2016) is a more recent example; they work out the implications of delay aversion for multiple priors

(and other models) representation of intertemporal preferences. The paper by Wakai (2008) also considers a max-min representation over the discount factor, but in his model the discount factor may be different in each time period. In that sense, his model is closer to Gilboa-Schmeidler's multiple-prior version of max-min. Wakai's focus is on obtaining a dynamically consistent version of the model with multiple and time-varying discount rates.

A crucial difference between all these papers and ours (specifically with Theorems 5 and 6 in our paper) is that *we show how the stationarity axiom imposes structure on the set of multiple priors*. In particular, stationarity makes certain priors update in a specific way, ensuring the memory-less property of the geometric distribution. This result is how we can go from multiple priors to multiple discount time-invariant discount rates. See Section 3.1 for a more detailed discussion.

The papers by Karni and Zilcha (2000), Higashi, Hyogo, and Takeoka (2009), Higashi, Hyogo, Tanaka, and Takeoka (2016), Pennesi (2015), and Lu and Saito (2016) all consider multiple but randomly chosen discount factors. This is of course quite different from our focus on robust conclusions with respect to a fixed set of discount factors.

Nishimura (2016) develops a general theory of intransitive preferences, and introduces the idea of a transitive core. He has an application to time preferences, the primitive being dated consumption, and building on the work on relative discounting of Ok and Masatlioglu (2007). Nishimura's Theorem 3 states that the transitive core of a relative discounting preference has a representation with multiple discount functions. His analysis is very different from ours, but has in common the proposal of multiple discount factors for the purpose of making normative welfare comparisons.

2. RESULTS

2.1. Definitions and notation. Our paper is a study of intertemporal choice. The objects of choice are sequences, or *streams*, $x = \{x_t\}_{t=0}^{\infty}$ of real numbers, where the natural numbers $\mathbf{N} = \{0, 1, \dots\}$ represent the time line. Let X be the universe of sequences under consideration. For some results we assume that $X = \ell_1$, the space of absolutely summable sequences. For other

results we assume that $X = \ell_\infty$, the space of all bounded sequences. The space ℓ_1 is endowed with the norm $\|x\|_1 = \sum_{t \in \mathbf{N}} |x_t|$, while ℓ_∞ has norm $\|x\|_\infty = \sup\{|x_t| : t \in \mathbf{N}\}$.

The streams in X are interpreted as sequences of *utility values*. Given a stream $x = \{x_t\}_{t=0}^\infty \in X$, x_t is how much utility is received in period t . One could work with a more primitive model, in which the objects of choice correspond to some physical outcomes over time. We focus our study on the problem of choosing among intertemporal streams of utils and abstract from how utilities are determined.

When $x \in \ell_\infty$, and $m \in \ell_1$ is a positive sequence, then we use the notation

$$x \cdot m = \sum_{t \in \mathbf{N}} m_t x_t.$$

Countably additive probability measures on \mathbf{N} are identified with sequences $m \in \ell_1$. Then $x \cdot m$ denotes the expectation of $x \in X$ with respect to the positive measure m .

The sequence $(1, 1, \dots)$, 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

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

takes the value θ for $t = 0, \dots, T - 1$ 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 $t \in \mathbf{N}$.

2.2. All discount factors: 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 $D = [0, 1]$.

Define the *discounting dominance* binary relation \succeq^d on ℓ_1 as follows. Let $x \succeq^d y$ if, for all $\delta \in [0, 1]$, $\sum_t \delta^t x_t \geq \sum_t \delta^t y_t$. Observe that \succeq^d 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^d from the four seemingly trivial observations:

- (1) $(1, 0, 0, \dots) \succeq^d 0$
- (2) If $x \succeq^d 0$, then $x \succeq^d (0, x) \succeq^d 0$
- (3) If $x \succeq^d y$, then $(x - y) \succeq^d 0$.

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 $(1, 0, 0, \dots) \succeq^d (0, 1, 0, 0, \dots)$. Then statement 3 implies that $(1, -1, 0, \dots) \succeq^d 0$. This is a first-order implication of impatience; let us work out a second-order implication: using 2, $(1, -1, 0, \dots) \succeq^d (0, 1, -1, 0, 0, \dots)$, from which 3 implies $(1, -2, 1, 0, 0, \dots) \succeq^d 0$. Observe that $(1, -2, 1, 0, 0, \dots) \succeq^d 0$ 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 $x \succeq^d 0$ implies $(0, x) \succeq^d 0$, but it is easy to see what happens when we do: the fact that $(1, -2, 1, 0, \dots) \succeq^d 0$ implies that $(0, 1, -2, 1, 0, \dots) \succeq^d 0$.

By pursuing all the implications of recursive impatience, we shall (essentially) exhaust all the situations in which $x \succeq^d y$. 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 $i \in \{s, \dots, s + t\}$ and $\eta(s, t)_i = 0$ otherwise. We shift the transformation $\eta(0, t)$ by s units of time to obtain $\eta(s, t)$: for example, $\eta(5, 1)$ is a shift of consumption on date $t = 6$ to $t = 5$. For a few examples, observe that $\eta(0, 0) = (1, 0, \dots)$, $\eta(2, 0) = (0, 0, 1, 0, \dots)$, $\eta(1, 1) = (0, 1, -1, 0, \dots)$, and $\eta(2, 3) = (0, 0, 1, -3, 3, -1, 0, \dots)$.

The inductive argument we sketch above guarantees that for all $s, t \in \mathbf{N}$, $\eta(s, t) \succeq^d 0$.

Observe that each $\eta(s, t)$ can be identified with shifting an unambiguously good stream backward one unit in time. For example, $\eta(0, 2) = (1, -2, 1, 0, \dots) \succeq^d 0$ reflects the fact $\eta(0, 1) \succeq^d (0, \eta(0, 1))$. Equivalently, $(1, -1, 0, \dots) \succeq^d (0, 1, -1, 0, \dots)$. More generally, for all $t > 0$, $\eta(s, t) \succeq^d 0$ reflects that $\eta(s, t-1) \succeq^d (0, \eta(s, t-1))$.

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 $x \succeq^d y$. When $x \succeq^d y$, 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 a sequence of shifts of unambiguously good streams backwards in time.

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

Theorem 1. *$y \succeq^d x$ 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^d and Theorem 1 presume that one allows for all $\delta \in [0, 1]$, but it is possible to extend the theorem.⁵ 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. Denote the derived relation by $\succeq_{a,b}^d$ (so that $\succeq_{0,1}^d = \succeq^d$).

In the three statements discussed above, properties 1 and 3 would remain unchanged. However, property 2 would be replaced. Consider what happens

⁵We thank Itai Sher for suggesting this question. Observe that Foster and Mitra (2003) perform a similar exercise.

when x dominates 0 for all $\delta \in [a, b]$. Instead of $(0, x) \succeq^d 0$, we can actually say more: we can say that $(0, x) \succeq_{a,b}^d ax$. Further, instead of $x \succeq^d (0, x)$, we can say more: we can say that $bx \succeq_{a,b}^d (0, x)$. So, we would replace 2 with the statement that $x \succeq_{a,b}^d 0$ implies

$$bx \succeq_{a,b}^d (0, x) \succeq_{a,b}^d ax.$$

Otherwise, the induction argument remains the same.

The following example illustrates Theorem 1.

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

In term of the transformations in Theorem 1,

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

2.3. Axioms. The remainder of our analysis, and the main contribution of our paper, is a characterization of the utilitarian (mean discounting), Pareto and maxmin criteria discussed in the introduction. The Pareto criterion will involve an endogenously determined set of discount factors.

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

2.3.1. 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 $x \gg y$ implies $x \succ y$.
- *Non-degeneracy:* There exist $x, y \in X$ for which $x \succ y$.

- *d-monotonicity*: $x \succeq^d y$ implies $x \succeq y$.
- *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 $x + z \succeq y + z$.
- *c-Translation Invariance*: $x \succeq y$ implies $x + \theta \succeq y + \theta$.
- *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.

Note that $x \succeq^d y$ presumes that $x, y \in \ell_1$. The relation \succeq^d is our basic dominance relation, so d-monotonicity is as reasonable as monotonicity. Moreover, if we have in mind a population of exponentially discounting agents, d-monotonicity just rules out certain Pareto dominated choices.

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. c-Translation Invariance weakens translation invariance to allow for meaningful intertemporal comparison of utility. 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 $x' \succeq y'$). In contrast, c-Translation Invariance does not impose such separability.

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.

2.3.2. 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 axioms weaken 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 $\lambda \in [0, 1]$,

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

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) .⁶ 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.

⁶See also Hayashi (2016).

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.⁷

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)).

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*

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

Compensation says that for any t there must exist 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 $t + 1, \dots$, 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: *For all $x \in X$, all θ , if $\theta \succeq (x_0, \dots, x_T, 0, \dots)$ for all T , then $\theta \succeq x$.*

2.4. Average discount factor/Utilitarian model. Jackson and Yariv (2015) have shown that a society of individuals aggregating exponentially discounted preferences in a time-separable and Paretian fashion must socially discount

⁷We should say that this direction of the axiom ($\lambda x + (1 - \lambda)(\theta, x) \succeq \theta \implies x \succeq \theta$) is only really needed to give a common stationarity axiom for the Pareto and maxmin models. For the maxmin model, where there is a meaningful notion of indifference, we do not need it. See the discussion of Indifference Stationarity in Section 3.2.

according to a weighted sum of period individual discount factors. Here, we characterize the full implications of this model using Theorem 1. Clearly, such a society must have a discount factor which respects d -monotonicity and translation invariance. It turns out that these are essentially the *only* requirements imposed on the model.

Theorem 4. *A weak order \succeq on ℓ_1 satisfies nondegeneracy, continuity, d -monotonicity, and translation invariance iff there is a Borel probability measure μ on $[0, 1]$ such that $x \succeq y$ iff $\sum_{t=0}^{\infty} (\int_0^1 \delta^t d\mu(\delta)) x_t \geq \sum_{t=0}^{\infty} (\int_0^1 \delta^t d\mu(\delta)) y_t$.*

The proof of Theorem 4 relies on similar ideas to Theorem 1 (see section 6). In particular it uses the Hausdorff moment problem, which allows us to dually describe \succeq^d . Translation invariance then provides the linearity of the functional form in the theorem.

2.5. Pareto criterion. We now turn to the Pareto criterion. The exercise is similar to Section 2.2, but now the set of discount factors is endogenous; and not given as part of the exercise. 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 the particular use of the criterion in the presence of a particular society or group of agents.

Theorem 5. *An ordering \succeq satisfies continuity, monotonicity, convexity, translation invariance, stationarity, compensation and continuity at infinity iff there is a nonempty closed⁸ set $D \subseteq (0, 1)$ such that $x \succeq y$ iff for all $\delta \in D$*

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

Furthermore, D is unique.

2.6. Max min.

⁸Closed 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.

Theorem 6. *The preference relation \succeq satisfies continuity, monotonicity, convexity, homotheticity, c -translation invariance stationarity, compensation and continuity at infinity iff there is a nonempty closed set $D \subseteq (0, 1)$ such that $x \succeq y$ iff $U(x) \geq U(y)$, with*

$$U(x) = \min\{(1 - \delta) \sum_{t=0}^{\infty} \delta^t x_t : \delta \in D\}.$$

Furthermore, D is unique.

As we explain in Section 3.2 below, stationarity could be replaced by “in-difference stationarity” in Theorem 6.

2.7. Maximal subrelations. We now focus on the following question. Theorem 5 axiomatizes a class of incomplete relations. However, many preference relations need not satisfy the axioms stated there. Motivated by (Cerrei-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 5.

We show that whenever there exists a subrelation satisfying the axioms of Theorem 5, there is a maximal such subrelation.

Theorem 7. *Let \succeq be a continuous and convex weak order satisfying that there exists $D^* \subset (0, 1)$ closed such that $\forall \delta \in D^*$, $\sum_{t=0}^{\infty} \delta^t x_t \geq \sum_{t=0}^{\infty} \delta^t y_t \implies (x - y) + z \succeq z$. 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 $\delta \in D$

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

3. DISCUSSION

3.1. On the proof of Theorems 5 and 6. Theorems 5 and 6 are obtained by first treating \mathbf{N} as a state space, and establishing a multiple prior representation, 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 5 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 . Similarly, the proof of Theorem 6 relies on a max-min multiple prior representation, as in Huber (1981, Proposition 2.1 of Chapter 10.2) and Gilboa and Schmeidler (1989). 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 11, which contains the core of the proofs of both Theorems 5 and 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.

3.2. 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, but c -translation invariance does not. So the Pareto model in Theorem 5 satisfies separability, but the following simple example illustrates that the max-min model in Theorem 6 may violate separability: Let the preference relation \succeq have a max-min representation with $D = \{1/5, 4/5\}$. Then $(0, 1, 0, \dots) \succ (0, 0, 2, 0, \dots)$ while $(5, 0, 2, \dots) \succ (5, 1, 0, \dots)$; a violation of

separability. In light of some experimental evidence against separability (see Loewenstein (1987) and Wakai (2008)), it may be interesting to note that the max-min model does not impose it.

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 2.3.2, 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 16). 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 , but we should stress that this direction of the axiom is only really needed in Theorem 5. For the max-min model of Theorem 6 we can use the following version of stationarity instead:

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

$$x \sim \theta \implies \lambda x + (1 - \lambda) \underbrace{(\theta, \dots, \theta, x)}_{t \text{ times}} \sim \theta.$$

The point of using Stationarity instead of Indifference stationarity in Theorem 6 is to provide a common notion of stationarity behind both theorems. It is easy to show that the other Indifference stationarity is implied by the other axioms.

Proposition 8. *Stationarity, continuity, and monotonicity imply indifference stationarity.*

Proof. Suppose that $x \sim \theta$. Then by stationarity, we can conclude that for any $t \in \mathbf{N}$ and any $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda) \underbrace{(\theta, \dots, \theta, x)}_{t \text{ times}} \succeq \theta$. Suppose by means of contradiction that $\lambda x + (1 - \lambda) \underbrace{(\theta, \dots, \theta, x)}_{t \text{ times}} \succ \theta$. Then, by continuity, there is $\varepsilon > 0$ for which $\lambda(x - \varepsilon \mathbf{1}) + (1 - \lambda) \underbrace{(\theta, \dots, \theta, x - \varepsilon \mathbf{1})}_{t \text{ times}} \succ \theta$. Conclude that $x - \varepsilon \mathbf{1} \succeq \theta$, by stationarity. By monotonicity, $x \succ \theta$, a contradiction. \square

We cannot use indifference stationarity for the Pareto representation of Theorem 5 because \succeq may be an incomplete ordering, which makes indifference unsuitable for our analysis. In the max-min model of Theorem 6, in contrast, we can base much of the analysis on the indifference relation \sim , and then Indifference stationarity can be used to replace stationarity in the theorem.

Finally, it is worth mentioning that all our models satisfy *impatience*, meaning that it is always desirable to obtain a positive outcome early; for example $(1, 0, \dots) \succeq (0, 1, 0, \dots)$. It is obvious that $(1, 0, \dots) \succeq^d (0, 1, 0, \dots)$, and therefore that the expected discounting model of Theorem 4 satisfies impatience. It is also true that the Pareto and max-min models satisfy impatience, but it is not the direct implication of any one of our axioms. Rather, impatience comes about because we obtain a multiple prior representation (see the discussion in 3.1), and stationarity and continuity at infinity imply discount factors that are in $(0, 1)$.

3.3. Non-convexity of D and Pareto optimality. The set of discount factors D in Theorems 5 and 6 does not need to be an interval in $(0, 1)$. The set of priors M is convex, but the set of discount factors does not need to be convex. Despite this lack of convexity, D has some of the same properties as the set of priors in models of multiple priors.

We will not spell out the details, but one can imagine an exchange economy in which ℓ_∞ is the commodity space, and with n agents, each of them with a set of discount factors D_i . Let M_i be the convex hull of the resulting exponential priors over \mathbf{N} . By results in Billot, Chateauneuf, Gilboa, and Tallon (2000) or Rigotti and Shannon (2005) the existence of smooth Pareto optimal outcomes relies on the existence of a point $m \in \bigcap_{i=1}^n M_i$.⁹ It is then easy to show that there is $\delta \in \bigcap_{i=1}^n D_i$ such that m corresponds to the exponential distribution over \mathbf{N} defined by δ .

4. CONCLUSION

This paper tackles the problem of multiplicity of discount rates when evaluating long-term projects. One might imagine solutions to the problem that

⁹These results rely on a multi-set generalization of the separating hyperplane theorem, usually attributed to Dubovitskii-Milyutin. See *e.g.* Holmes (1975), exercise 2.47.

do away with the whole concept of exponential discounting, but the idea of present value calculations, and the accompanying notion of stationarity, is so ingrained that it seems difficult to do without discounting. Indeed present-value calculations are taught to highschool students (at least in the United States), and are pervasive in the practice of private and public project evaluation. A theory that does not use exponential discounting might be viable, but it would surely be a hard sell to the economics profession.

We propose instead to develop a theory of discounting that is robust to discount factors by considering a set-valued concept. Our first result provides a language for discussing unambiguous ranking in the presence of discounting when the discount factor is either unknown, or there is disagreement about the discount factor. The model is a version of the Pareto model, with an exogenous discount factor. Our second result provides the implications for a discounter who places a probabilistic assessment on exponential discounting. The model involves a utilitarian social choice function. Finally, our final two results describe an endogenously derived set-valued concept. The two models obtained are a Pareto model with an endogenous set of discount factor, and the maxmin model where each stream is evaluated according to a worst-case present-value calculation.

5. PROOF OF THEOREM 1

To establish the theorem, we need a preliminary definition.

Given $\gamma \in l_\infty$, define the *difference function* $\Delta_\gamma : \mathbf{N}^2 \rightarrow \mathbf{R}$ 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*.¹⁰

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 $x \succeq^d y$ if and only if for all $\gamma \in \mathcal{T}$, $\gamma \cdot x \geq \gamma \cdot y$.¹¹ 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 $x \succeq^d y$. Conversely, suppose that $x \succeq^d y$. Let $\gamma \in \mathcal{T}$. Then let μ be the Borel over $[0, 1]$ associated with γ . Since $x \succeq^d y$, 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 $x \succeq^d y$ 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 $m, t \in \mathbf{N}$.¹² So $x \succeq^d y$ being false is equivalent to the consistency of the set of linear inequalities:

- $\gamma \cdot (y - x) \geq 1$
- $\gamma \cdot \eta(m, t) \geq 0$ 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),

¹⁰Observe that this result is closely related to the characterization of belief functions as those capacities which are totally monotone, e.g. Shafer (1976).

¹¹We use the notation $\gamma \cdot x = \sum_t \gamma(t)x_t$.

¹²The 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.

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 $\|\frac{1}{a}z\|_1 < \epsilon$.¹³ The vector $(\frac{1}{a}z, 1)$ 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 $\gamma : \mathbf{N} \rightarrow \mathbf{R}$, 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 $P : \mathbf{R} \rightarrow \mathbf{R}$, 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 .

6. PROOF OF THEOREM 4

That the axioms are necessary is obvious. Conversely, suppose that the axioms are satisfied. Since ℓ_1 is separable (Theorem 15.21 of Aliprantis and Border (1999)), and since \succeq is a continuous weak ordering, by Debreu's representation theorem (Debreu (1954)), there is a continuous utility function $U : \ell_1 \rightarrow \mathbf{R}$ which represents \succeq .

¹³For 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 $\|\frac{1}{a}z\|_1 \leq |\frac{1}{a}|\|z\|_1 < \nu^2 + \nu < \epsilon$.

First note that $x \succeq y$ implies that $nx \succeq ny$ for any $n \geq 1$. The proof is by induction: suppose that $(n-1)x \succeq (n-1)y$, then using translation invariance twice we obtain that $nx = x + (n-1)x \succeq x + (n-1)y \succeq y + (n-1)y = ny$. In particular, this means that $nx \sim ny$ whenever $x \sim y$.

Second, we argue that there is a scalar θ such that $(\theta, 0, \dots) \succ 0$. By non-degeneracy there is $x, y \in \ell_1$ with $x \succ y$. By translation invariance, we may without loss suppose that $y = 0$, so we obtain that $x \succ 0$. We may also suppose without loss that $x \geq 0$, since monotonicity (implied by d -monotonicity) implies $(x \vee 0) \succeq x$.¹⁴ Now using d -monotonicity, note that $(\theta, 0, \dots) \succeq x$ for any scalar $\theta \geq \|x\|_1$.

These two facts, that $nx \sim ny$ whenever $x \sim y$ and that $(\theta, 0, \dots) \succ 0$, imply that if $\gamma > \gamma'$ then $(\gamma, 0, \dots) \succ (\gamma', 0, \dots)$. The reason is that $(\gamma, 0, \dots) \succeq (\gamma', 0, \dots)$ by monotonicity (again implied by d -monotonicity) and that $(\gamma, 0, \dots) \sim (\gamma', 0, \dots)$ would mean that $(n\gamma, 0, \dots) \sim (n\gamma', 0, \dots)$ for any $n \geq 1$. But if we choose n with $n(\gamma - \gamma') > \theta$ then $(n\gamma, 0, \dots) \sim (n\gamma', 0, \dots)$ would mean that $(\theta, 0, \dots) \succ 0 \sim (n\gamma - \gamma', \dots)$ and contradict monotonicity.

Given that we have shown that $\gamma > \gamma'$ implies $U(\gamma, 0, \dots) > U(\gamma', 0, \dots)$ we may without loss of generality assume that $U(\gamma, 0, \dots) = \gamma$ for $\gamma \in \mathbf{R}$.¹⁵

We claim that U is a linear functional. By definition of U , we know that for all $x, y \in \ell_1$, $U(U(x), 0, 0, \dots) = U(x)$, so $x \sim (U(x), 0, 0, \dots)$; and $y \sim (U(y), 0, 0, \dots)$. Hence, we know that $x + y \sim (U(x), 0, 0, \dots) + y \sim (U(x), 0, 0, \dots) + (U(y), 0, 0, \dots)$, by a double application of translation invariance. Conclude that $U(x + y) = U((U(x) + U(y), 0, 0, \dots))$. Since $U((U(x) + U(y), 0, 0, \dots)) = U(x) + U(y)$, we have that $U(x + y) = U(x) + U(y)$. That $U(\alpha x) = \alpha U(x)$ for any $\alpha \in \mathbf{R}$ follows from the preceding and the continuity of U . Hence, U is a continuous, monotone linear functional representing \succeq . Moreover, $U(1, 0, 0, \dots) = 1$.

The dual space of ℓ_1 coincides with l_∞ (Theorem 12.28 of Aliprantis and Border (1999)), so that there is some bounded function $\gamma : \mathbf{N} \rightarrow \mathbf{R}$ for which

¹⁴The notation $(x \vee 0)$ refers to the element-by-element maximum.

¹⁵That this normalization is valid relies on the fact that $U(\{\theta, 0, 0, \dots : \theta \in \mathbf{R}\}) = U(\ell_1)$. This latter property holds since, by continuity, for any $y \in \ell_1$ and any $\theta > 0$, there is $n > 0$ large so that $(\theta, 0, 0, \dots) \succ \alpha n^{-1}y$. This can be seen to imply that $(n\theta, 0, 0, \dots) \succ y$. Similarly, there is θ' for which $y \succ (\theta', 0, 0, \dots)$. Hence by continuity there is θ^* for which $(\theta^*, 0, 0, \dots) \sim y$, establishing the claim.

for all $x \in \ell_1$, $U(x) = \sum_t \gamma(t)x_t$. Observe now that by d-monotonicity and Theorem 1, for each $s, t \in \mathbf{N}$, we have $U(\eta(s, t)) \geq 0$. In other words,

$$0 \leq U(\eta(s, t)) = \gamma \cdot \eta(s, t) = \Delta_\gamma(s, t)$$

for all $s, t \in \mathbf{N}$. Thus, γ is totally monotone. The result then follows from Proposition 9 and the fact that $U(1, 0, 0, \dots) = 1$.

7. PROOF OF THEOREMS 5 AND 6

The following lemma is useful.

Lemma 10. *The function $m : [0, 1) \rightarrow \ell_1$ given by $m(\delta) = (1 - \delta)(1, \delta, \delta^2, \dots)$ is norm-continuous.*

Proof. First, we show that the map $d : [0, 1) \rightarrow \ell_1$ given by $d(\delta) = (1, \delta, \delta^2, \dots)$ is continuous. The result will then follow as $m(\delta) = (1 - \delta)d(\delta)$.¹⁶

So, let $\delta_n \rightarrow \delta^*$. Then $\|d(\delta_n) - d(\delta^*)\|_1 = \sum_t |\delta_n^t - (\delta^*)^t|$. Observe that for each t , $|\delta_n^t - (\delta^*)^t| \rightarrow 0$. 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 $\|d(\delta_n) - d(\delta^*)\|_1 \rightarrow 0$. \square

Lemma 11, 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 both the maxmin representation, and 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

¹⁶The latter is easily deemed continuous. By a simple application of the triangle inequality, if $\delta_n \rightarrow \delta^*$, we have $\|(1 - \delta_n)d(\delta_n) - (1 - \delta)d(\delta)\|_1 \leq |(\delta - \delta_n)|\|d(\delta_n)\|_1 + (1 - \delta)\|d(\delta_n) - d(\delta)\|_1$.

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

- (1) P is a ℓ_∞ -closed, convex cone.
- (2) There is $p \notin P$.
- (3) $\ell_\infty^+ \subseteq P$.
- (4) $p \in \text{bd}(P)$ implies $(0, 0, \dots, 0, p) \in P$ and $p + (0, 0, \dots, 0, p) \in \text{bd}(P)$.
- (5) For all $\theta \in [0, 1)$, there is T so that

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

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

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 10 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, $m(\delta) \cdot \underbrace{(0, \dots, 0, x)}_{T \text{ times}} = 0$, and hence $m(\delta) \cdot (x + \underbrace{(0, \dots, 0, x)}_{T \text{ times}}) = 0$. This means that $x + (0, \dots, 0, x) \in \text{bd}(P)$.

Properties (5) and (6) obtain as $0 < \inf D \leq \sup D < 1$. First, $m(\delta) \cdot (1 - \theta, \dots, 1 - \theta, -\theta, \dots) = (1 - \delta^T) - \theta$. 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 $m(\delta) \cdot (-\varepsilon, \dots, -\varepsilon, 1 - \varepsilon, \dots) = \delta^T - \varepsilon \geq 0$ for all $\delta \in D$. This means that if $\|x - (0, \dots, 0, 1, \dots)\| < \varepsilon$ 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 $(\ell_\infty, (\text{ba})(\mathbf{N}))$ 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, $P = \bigcap_{x \in M^*} \{p : x \cdot p \geq 0\}$.¹⁷ 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 $m(\mathbf{1}) > 0$.¹⁸ 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 $m(\{T, \dots\}) > 0$ 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 $\underbrace{(1 - \theta, \dots, 1 - \theta, -\theta, -\theta, \dots)}_{T \text{ times}} \in P$ (property (5)), it follows that for all $m \in M$, $m(\underbrace{\{0, \dots, T-1\}}_{T \text{ times}}) \geq \theta$. Conclude that $\lim_{t \rightarrow \infty} m(\{0, \dots, t\}) = m(\mathbf{N})$, so that countable additivity is satisfied.¹⁹ So we write $m(z) = m \cdot z$.

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

¹⁷One 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 $(\text{ba})(\mathbf{N})$ are the ℓ_∞ continuous linear functionals by Aliprantis and Border (1999), Theorem 12.28.

¹⁸Otherwise, we would have $m(x) = 0$ for all $x \in [0, \mathbf{1}]$, which would imply $m = 0$.

¹⁹For example, see Aliprantis and Border (1999), Lemma 9.9. Suppose $E_k \subseteq \mathbf{N}$ is a sequence of sets for which $\bigcap_k E_k = \emptyset$ and $E_{k+1} \subseteq E_k$. Then for each k , there is $t(k) \in \mathbf{N}$ such that $E_k \subseteq \{t(k), t(k) + 1, \dots\}$ and for which $t(k) \rightarrow \infty$. Without loss, take t to be nondecreasing. The result then follows as $m(E^k) \leq m(\{t(k), t(k) + 1, \dots\}) \rightarrow 0$.

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

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

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

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.²⁰ 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 $(1 - \delta)(1, \delta, \delta^2, \dots)$ 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 $m' \in M \setminus \{m\}$. Clearly it is without loss to suppose that $x \cdot m = 0$.²¹ Since $x \cdot m = 0$, it follows that x

²⁰Namely, the sets $\{m : p \cdot m \geq 0\}$ for $p \in P$ and $\{m : \mathbf{1} \cdot p = 1\}$.

²¹If $x \cdot m > 0$, observe that $x - (x \cdot m)\mathbf{1}$ satisfies $0 = (x - x \cdot m\mathbf{1}) \cdot m < (x - x \cdot m\mathbf{1}) \cdot m'$.

is on the boundary of P . Therefore, for any T , $x + \underbrace{(0, \dots, 0, x)}_{T \text{ times}}$ is also on the boundary of P (property 4). Since $x + \underbrace{(0, \dots, 0, x)}_{T \text{ times}}$ is on the boundary, it has a supporting hyperplane $m^x \in M^*$ passing through the origin, for which for all $y \in P$,

$$0 = m^x \cdot (x + \underbrace{(0, \dots, 0, x)}_{T \text{ times}}) \leq m^x \cdot y.^{22}$$

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 $0 = m^x \cdot (\underbrace{(0, \dots, 0, x)}_{T \text{ times}} + x)$. But observe that, since $x \in P$ and $\underbrace{(0, \dots, 0, x)}_{T \text{ times}} \in P$, $m^x \cdot x \geq 0$ and $m^x \cdot \underbrace{(0, \dots, 0, x)}_{T \text{ times}} \geq 0$. Then $0 = m^x \cdot \underbrace{(0, \dots, 0, x)}_{T \text{ times}} + m^x \cdot x$ means that $m^x \cdot x = 0$ and $m^x \cdot \underbrace{(0, \dots, 0, x)}_{T \text{ times}} = 0$. 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 $m \cdot \underbrace{(0, \dots, 0, x)}_{T \text{ times}} = 0$ as well.

Let

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

(recall that we established that $m(\{T-1, \dots\}) > 0$.) We shall first show that $m^T \in M$. Let $p \in P$. It is enough to show that $\underbrace{(0, \dots, 0, p)}_{T \text{ times}} \in P$, as $m^T \cdot p = m \cdot \underbrace{(0, \dots, 0, p)}_{T \text{ times}} \geq 0$ 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 $\underbrace{(0, \dots, 0, p - c\mathbf{1})}_{T \text{ times}} \in P$. Property (3) implies that $\underbrace{(0, \dots, 0, p)}_{T \text{ times}} \in P$.

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 $h(s) = m(\{s, s+1, \dots\})$. Then we have

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

²²That 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.

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 $\delta = h(1) = m^*(\{1, 2, \dots\})$, we have $m^*(\{t, \dots\}) = \delta^t$ for all $t \geq 1$, and $m^*(\{0\}) = 1 - m^*(\{1, \dots\}) = 1 - \delta$. Finally, observe $\delta > 0$ as $m(\{T, \dots\}) > 0$ for all T .

So, conclude that each exposed point of M takes the form $(1-\delta)(1, \delta, \delta^2, \dots)$ 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 10, we may take D to be closed. Moreover, $0 \notin \delta$, since for any $m \in M$ and any T , $m(\{T, \dots\}) > 0$. \square

7.1. Proof of Theorem 6. Let us denote $\{x : x \succeq 0\}$ by $\mathcal{U}(0)$. The theorem follows from an application of Lemma 11.

Lemma 12. *The set $\mathcal{U}(0)$ satisfies all of the properties listed in Lemma 11.*

Proof. Verification of most of these properties is simple. That $\mathcal{U}(0)$ is a closed convex cone follow from continuity, convexity, and homotheticity of \succeq . That $\ell_\infty^+ \subseteq \mathcal{U}(0)$ follows from monotonicity and continuity of \succeq . That there is $p \notin \mathcal{U}(0)$ follows from monotonicity, as $0 \succ -\mathbf{1}$.

Let us now show property 4 of Lemma 11, that $x \in \text{bd}(\mathcal{U}(0))$ implies $(0, \dots, 0, x) \in \mathcal{U}(0)$ and $x + (0, \dots, 0, x) \in \text{bd}(\mathcal{U}(0))$. By Proposition 8, we may assume that \succeq satisfies indifference stationarity. Observe that, by continuity and monotonicity, $x \in \text{bd}(\mathcal{U}(0))$ if and only if $x \sim 0$: If $x \sim 0$, then for any ϵ , $x + \epsilon\mathbf{1} \succ x$ and $x \succ x - \epsilon\mathbf{1}$, so $x \in \text{bd}(\mathcal{U}(0))$. On the other hand, if $x \in \text{bd}(\mathcal{U}(0))$, then any open ball about x intersects both $\{y : y \succeq 0\}$ and $\{y : 0 \succ y\}$, so it follows by continuity that $x \sim 0$. So, to establish that property 4 holds, let $x \in \text{bd}(\mathcal{U}(0))$. Then $x \sim 0$, so indifference stationarity implies that $(0, 0, \dots, 0, x) \sim 0$, and $(1/2)x + (1/2)(0, \dots, 0, x) \sim 0$. Then $(0, \dots, 0, x) \in \mathcal{U}(0)$, and, using homotheticity, $x + (0, \dots, 0, x) \sim 0$, so $x + (0, \dots, 0, x) \in \text{bd}(\mathcal{U}(0))$.

Now turn to property 5. We show that for all $\theta \in [0, 1)$, there is T so that

$$\underbrace{(1 - \theta, \dots, 1 - \theta, -\theta, -\theta, \dots)}_{T \text{ times}} \in \mathcal{U}(0).$$

Suppose false, so that (by using c -additivity), there is some $\theta \in [0, 1)$ such that for all T , $\theta \succ \underbrace{(1, \dots, 1, 0, \dots)}_{T \text{ times}}$. Then monotone continuity implies $\theta \succeq \mathbf{1}$, contradicting monotonicity.

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

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

(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 $\underbrace{(0, \dots, 0, 1, \dots)}_{T \text{ times}} \succ$

0. Property 6 then follows from the continuity of \succeq . \square

We proceed to proving the theorem.

Let D be the set of discount factors provided by Lemma 11 for $\mathcal{U}(0)$. We claim that the function U defined by $U(x) = \min_{\delta \in D} (1 - \delta) \sum_t \delta^t x_t$ represents \succeq . To this end, we first establish that $U(x) = 0$ if and only if $x \sim 0$. To see this, suppose $U(x) = 0$. Then, by definition of U , $x \in \mathcal{U}(0)$; thus $x \succeq 0$. To establish that $x \sim 0$ we rule out that $x \succ 0$. So suppose that $x \succ 0$. Then we would have by continuity of \succeq that there is $\epsilon > 0$ small so that $x \succeq \epsilon \mathbf{1}$. But then $x - \epsilon \mathbf{1} \in \mathcal{U}(0)$ (by c -translation invariance), so that $U(x - \epsilon \mathbf{1}) \geq 0$, and then clearly $U(x) \geq \epsilon > 0$, a contradiction. So $U(x) = 0$ implies $x \sim 0$.

Conversely, suppose that $x \sim 0$. It follows that $x \in \mathcal{U}(0)$, from which we obtain $U(x) \geq 0$. If in fact $U(x) > 0$, then let $\epsilon > 0$ be such that $U(x) \geq \epsilon$, and hence $U(x - \epsilon \mathbf{1}) \geq 0$. Thus $x - \epsilon \mathbf{1} \in \mathcal{U}(0)$, so that $x \succ x - \epsilon \mathbf{1} \succeq 0$, or $x \succ 0$, a contradiction.

So now let $x \in \ell_\infty$ be arbitrary. We claim that $x \sim U(x)\mathbf{1}$. But this follows directly from c -translation invariance, as $U(x - U(x)\mathbf{1}) = 0$ if and only if $x - U(x)\mathbf{1} \sim 0$ if and only if $x \sim U(x)\mathbf{1}$.

Finally the result follows a classical textbook argument; if $x \succeq y$, $x \sim U(x)\mathbf{1}$, and $y \sim U(y)\mathbf{1}$, it must be that $U(x) \geq U(y)$, otherwise we would have

$U(y)\mathbf{1} \gg U(x)\mathbf{1}$, and hence $U(y)\mathbf{1} \succ U(x)\mathbf{1}$ by monotonicity. Conversely if $U(x) \geq U(y)$, we have by monotonicity that $U(x)\mathbf{1} \succeq U(y)\mathbf{1}$, so that $x \succeq y$.

The necessity of the axioms is straightforward and omitted. We only provide the calculations showing that the representation satisfies stationarity. Let $x \sim \theta$, so $\theta = U(x) = \min_{\delta \in D} (1 - \delta) \sum_t \delta^t x_t$, where the minimum is achieved for some $\delta \in D$.

Let $z = \lambda x + (1 - \lambda)(\theta, \dots, \theta, x)$. Then for any δ

$$(1 - \delta) \sum_t \delta^t z_t = \lambda(1 - \delta^T)\theta + [\lambda + (1 - \lambda)\delta^T](1 - \delta) \sum_t \delta^t x_t.$$

But $\theta = \lambda(1 - \delta^T)\theta + [\lambda + (1 - \lambda)\delta^T]\theta$, so for $\delta \in D$, $(1 - \delta) \sum_t \delta^t z_t \geq \theta$ if and only if $(1 - \delta) \sum_t \delta^t x_t \geq \theta$. A similar statement holds for equalities. This implies that $U(z) = \theta$.

7.2. Proof of Theorem 5. 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 11 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 11.*

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

and $x' \notin P$. Note that

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

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

The proof that P satisfies properties 5 and 6 are the same as the corresponding proofs in Lemma 12 (in Lemma 12 we use c-translation invariance, which is weaker than what we assume for Theorem 5). \square

Now we turn to the necessity of the axioms. It is clear that continuity at infinity is necessary, as $\theta \geq (1 - \delta) \sum_{t=0}^T x_t$ for all $\delta \in D$ implies that $\theta \geq (1 - \delta) \sum_{t=0}^\infty x_t$ for all $\delta \in D$. Compensation is also a simple consequence of D being bounded away from 1.

Lemma 14. *Stationarity is necessary.*

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

7.3. Uniqueness. The uniqueness argument is common to Theorem 5 and Theorem 6, so we put its proof here:

Proof. By Lemma 10, $m(D)$ 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 $m(\delta_1), \dots, m(\delta_n)$).

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 $(x - y) \succ' 0$. \square

8. PROOF OF THEOREM 7

Let \mathcal{P} be the set of all cones P in ℓ_∞ that satisfy the properties listed in Lemma 11, 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 $\lambda_i^\epsilon, z_i^\epsilon$ where $\|(x - y) - \sum_i \lambda_i^\epsilon z_i^\epsilon\|_\infty < \epsilon$, and $z_i^\epsilon \in \mathcal{P}$. 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 11. So Lemma 11 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 11. Then there is a nonempty closed $D \subseteq (0, 1)$ so that*

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

Proof. Let \tilde{m} denote the function defined in Lemma 10.

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 $m^n \rightarrow_1 m$. 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), $\lambda^n \rightarrow_{w^*} \mu_P$. This implies that for each t ,

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

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 $m = m(\delta)$. \square

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.
- BERTSIMAS, D., I. POPESCU, AND J. SETHURAMAN (2000): “Moment problems and semidefinite optimization,” in *Handbook of semidefinite programming*, ed. by H. Wolkowicz, R. Saigal, and L. Vandenberghe, pp. 469–509. Springer.
- BEWLEY, T. F. (2002): “Knightian decision theory. Part I,” *Decisions in Economics and Finance*, 25(2), 79–110.
- BILLOT, A., A. CHATEAUNEUF, I. GILBOA, AND J.-M. TALLON (2000): “Sharing Beliefs: Between Agreeing and Disagreeing,” *Econometrica*, 68(3), 685–694.
- CERREIA-VIOGLIO, S., P. GHIRARDATO, F. MACCHERONI, M. MARINACCI, AND M. SINISCALCHI (2011): “Rational preferences under ambiguity,” *Economic Theory*, 48(2-3), 341–375.
- CHATEAUNEUF, A., F. MACCHERONI, M. MARINACCI, AND J.-M. TALLON (2005): “Monotone continuous multiple priors,” *Economic Theory*, 26(4), 973–982.
- DEBREU, G. (1954): “Representation of a preference ordering by a numerical function,” in *Decision Processes*, ed. by R. Thrall, R. L. Davis, and C. H. Coombs, pp. 159–165. Wiley.
- FOSTER, J. E., AND T. MITRA (2003): “Ranking investment projects,” *Economic Theory*, 22(3), 469–494.
- GILBOA, I. (1989): “Expectation and variation in multi-period decisions,” *Econometrica*, pp. 1153–1169.
- GILBOA, I., AND D. SCHMEIDLER (1989): “Maxmin expected utility with non-unique prior,” *Journal of Mathematical Economics*, 18(2), 141–153.
- 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.
- HIGASHI, Y., K. HYOGO, AND N. TAKEOKA (2009): “Subjective random discounting and intertemporal choice,” *Journal of Economic Theory*, 144(3), 1015 – 1053.
- HIGASHI, Y., K. HYOGO, H. TANAKA, AND N. TAKEOKA (2016): “Comparative impatience under random discounting,” *Economic Theory*, p. forthcoming.
- HOLMES, R. B. (1975): *Geometric functional analysis and its applications*. Springer-Verlag New York.
- HUBER, P. J. (1981): *Robust statistics*. Springer.
- JACKSON, M. O., AND L. YARIV (2015): “Collective dynamic choice: the necessity of time inconsistency,” *American Economic Journal: Microeconomics*, 7(4), 150–178.
- KARNI, E., AND I. ZILCHA (2000): “Saving behavior in stationary equilibrium with random discounting,” *Economic Theory*, 15(3), 551–564.
- KOOPMANS, T. C. (1960): “Stationary ordinal utility and impatience,” *Econometrica*, 28(2), 287–309.
- LOEWENSTEIN, G. (1987): “Anticipation and the valuation of delayed consumption,” *The Economic Journal*, 97(387), 666–684.
- LU, J., AND K. SAITO (2016): “Random Intertemporal Choice,” mimeo, Caltech.
- 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.
- NISHIMURA, H. (2016): “The transitive core: inference of welfare from non-transitive preference relations,” *Theoretical Economics*, forthcoming.
- OK, E. A., AND Y. MASATLIOGLU (2007): “A theory of (relative) discounting,” *Journal of Economic Theory*, 137(1), 214–245.
- PENNESI, D. (2015): “Uncertain discount and hyperbolic preferences,” Université de Cergy-Pontoise, THEMA Working Paper.

- PINDYCK, R. S. (2013): “Climate change policy: What do the models tell us?,” *Journal of Economic Literature*, 51(3), 860–872.
- RIGOTTI, L., AND C. SHANNON (2005): “Uncertainty and Risk in Financial Markets,” *Econometrica*, 73(1), 203–243.
- SHAFFER, 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.
- STERN, N. (2007): *The Economics of Climate Change: the Stern Review*. Cambridge University Press.
- VARIAN, H. R. (2006): “Recalculating the Costs of Climate Change,” *New York Times*.
- VILLEGAS, C. (1964): “On Qualitative Probability σ -Algebras,” *The Annals of Mathematical Statistics*, pp. 1787–1796.
- WAKAI, K. (2008): “A Model of Utility Smoothing,” *Econometrica*, 76(1), 137–153.
- WEITZMAN, M. L. (2001): “Gamma Discounting,” *American Economic Review*, 91(1), 260–271.