Abstract

Time preference is modelled as a current self that overcomes selfishness by incurring a cognitive cost of empathizing with her future selves. Such a model unifies disparate well-known experimental findings. Behavioral foundations are provided by exploiting the idea that higher stakes provide an incentive for the exertion of higher effort, so that changes in the agent’s impatience with respect to the scale of outcomes pin down the underlying process. The behavioral content of limited cognitive resources is shown to lie in violations of Separability.

1 Introduction

Impatience is a central feature of intertemporal decision-making. Most models in economics treat it as a given feature of an agent’s preference. This

*Noor (the corresponding author) is at the Dept of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Email: jnoor@bu.edu. Takeoka is at the Dept of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. Email: norio.takeoka@r.hit-u.ac.jp. The authors are grateful to Andrew Ellis, Larry Epstein, Bart Lipman, Jay Lu, Samuel McClure, Marco Mariotti, Anna Rubinchik, Bruno Strulovici for very useful discussions, and to seminar participants at Boston University, Yokohama National University, HEC Paris, Queen Mary University of London, Caltech, UC San Diego, Otaru University of Commerce, Brown University, University of Almeria, Hitotsubashi University and the audience at RUD (University of Paris, Dauphine), Bounded Rationality in Choice (Queen Mary), SAET (IMPA, Rio de Janeiro), SAET (Faro, Portugal) and Asian Meeting of the Econometric Society (Doshisha University, Kyoto) for helpful comments. Noor gratefully acknowledges financial support from the NSF (grant SES-1627906). Last but certainly not least, we thank Kevin Cooke and Nick Saponara for expert research assistance. The usual disclaimer applies.
paper explores the idea that impatience may arise endogenously from factors more basic than preference.

The initial motivation for this study comes from the robust finding in intertemporal choice experiments that subjects are more patient when dealing with larger rewards than smaller ones. This finding, known as the *magnitude effect*, was demonstrated by Thaler (1981) in the following average choices of subjects in his experiment:

\[
\begin{align*}
15 & \sim 60 \\
3000 & \sim 4000
\end{align*}
\]

which imply that, under the assumption that the utility \( u \) of money is approximately linear, the discount factor applied to the small reward is \( \delta_{60} \approx 0.25 \) and that to the larger reward is \( \delta_{4000} \approx 0.75 \). Such results have been found also in experiments that use real rewards (see Fredrick et al [13] for a preview and more recently Sun and Potters [36] and Hardisty et al [19]), and also have support from the field (Andersen et al [2]) and is reflected in historical data (Galor and Özak [29]).

We hypothesize that the dependence of impatience on the size of a reward may be an outcome of some cognitive process where a larger reward incentivizes higher cognitive effort that gives rise to less impatience. Introspection suggests that our knowledge of our future selves is not of the same quality as our knowledge of our current self. In fact, our ability to appreciate the well-being of, say, our retired future self requires a process that is analogous to – if not identical to – the process by which we appreciate the well-being of other people: we achieve empathy by imagining ourselves in the other person’s shoes.\(^1\) Our model is a multiple selves model where the current self has a limited capacity for empathy that she allocates over future selves. Empathy is achieved through the costly cognitive effort of imagining herself in the other selves’ shoes. Taking a preference over streams of consumption as our primitive, we provide such a model and study its behavioral foundations and ability to explain what we know about time preference.

It is worth noting that there is some precedence for the idea that empathy may be a relevant component of time preference. In a recent neuroscience experiment, Soutschek et al [34] show that disabling the part of the brain

\(^1\)See Hershfield [20] for a review of studies that show how savings increase with the visualization of the future self.
that is responsible for social cognition (that is, the ability of an agent to take on the perspective of other people) causes subjects not only to become more selfish in dictator games but also more impatient. See Parfit [31] for thought experiments that suggest that future incarnations of an individual are literally, and not just metaphorically, separate selves. See Curry et al [7] and Pronin et al [33] for evidence favoring a positive relationship between patience and generosity. A natural hypothesis is that time preference are a subset of an agent’s social preference. In particular, impatience may be a reflection of the agent’s altruism. We leave it to future research to explore this message.

The main contributions of this paper are as follows.

a) While most models take the discount function as a given feature of preference, we provide a model in which it is determined endogenously.2

b) The model contributes to the literature on time preference by unifying behavioral findings such as magnitude effect, preference reversal, time non-separability, etc. This suggests that there may exist some relations among these anomalies though they have been investigated separately. It also relates to evidence on the role of cognitive abilities for time preference (Dohmen et al [8]), the possible connection between time and social preference (Soutschek et al [34], Pronin et al [33]), and to the multiple selves view of the individual (Parfit [31], Strotz [35], Ainslie [1]). It adds to the theoretical literature on magnitude-dependent discount functions (Noor [27], Baucells and Heukamp [3], Wakai [38], Epstein and Hynes [10]).

c) Outside the time preference literature, there are several models that incorporate subjective optimization, such as those of optimal expectations (Brunnermeier and Parker [6]), optimal contemplation (Ergin and Sarver [11]) and optimal attention (Ellis [9], Gabaix [15]). To our knowledge, neither the behavioral economics nor decision theory literature has explicitly studied constrained subjective optimization, as we do in this paper. One of the key findings in this paper is that binding cognitive constraints for large stakes express themselves behaviorally as violations of Separability.

d) Our perspective on impatience can also enrich our understanding of the multiple selves model. It is recognized in the literature that the well-known hyperbolic discounting/multiple selves model (Strotz [35], Ainslie [1],

---

2Becker and Mulligan [4] also provide a model of endogenous impatience, where the discount function can be altered by investment in education, etc. In our model, the investment is psychological rather than physical.
Laibson [21], O’Donoghue and Rabin [30]) does not include an expression of the notion of “self-control”, which entails effort to reduce the impact of urges on choice. We observe that the magnitude effect can be interpreted as an expression of self-control in the multiple selves framework. Intuitively, the current self can be thought of as being subject to the temptation to be completely selfish, that is, she would set her discount function to $D = 0$ if she exerts zero effort. However she also has an incentive to be empathic toward her future selves, and this motivates her to think about the consequences of her actions on her future selves. Thinking about her future selves gives rise to self-control in her choices as she uses some $D > 0$ to evaluate her stream. The cost of thinking is the cost of exerting self-control, and the magnitude effect is the behavioral expression of self-control.

The remainder of the paper proceeds as follows. Section 2 describes our model. Sections 3-4 provide behavioral foundations for the model on the set of positive streams and Section 5 extends it to all streams. Section 6 relates the model to empirical findings whereas Section 7 discusses an application to procrastination. All proofs are relegated to appendices.

Noor and Takeoka [28], a supplementary appendix of this paper, provide two related results: (1) compatibility of our main representation with the stationarity axiom, and (2) an axiomatization for a more general representation, called the General Costly Empathy representation.

## 2 Model

The time horizon is $T + 1 < \infty$. The consumption space $C$ is a metric space endowed with the metric topology and also a mixture space – in particular, a consumption alternative $c \in C$ can be some continuous variable like money, or it can be a lottery (with consequences in a different space of outcomes). Consider the space of consumption streams $X = C^{T+1}$, endowed with the product topology. A typical element in $X$ is denoted by $x = (x_0, x_1, \ldots, x_T)$. The primitive of our model is a preference $\succsim$ over $X$. 

---

3For a single-self (dynamically consistent) model with time-inconsistent urges and self-control, see Gul and Pesendorfer [18], Noor [26], Fudenberg and Levine [14].
2.1 Costly Empathy

Fix some consumption alternative that represents some base-line consumption level, and denote it by $0 \in C$. Say that a tuple $(u, \{\varphi_t\}_{t=0}^T, K)$ is regular if

(i) $u : C \rightarrow \mathbb{R}$ is continuous and mixture linear with $u(0) = 0$,

(ii) for each $t$, a cost function $\varphi_t : [0, 1] \rightarrow \mathbb{R}_+ \cup \{\infty\}$ takes the form

$$\varphi_t(d) = \begin{cases} a_t \cdot d^m & \text{if } d \in [0, \overline{d}_t], \\ \infty & \text{if } d \notin (\overline{d}_t, 1], \end{cases}$$

where

- $m > 1$,
- $a_0 = 0$ and $\overline{d}_0 = 1$,
- $a_t > 0$ is increasing in $t \geq 1$, and
- $0 < \overline{d}_t \leq 1$ is decreasing in $t \geq 1$,

(iii) Either $K = \infty$, or $\varphi_t(\overline{d}_t) = a_t \overline{d}_t^m = K$ for all $t \geq 1$.

Condition (i) requires that the utility from consumption should have familiar properties. Condition (ii) requires $\{\varphi_t\}$ to be a family of convex power cost functions that represent the cost of cognitive effort of appreciating future consumption. The degree of appreciation of consumption at time $t$ is given by the period $t$ discount factor $d$, and the maximum possible degree of appreciation (beyond which the costs are infinite) is $\overline{d}_t$. There is no cost of appreciating immediate consumption, $\varphi_0 = 0$. The idea that farther consumption is harder to appreciate is expressed by the fact that $a_t$ is increasing with $t$ and that the maximal achievable period $t$ discount factor $\overline{d}_t$ is decreasing with $t$.

Condition (iii) introduces the stock $K$ of cognitive resources. There are two cases. The first is where the resources are infinite. The second is where resources are limited, but are just enough to achieve $\overline{d}_t$ in any given period $t$ (see below for further discussion on this). Indeed, the resources are such that $\overline{d}_t$ cannot be achieved at two or more periods of consumption.

Our model is defined as follows.

**Definition 1 (CE Representation)** A Costly Empathy (CE) representation is a regular tuple $(u, \{\varphi_t\}, K)$ such that $\succsim$ is represented by the function $U : X \rightarrow \mathbb{R}$ defined by

$$U(x) = \sum_{t \geq 0} D_x(t) \cdot u(x_t), \quad x \in X,$$  

(1)
where \( D_x = \arg \max_{D \in [0,1]^{T+1}} \left\{ \sum_{t \geq 0} D(t) \cdot |u(x_t)| - \sum_{t \geq 0} \varphi_t(D(t)) \right\}, \) 

subject to \( \sum_{t \geq 0} \varphi_t(D(t)) \leq K. \)

The CE representation is constrained if \( K < \infty \) and unconstrained otherwise. The CE representation is positive if \( u(C) = \mathbb{R}_+. \)

To interpret, consider a consumption stream \( x = (x_0, ..., x_T). \) The period 0 self evaluates the value of the stream \( x \) via the discounted utility formula (1) where the discount function \( D_x \) depends on the stream. We interpret \( D_x \) in terms of her degree of selfishness, with lower values of \( D_x(t) \in [0,1] \) expressing greater selfishness towards future self \( t. \) The degree of selfishness is a cognitive choice in this model, with the following elements:

- Empathy is a cognitively difficult task, involving the cost of imagining oneself in the other’s shoes. We assume that the cost of any discount function \( D \) is additive:

\[ \varphi(D) := \sum_{t \geq 0} \varphi_t(D(t)) \]

where each \( \varphi_t, t \geq 1, \) is strictly increasing and strictly convex.

- The agent potentially has a cap \( K \) on her capacity for empathy when evaluating any given stream. Thus, in order to evaluate \( x, \) the process is limited to the set of discount functions given by:

\[ \{ D \in [0,1]^{T+1} : \varphi(D) \leq K \}. \]

This is referred to as the capacity constraint.

- The agent’s objective in the cognitive stage is to choose a discount function that maximizes

\[ D \cdot |u(x)| := \sum_{t \geq 0} D(t) \cdot |u(x_t)| \]

subject to costs and constraints. Without costs and constraints, she would set \( D(t) = \bar{d}_t \) for each \( t. \) When trading off consumption across selves, however, she places greater weight on the more extreme (positive or negative) consequences of the stream.

According to the model, the discount function \( D_x \) is determined optimally as in (2)-(3). Since \( \varphi_t \) is strictly convex, \( D_x \) is unique. Because \( \varphi_0 = 0, \) the agent always sets \( D(0) = 1. \)
Let the effective domain of $\varphi$ be denoted by
\[
eff(\varphi) := \{ D : \varphi(D) < \infty \} = \prod_{t \geq 0} [0, \bar{d}_t].
\]
Since choosing a discount function $D \notin \eff(\varphi)$ is prohibitively costly, $\eff(\varphi)$ is regarded as the feasible set of discount functions. We call $D(t) \leq \bar{d}_t$ the boundary constraint. In the constrained CE model, the capacity constraint imposes a further restriction on feasible discount functions such as
\[
\{ D \in [0, 1]^{T+1} : \varphi(D) \leq K \} \subset \eff(\varphi).
\]
In contrast, if $K$ is large enough that $\sum_t \varphi_t(\bar{d}_t) \leq K$, the capacity constraint is superfluous and does not restrict the feasible set of discounting functions in any way. In this case the cognitive optimization is effectively unconstrained and we set $K = \infty$ wlog to define the unconstrained CE model. The unconstrained CE model is thus defined by the representation:
\[
U(x) = D_x \cdot u(x),
\]
where $D_x = \arg \max_{D \in [0, 1]^{T+1}} \{ D \cdot |u(x)| - \varphi(D) \}$.

We clarify a role of the condition $\varphi_t(\bar{d}_t) = K$ in regularity (iii). In general, one can imagine a version of the model with the weaker restriction $\varphi_t(\bar{d}_t) \leq K$. However, assuming equality makes the model more tractable because meeting the capacity constraint becomes sufficient to meet the boundary constraint: when the capacity constraint is satisfied, the discount function satisfies, for any given $t$,
\[
\varphi_t(D_x(t)) \leq \varphi(D_x) \leq K = \varphi_t(\bar{d}_t),
\]
which implies $D_x(t) \leq \bar{d}_t$. The tractability comes from the fact that, both in constructing the model and in its applications, one can effectively ignore the boundary constraint.

As we will see in the axiomatization (Section 3.1), the model satisfies an Impatience axiom. This axiom implies that an optimal discount function for $x$ satisfies
\[
D_x(t) \leq 1,
\]

Footnote 4: If $K$ is exhausted before achieving $\bar{d}_t$, that is, $\varphi_t(\bar{d}_t) = a_t d_t^{m} > K$, then there is $\tilde{d}_t < \bar{d}_t$ such that $a_t d_t^{m} = K$. Since $\varphi_t$ cannot be observed beyond the capacity constraint, there is no behavioral expression of $\varphi_t$ beyond $\tilde{d}_t$. Without loss of generality, we can assume $\bar{d}_t = \tilde{d}_t$. 

7
or $d_t \leq 1$. Though it does not play a significant role in the construction of our representation, we impose the Impatience axiom because of its central place in the literature, and because we feel that its violation (such as due to anticipation) is not best understood in terms of empathy.\footnote{See the discussion at the end of Section 6.5}

2.2 Discussion

1. A few words are in order about the capacity constraint $K$. The model should be viewed as one where the agent is in fact not constrained in her cognitive resources, but there exists a cap on how much of these resources are used for each stream in her menu. This is different from a model where the agent has a limited grand pool, and on facing a menu of streams, decides how to optimally allocate these resources across future selves, giving rise to a menu-dependent (as opposed to stream-dependent) discount function. While such an extension of our model is straightforward to write and technically feasible to provide foundations for, we eschew it because it would generically violate the Weak Axiom of Revealed Preference. We opt to house our ideas in a utility model so as to stay close to the time preference literature, and also for its tractability for applications. We leave it to future research to extend and analyze the model in the noted way.

2. For perspective, it is worth considering why we adopt a cognitive story involving empathy. A cognitive story for the magnitude effect must answer the question: what kind of cognitive effort makes a future reward more appealing to the agent? A natural first pass at an answer is to exploit the classic quote by Pigou [32]: “[O]ur telescopic faculty is defective, and we, therefore, see future pleasures, as it were, on a diminished scale”. One could hypothesize accordingly that higher cognitive effort can enhance telescopic faculties. However, this does not necessitate a magnitude effect: in principle, improved telescopic clarity could make a future reward look less attractive: with faulty telescopic faculty, a consumption alternative may look better from a distance, and enhanced visibility can reveal its shortcomings. Finding a suitable cognitive story was the main conceptual challenge in this paper and the empathy story served the purpose, while also having some indirect support from neuroscience and philosophy.

3. Given the analogy between allocation of a resource across future selves versus allocation across other individuals, it is natural to inquire about
the differences between the self-other relation in the time context and the social context. We observe that in a social context a person may find it aversive if others are better off than her, whereas in the time context it may well be a source of well-being. This suggests that the self-other relationship in the time context is likely purely altruistic whereas other social emotions are at play in the social context. It would be natural to investigate the similarities and differences between time preference and social allocation decisions when the agent has no stake in the decision or is better off than others.

2.3 Reduced Form

We explore some properties of the model here. Assume that $\succsim$ admits a positive CE representation. As noted earlier, because of strict convexity of the cost function, the cognitive optimization problem has a unique solution. It is instructive to analyze how this optimal discount function changes along rays of streams. Given the additive separability of the cost function, it is useful to consider the special stream referred to as the dated reward: a stream that pays some consumption $c$ at time $t$ and 0 otherwise, denoted $(c)^t$. Denote by $(\lambda c)^t$, the dated reward $c$ scaled up or down by $\lambda > 0$ (for instance, think of a lottery that pays $c$ at $t$ with probability $\lambda$ and 0 otherwise). For all values of $\lambda$ for which the capacity constraint is slack when evaluating $(\lambda c)^t$, the first order condition is given by:

$$u(\lambda c) = \varphi_t'(D(t)),$$

and consequently the optimal discount function is given by:

$$D_{\lambda c}(t) = \left(\frac{u(\lambda c)}{ma_t}\right)^\frac{1}{m-1} := \gamma(t)u(\lambda c)^\frac{1}{m-1}.$$

In particular, $D_{\lambda c}$ is increasing in $u(\lambda c)$. Intuitively, as consumption at $t$ improves, the agent has more incentive to exert effort to overcome selfishness, which leads to a higher $D$.

We just saw that $D_{\lambda c}$ is increasing in $\lambda$ as long as the capacity constraint is lax. There is some threshold $\lambda_c$ such that the capacity constraint binds and the discount function ceases to increase in $\lambda$. 

9
Say that a stream $x$ is small if the capacity constraint is slack at the optimal $\lambda x$ and large if it is binding. In the model, a stream $x$ is small if and only if scaling down leads to a magnitude effect: $D_{\lambda x} < D_x$ for any $\lambda < 1$. It is large otherwise. We use this observation in the sequel to identify small and large streams behaviorally.

A very useful feature of the model is that it admits a clean way of distinguishing small and large streams in terms of the representation: a stream $x$ is small if and only if scaling down leads to a magnitude effect: $D_{\lambda x} < D_x$ for any $\lambda < 1$. It is large otherwise. We use this observation in the sequel to identify small and large streams behaviorally.

Write $\gamma(t) := (ma_t)^{-\frac{1}{m-1}}$. Since $a_t$ is increasing, $\gamma(\cdot)$ is a weakly decreasing function.

**Proposition 1** If $\succsim$ admits a positive CE representation $(u, \{a_tD(t)^m\}_{t=1}^T, K)$, then

$$U(x) = \begin{cases} 
    u(x_0) + \sum_{t=1}^T \gamma(t)u(x_t)^{\frac{m}{m-1}} & \text{if } U(x) - u(x_0) \leq mK \\
    u(x_0) + (mK)^{\frac{1}{m}} \left\{ \sum_{t=1}^T \gamma(t)u(x_t)^{\frac{m}{m-1}} \right\}^{\frac{m-1}{m}} & \text{if } U(x) - u(x_0) > mK
\end{cases}$$

This result reveals that for small magnitudes, the utility function is additively separable, and future utility from lotteries is a power transformation of immediate utility. If $c$ is a lottery, then $u(c)$ is an expected utility and risk preferences are unchanged over $t$. However, the parameter $m$ will affect intertemporal substitution. For large stakes we find that the utility function is no longer additively separable. Future utility is evaluated using a concave aggregator.

One more remark is that the positive CE representation on large streams can be interpreted as a maxmin-type (more precisely, maxmax-type) representation à la Gilboa and Schmeidler [17]. For positive large streams, (3) is binding, that is, $\varphi(D) = K$. Then, (2) is reduced to

$$D_x = \arg \max_D \{D \cdot u(x) - K\} = \arg \max_D D \cdot u(x)$$

subject to $D \in \mathcal{D} := \{D \in [0,1]^{T+1} : \varphi(D) = K\}$. The CE representation on the large streams is also written as

$$U(x) = D_x \cdot u(x) = \max_{D \in \mathcal{D}} D \cdot u(x),$$

(4)
that is, given each stream, the agent optimally chooses a discount function
so as to maximize discounted utilities within the capacity constraint.\footnote{To explain consumption smoothing in an intertemporal decision making, Wakai [38] studies an agent who minimizes discounted utilities over a set of discount functions.}

It follows from (4) that a CE representation satisfies convexity for large
streams, while it is not necessarily the case for small streams. However, we
can show that a CE representation must be star-shaped, $\alpha U(x) \geq U(\alpha x)$
(for positive streams), which is a property weaker than convexity. Since our
model violates convexity, it goes beyond models of convex preferences in the
literature (such as Maccheroni et al [24]). See Noor and Takeoka [28] for
more details.

3 Basic Behavioral Framework

Consider a binary relation $\succsim$ over the space $X = C^{T+1}$ of consumption
streams as defined in Section 2.

3.1 Basic Axioms

Part (c) of our composite Regularity axiom below asserts the existence of an
alternative, denoted 0. Abusing notation, we use $c$ to denote both consump-
tion $c \in C$ and a stream $(c, 0, ..., 0) \in X$ that delivers $c$ immediately and 0 in
every subsequent period (so 0 also denotes the stream $(0, ..., 0)$). Relatedly,
denote by $(c)^t$ the stream that pays $c \in C$ at time $t$ and 0 in all other periods.
Such a stream is called a dated reward.

Say that a stream $x$ is positive if $x_t \succsim 0$ for all $t$, and it is negative if
$x_t \precsim 0$ for all $t$. Let $X_+$ denote the set of positive streams, that is,

$$X_+ := \{x \in X | x_t \succsim 0 \text{ for all } t\}.$$ 

Note that via the identification between $c$ and $(c, 0, ..., 0)$, it is meaningful to
say that $c \in C$ is positive or negative.

**Axiom 1 (Regularity)** (a) (Order). $\succsim$ is complete and transitive.
(b) (Continuity). For all $x \in X$, $\{y \in X : y \succsim x\}$ and $\{y \in X : x \succsim y\}$ are
closed.
(c) (Base-line). There exists $0 \in C$ s.t. for any $c \in C$,

$$c \sim 0 \implies (c)^t \sim 0 \text{ for all } t.$$
(d) (Impatience). For any positive $c$ and $t < t'$,

$$(c)^t \succsim (c)^{t'}.$$

(e) (Monotonicity) For any $x, y \in X$,

$$(x, 0, \ldots, 0) \succsim (y, 0, \ldots, 0) \text{ for all } t \implies x \gtrsim y.$$

Moreover, if $(x, 0, \ldots, 0) \succ (y, 0, \ldots, 0)$ for some $t$, then $x \succ y$.

(f) (Present Equivalents). For any stream $x$ there exist $c, c' \in C$ s.t.

$$c \gtrsim x \gtrsim c'.$$

(g) (Risk Preference). For any $c, c', c'' \in C$, and $\alpha \in [0, 1]$,

$$c \succ c' \implies \alpha c + (1 - \alpha)c'' \succ \alpha c' + (1 - \alpha)c''.$$

Order, Continuity and Risk Preference are standard. Impatience requires that positive outcomes are weakly preferred sooner rather than later, and Monotonicity requires that point-wise preferred streams are preferred.

Present Equivalents states that for any positive stream, there are immediate consumption levels that are better and worse than $x$. Given Order and Continuity, this ensures that each stream $x$ has a present equivalent $c_x = (c, 0, \ldots, 0)$ defined by\footnote{Since $c$ is an element of a mixture space, there may be multiple present equivalents of a given stream. However, this plays no role in all that follows.}

$$c_x \sim x,$$

which will be used instrumentally below. Note, however, that Present Equivalents rule out the existence of a “best consumption”: If there was a best consumption $\overline{c}$, then we would obtain $(c, 0, \ldots, 0) \prec (\overline{c}, \overline{c}, \ldots, \overline{c})$ for any $c$, which violates Present Equivalents.

Next we state a separability axiom. For notational convenience, for all streams $x, y \in X$ and $S \subset \{0, 1, \ldots, T\}$, let $x|_{S}y$ denote the stream that pays according to $x$ at $t \in S$ and according to $y$ otherwise. In particular, if $S = \{t\}$, the stream is denoted by $x|_{t}y$ instead of $x\{t\}y$. As before, let $c_x$ denote the present equivalent of $x$.

For any $Z \subset X$, define...
Axiom 2 (Z-Separability) For all $x \in Z$ and all $t$,

$$\frac{1}{2}c_{xt0} + \frac{1}{2}c_{0xt} \sim \frac{1}{2}c_{x} + \frac{1}{2}c_{0}.$$  

We will vary the subset $Z$ on which Separability holds for different results. Given that preferences over immediate consumption satisfy Independence and that this axiom considers lotteries over present equivalents, the axiom can be interpreted in a standard way via its analogy with the usual Independence condition applied for a hypothetical preference that is defined over lotteries over streams.\(^8\)

One can imagine that if an agent establishes empathy for self $t$ then she may costlessly empathize with adjacent selves $t-1$ and $t+1$. This would violate separability. But we could alternatively view the duration of a period as sufficiently long that such intertemporal complementarities disappear.

3.2 Discounted Utility

We first characterize the standard Discounted Utility (DU) model in our setup to help fix some key ideas.

In our setting the DU model is defined as:

**Definition 2 (Discounted Utility Representation)** A Discounted Utility (DU) representation is a tuple $(u, D)$ where $u : C \to \mathbb{R}$ is continuous and mixture linear with $u(0) = 0$ and $D : \{1, \ldots, T\} \to [0, 1]$ is weakly decreasing, such that $\succsim$ is represented by the function $U : X \to \mathbb{R}$ defined by

$$U(x) = u(x_0) + \sum_{t \geq 1} D(t) \cdot u(x_t), \quad x \in X.$$  

This model exhibits no magnitude effect, in that $D$ is independent of the outcomes. To understand the behavioral content of magnitude-independent

\(^8\)To illustrate, consider:

$$\frac{1}{2} \circ (0, c', 0) + \frac{1}{2} \circ (c, 0, c'') \sim \frac{1}{2} \circ (c, c', c'') + \frac{1}{2} \circ (0, 0, 0).$$

This says that the agent only cares about the distribution of consumption across periods, and not the possible correlation across periods.
discounting, consider the following. For any \( c \) and \( \alpha \in [0, 1] \) define the mixture \( \alpha c := \alpha c + (1 - \alpha)0 \). For any stream \( x = (x_0, \ldots, x_T) \) define
\[
\alpha x := (\alpha x_0, \ldots, \alpha x_T).
\]
Intuitively, the stream \( \alpha x \) uniformly “scales down” the consumption offered by \( x \) in every period. Abusing notation, write \( \alpha c \) for the stream \((\alpha c, 0, \ldots, 0)\).

Consider a stream \( x \) and its present equivalent, \( c_x \).

Note that the agent’s evaluation of immediate consumption \( c_x \) does not rely on impatience whereas that of a stream \( x \) does. Then if impatience does not change in response to scaling down \( x \) by \( \alpha \), then it must be that:
\[
\alpha c_x \sim \alpha x,
\]
since the scaling down affects the evaluation of consumption equally for the immediate reward and the stream.

The behavioral content of magnitude-independent discounting is therefore:

**Axiom 3 (Homotheticity)**  For any \( x \in X \) and any \( \alpha \in (0, 1) \),
\[
\alpha c_x \sim \alpha x.
\]

We establish:

**Theorem 1**  A preference \( \succeq \) over \( X \) satisfies Regularity, X-Separability and Homotheticity if and only if \( \succeq \) admits a DU representation.

**Theorem 2**  If there are two DU representations \((u^i, D^i)\), \( i = 1, 2 \) of the same preference \( \succeq \), then \( D^1 = D^2 \) and there exists \( \alpha > 0 \) such that \( u^2 = \alpha u^1 \).

The uniqueness of \( u \) up to scalar multiplication rather than a more general affine transformation because of the model’s requirement that \( u(0) = 0 \).
4 Behavioral Foundations: Positive Model

4.1 Identifying a Magnitude Effect

If Homotheticity is a behavioral definition for magnitude-independent discounting, then magnitude effects correspond to a violation of Homotheticity. Consider a positive stream $x$ and its present equivalent,

$$c_x \sim x.$$

As before, the agent’s evaluation of immediate consumption $c_x$ does not rely on impatience whereas that of a stream $x$ does. It is intuitive that if impatience changes in response to scaling down $x$ by $\alpha$ (i.e. there is a strict magnitude effect) we should observe that the stream loses value faster than the immediate reward:

$$\alpha c_x \succ \alpha x.$$

Therefore a weak magnitude effect (where impatience is weakly decreasing with scale) for positive streams, can be defined by the condition that for any positive stream, that is,

$$c_x \sim x \implies \alpha c_x \succeq \alpha x$$

for any $\alpha \in (0, 1]$.

Given Regularity, this condition implies that either $\alpha c_x \sim \alpha x$ for $\alpha$ close to 1 or that $\alpha c_x \succ \alpha x$ for all $\alpha \in (0, 1)$, that is, for any stream $x$, either there is no magnitude effect with a slight reduction in scale, or that there is a strict magnitude effect.\(^9\) It will be useful to distinguish between these two cases: Say that a positive stream $x$ is small if it is subject to a strict magnitude effect with any scaling down of the stream:

$$\alpha c_x \succ \alpha x \text{ for all } \alpha \in (0, 1).$$

Similarly say that a positive stream $x$ is large if scaling it down slightly does

\(^9\)To establish this, we show that if $\alpha c_x \succ \alpha x$ for any $\alpha$, then $\beta c_x \succ \beta x$ for $\beta \in (0, 1)$. First note that, $\alpha c_x \succ \beta c_x$, and the present-equivalent $\alpha x$ satisfies $\alpha c_x \succ \alpha x \sim c_{\alpha x}$. By Risk Preference, $\beta c_x \succ \beta c_{\alpha x}$ for any $\beta \in (0, 1)$. By hypothesis, $c_{\alpha x} \sim \alpha x$ implies $\beta c_{\alpha x} \succeq \beta x$ for any $\beta \in (0, 1)$. But then $\beta c_x \succ \beta c_{\alpha x} \succeq \beta x$ and thus $\beta c_x \succ \beta x$ for $\beta \in (0, 1)$, as desired.
not change impatience:

\[ \alpha c_x \sim \alpha x \text{ for some } \alpha \text{ close to } 1. \]

In terms of discount functions, these conditions say that \( x \) is small if \( D_x > D_{\alpha x} \) for all \( \alpha \in (0, 1) \), and large if \( D_x = D_{\alpha x} \) for \( \alpha \) close to 1. Intuitively, small streams are those streams for which the agent’s cognitive constraint is not binding. Consequently, any reduction in scale induces lower cognitive effort and therefore a strict magnitude effect. In contrast, if the agent’s cognitive constraint is binding for a stream \( x \), the agent may continue to exert maximum cognitive effort even if the stream is scaled down by a little, and so there would be no magnitude effect.

Define the set of positive small streams by:

\[ X_s = \{ x \in X_+ : \alpha c_x \succ \alpha x \text{ for all } \alpha \in (0, 1) \}. \]

Note that, given Risk Preference, small streams are elements of \( X \setminus C \).

4.2 Unconstrained CE Representation

4.2.1 Axioms

As noted above, the magnitude effect is a violation of Homotheticity, as \( c_x \sim x \leftrightarrow \alpha c_x \sim \alpha x \). The next axiom weakens Homotheticity but retains substantial structure on the preference:

**Axiom 4 (Weak Homotheticity)** (a) For all positive streams \( x \in X \) and all \( \alpha \in (0, 1) \),

\[ \alpha c_x \succsim \alpha x. \]

(b) For any positive stream \( x \in X \setminus X_s \) and any \( \alpha \in (0, 1) \) s.t. \( \alpha x \in X \setminus X_s \),

\[ \alpha c_x \sim \alpha x. \]

(c) For any \( x, y \in X_s \) s.t. \( x_0 \sim y_0 \sim 0 \) and any \( \alpha, \beta \in (0, 1) \),

\[ \beta c_x \sim \alpha x \implies \beta c_y \sim \alpha y. \]

\(^{10}\)The intuitive meaning of “large stream” requires the property only for nonimmediate streams \( x \not\in C \). However, as stated, the definition implies that immediate rewards are large (by Risk Preference). We maintain the definition nonetheless because it is simpler to state.
Part (a) implies a weak magnitude effect with respect to scale, as motivated in Section 4.1. Part (b) imposes homotheticity outside \( X_s \), while homotheticity will be violated when considering streams in \( X_s \). Part (c) of the axiom places structure on homotheticity violations. It imposes a substantive simplification that requires homogeneity of magnitude effects in the sense that for any two small streams, scaling down each by \( \alpha \) causes its present-equivalent to be scaled down by some \( \beta \), where \( \beta \) depends on \( \alpha \) but not the stream.\(^{11}\)

It is obvious that Homotheticity implies part (a) and part (b). If Homotheticity holds then \( X_s = \emptyset \), and hence, it implies part (c) trivially.

Weak Homotheticity does not require that \( \succsim \) exhibits a strict magnitude effect. The next axiom implies that \( X_s \) is non-empty.

**Axiom 5 (Absorbing)** For all positive streams \( x \in X \setminus C \),

(a) \( \alpha c_x \succ \alpha x \) for some \( \alpha \in (0,1) \), and
(b) if \( \alpha c_x \succ \alpha x \) for some \( \alpha \in (0,1) \), then \( \beta c_{\alpha x} \succ \beta \alpha x \) for all \( \beta \in (0,1) \).

Part (a) states that any positive stream \( x \in X \setminus C \), there always exists some scaling down that causes a strict magnitude effect. Part (b) states that if scaling \( x \) down by \( \alpha \) causes a strict magnitude effect, then \( \alpha x \) must be a small stream, that is, any further scaling down leads to a further strict change in impatience.

### 4.2.2 Representation Theorem

Our first result is formulated for positive streams only – see Section 5 for an extension to negative outcomes. We impose

**Axiom 6 (Positivity)** For all \( c \in C \), \( c \succsim 0 \), with strict preference for some \( c' \in C \).

**Theorem 3** A preference \( \succsim \) on \( X \) satisfies Positivity, Regularity, \( X \)-Separability, Weak Homotheticity, and Absorbing if and only if it admits a positive unconfined CE representation.

Absorbing implies \( X_s \neq \emptyset \). Without this axiom, we may have \( X_s = \emptyset \), in which case Weak Homotheticity is reduced to Homotheticity. Hence, the axioms of Theorem 3 without Absorbing characterize the DU model as a special case.

\(^{11}\)It is easy to see that, given Risk Preference, part (c) implies homotheticity on \( X_s \), that is, if \( x, y \in X_s \) then \( x \sim y \implies \alpha x \sim \alpha y \).
4.3 Constrained CE Representation

4.3.1 Axioms

Unlike the positive unconstrained CE model, the positive constrained CE model has a non-trivial constraint on cognitive optimization. For instance, we may have $\varphi_t(1) = K < \infty$ for each $t \geq 1$ so that the completely patient discount function $1^{T+1}$ has finite cost $\varphi(1^{T+1}) < \infty$. In this case $1^{T+1} \in \text{eff}(\varphi)$. But it could be that $\varphi(1^{T+1}) > K$ and so $1^{T+1}$ may not be feasible because of the capacity constraint.

The cognitive constraints express themselves behaviorally in the boundary of $X_s$. We strengthen Absorbing to give structure to this boundary.\(^\text{12}\) Denote by $(0, x_0)$ the stream that pays 0 in period 0 and then according to $x$ in subsequent periods. We require that $X_s$ is related to the lower contour set of some indifference curve.

**Axiom 7 (Restricted Boundary)** There is some $c > 0$ such that for all positive $x$,

\[ x \in X_s \iff (0, x_0) \not\succ c. \]

The axiom implies a simple criterion for distinguishing between small and large streams. The axiom is best viewed as a convenient property we would like in our model, as opposed to a behavior we expect to be associated with the magnitude effect.

The theorem stated in this section – the main result of the paper – reveals, however, that the key behavioral content of a non-trivial constraint is a relaxation of $X$-Separability to $X_s$-Separability. That is, when an agent has cognitive constraints, facing high magnitudes does not lead her to behave like a standard agent, as she will violate Separability when her cognitive constraint binds.

4.3.2 Representation Theorem

We establish that:

**Theorem 4** A preference $\succeq$ on $X$ satisfies Positivity, Regularity, $X_s$-Separability, Weak Homotheticity, and Restricted Boundary if and only if it admits a positive constrained CE representation.

\(^{12}\)As shown in Lemma 19 in Appendix E, Restricted Boundary implies Absorbing.
Before providing a proof sketch, we note that the CE (constrained and unconstrained) model has strong uniqueness properties, inherited from the separability of the representation on the subdomain \( X_s \) and because \( u(0) = 0 \) is featured in the representation.

**Theorem 5** If there are two positive constrained CE representations or two positive unconstrained representations \( (u^i, \{\varphi^i_t\}, K^i), i = 1, 2 \) of the same preference \( \succsim \), then there exists \( \alpha > 0 \) such that (i) \( u^2 = \alpha u^1 \), (ii) \( \varphi^2_t = \alpha \varphi^1_t \), and (iii) \( K^2 = \alpha K^1 \).

As in the case of the DU model, the optimal discount function \( D_x \) for each \( x \) is uniquely determined from preference. Specifically, \( D_x(t) \) is identified by \( \beta c \sim c^i \), for \( c = x_t \). The theorem also ensures that the curvature or elasticity of \( \varphi_t \) and also the empathy cap \( K \) are uniquely derived from preference.

### 4.4 Proof Outline

A proof sketch is as follows. Regularity and \( X_s \)-Separability yields an additively separable representation \( U(x) = \sum U_t(x_t) \) on the space of small rewards \( X_s \). This representation can be rewritten in the obvious way (that is, define \( u(c) = U_0(c) \) and \( D_x(t) = \frac{U_t(x_t)}{u(x_t)} \)) so that it looks like a discounted utility as in the desired representation, with the discount function \( D_x \) dependent on the stream. Since \( u \) and \( D_x \) is given, we can use the first order condition \( u(x_t) = \varphi'(D_x(t)) \) for each \( t \) to obtain an additive cost function \( \varphi = \sum \varphi_t \) for which \( D_x \) is optimal. This yields a representation close to the desired one on the space of small rewards \( X_s \) (the constraint remains to be defined).\(^{13}\) The remaining problem is as follows. For any positive stream, Weak Homotheticity implies that along a ray \( \{\lambda x : \lambda > 0\} \), as \( \lambda \) increases, \( D_x \) should first strictly increase and eventually become constant once \( \lambda x \) crosses the boundary of \( X_s \). The main step in proving the theorem is to find a cognitive constraint \( \Lambda_x \) for which the optimal \( D \) has this property. An arbitrary closed and convex set \( \Lambda_x \subset [0,1]^T \) satisfying \( \Lambda_x = \Lambda_{\lambda x} \) for all \( \lambda \) does not define a model that is consistent with Weak Homotheticity.\(^{14}\)

---

\(^{13}\)See Section 5 for details about the extension to non-positive streams.

\(^{14}\) Fix \( x \) and let \( \Lambda_x = \Lambda_{\lambda x} = \Lambda \). Weak Homotheticity requires that as stakes are increased, the discount function eventually ceases to change. But without additional structure on \( \Lambda \), this property may not be obtained. To see this suppose \( D_x \) is optimal for
find that the cognitive constraint

$$\Lambda_x = \{ D \in [0, 1]^{T+1} : \sum_{t \geq 0} \varphi_t(D(t)) \leq K_x \}$$

does the job, and arrive at the following general representation:

**Definition 3 (General CE Representation)** A (positive) General CE representation is a tuple $(u, \{ \varphi_t \}, K_x)$ such that $\succeq$ is represented by the function $U : X \to \mathbb{R}$ defined by

$$U(x) = D_x \cdot u(x),$$

s.t. $D_x = \arg \max_{D \in [0, 1]^{T+1}} \{ D \cdot |u(x)| - \varphi(D) \}$ subject to $\varphi(D) \leq K_x$.

Here, the cost function $\varphi_t$, $t \geq 1$, is increasing and convex (and not necessarily CRRA) and the cognitive resource is a function $K : X \to \mathbb{R}_{++}$. The cognitive resource $K_x$ can be computed via integrals of functions derived from behavior, but it is difficult to find an elegant axiom that would imply that $K_x$ is a constant function. However, if the cost function is a homogeneous function then we find that Restricted Boundary imposes that $K_x$ is constant: Weak Homotheticity (c) restricts the cost function to be a homogeneous function, and as noted earlier in the context of the reduced form of the model, a very convenient property of the resulting model is that the cost $\sum_{t=1}^T \varphi_t(D_x(t))$ of a small reward $x$ is proportional to $U(x) - u(x_0)$. Consequently, Restricted Boundary requires that the capacity constraint must be independent of $x$, as desired.

$x$, that is, it satisfies $D_x \cdot u(x) - \varphi(D_x) > D \cdot u(x) - \varphi(D)$ for all $D \in \Lambda$ (the strict inequality comes from the strict convexity of the cost function, which yields a unique maximizer). This can be rewritten as

$$D_x \cdot u(x) - D \cdot u(x) > \varphi(D_x) - \varphi(D)$$

for all $D \in \Lambda$. However, suppose $D_x \cdot u(x) - D \cdot u(x) < 0$ for some $D \in \Lambda$. Exploiting the linearity of $u$, it is readily seen that scaling up $x$ to $\lambda x$ for $\lambda > 1$ can lead to the inequality

$$D_x \cdot u(\lambda x) - D \cdot u(\lambda x) < \varphi(D_x) - \varphi(D).$$

Consequently, even if $D_x$ is on the boundary of $\Lambda$, scaling up rewards may change the agent’s discount function in a way inconsistent with Weak Homotheticity.
4.5 Limiting Cases

4.5.1 DU Model

The DU model is characterized by Theorem 1, which corresponds to the case where $X_s = \emptyset$. Since the CE model assumes $X_s \neq \emptyset$, they are not nested to each other. However, the DU model can be interpreted as a limiting case of both the unconstrained and constrained CE model.

**Proposition 2** Assume that $U : X \to \mathbb{R}$ is a positive unconstrained CE representation $(u, \{\varphi_t\}, K)$ with $\text{eff}(\varphi_t) = [0, \overline{d}_t]$. If either (a) $a_t \to 0$ for all $t \geq 1$ or (b) $\overline{d}_t < 1$ for all $t \geq 1$ and $m \to \infty$, then, for all positive streams $x$, the optimal discount factor $D_{u(x_1)}(t)$ satisfies

$$D_{u(x_1)}(t) = \overline{d}_t$$

for all sufficiently small $a_t$ or sufficiently large $m$, respectively.

If either (a) or (b) holds, the cost function converges to zero on the effective domain, and the agent eventually chooses a maximal discount factor $\overline{d}_t$. Intuitively, $X_s$ converges to an empty set as either $a_t \to 0$ or $m \to \infty$. Thus, as suggested by Theorem 1, the DU model can be obtained as a limiting case of a positive unconstrained CE representation.

We also have an approximation result for the constrained CE model. Because of the joint restriction on the parameters of the constrained CE model (condition (iii) in Section 2.1), the parameters $((a_t), m, (\overline{d}_t), K)$ cannot be chosen independently. From condition (iii), $a_t$ may be removed from the model as $a_t = K / \overline{d}_t^m$ for all $t \geq 1$. By substituting it into the cost function, we have

$$\varphi_t(d) = K \left( \frac{d}{\overline{d}_t} \right)^m.$$  \hspace{1cm} (5)

Moreover, the capacity constraint is reduced to

$$\sum_{t \geq 1} \left( \frac{D(t)}{\overline{d}_t} \right)^m \leq 1.$$

Thus, $K$ is independent of the capacity constraint and is interpreted as a cost parameter.
Proposition 3 Assume that $U : X \to \mathbb{R}$ is a positive constrained CE representation $(u, \{\varphi_t\}, K)$ with $\text{eff}(\varphi_t) = [0, \overline{d}_t]$. Then, for all positive streams $x$, the optimal discount factor $D_{u(x_1)}(t)$ satisfies

$$D_{u(x_1)}(t) \to \overline{d}_t$$

as $m \to \infty$ while holding $K$ and $\overline{d}_t$ fixed.

This proposition is a counterpart of part (b) of Proposition 2. Since $\varphi_t(d) = K \left(\frac{d}{m}\right)^m$, it converges to zero on the effective domain except for $\overline{d}_t$. Moreover, all discount functions except for $(\overline{d}_t)_{t=1}^T$ become feasible in the capacity constraint as $m \to \infty$. Thus, an optimal discount function for any positive stream can get arbitrarily close to $(\overline{d}_t)_{t=1}^T$. 

4.5.2 Myopic Agent

Next, we consider under what conditions on parameters the agent becomes more impatient. Consider higher cognitive costs ($a_t \to \infty$ or $\overline{d}_t \to 0$) or lower cognitive capacity for empathy ($K \to 0$). By condition (iii) in the definition of a regular tuple, we have $a_t = K / \overline{d}_t^m$, $t \geq 1$, and can consider the case where $a_t \to \infty$ and $\overline{d}_t \to 0$ while holding $K$ fixed. As stated above, $\varphi_t(d) = K \left(\frac{d}{m}\right)^m$, and the capacity constraint is independent of $K$. Proposition 1 implies that for a positive small stream, its optimal discount factor at time $t \geq 1$ satisfies

$$D_{u(x_1)}(t) = \left(\frac{\overline{d}_t^m u(x_1)}{mK}\right)^{\frac{1}{m-1}}. \quad (6)$$

As $\overline{d}_t \to 0$, $D_{u(x_1)}(t) \to 0$. Since $D_x(0) = 1$, this limit case corresponds to a completely myopic agent. For large streams, it is easy to see that the representation can be written as

$$U(x) = u(x_0) + \left\{ \sum_{t \geq 1} (\overline{d}_t u(x_t))^{\frac{m}{m-1}} \right\}^{\frac{m-1}{m}}. \quad (7)$$

Again, the agent becomes completely myopic as $\overline{d}_t \to 0$. 

22
If $K \to 0$ and $\bar{d}_t \to 0$ while holding $a_t$ fixed, Proposition 1 implies that all positive streams become large streams. From (7), the agent becomes myopic as $\bar{d}_t \to 0$.

Finally, consider the case where $a_t \to \infty$ and $K \to \infty$. Proposition 1 implies that all positive streams become small streams. From (6), we see that the agent becomes myopic as $a_t = K/\bar{d}_t \to \infty$.

4.5.3 Max-max-type Model

If $K \to 0$ and $a_t \to 0$ while holding $\bar{d}_t$ fixed, the constrained CE model is reduced to the case where cost functions vanish to zero and only the capacity constraint remains the same, that is,

$$U(x) = D_x \cdot u(x), \text{ where } D_x = \arg \max D \cdot u(x)$$

subject to $D = \left\{ D \in [0, 1]^{r+1} : \sum_{t \geq 1} \left( \frac{D(t)}{d_t} \right)^m \leq 1 \right\}$,

or $U(x) = \max_{D \in D} D \cdot u(x)$. It is also clear from Proposition 1 because all positive streams become large if $K \to 0$. Proposition 1 also implies that (8) is more explicitly written as (7).

5 Extension to Negative Outcomes

5.1 Axioms

First we replace Positivity with:

Axiom 8 (Non-Triviality) There exist $c, c' \in C$ s.t. $c \succ 0 \succ c'$.

The vNM structure implied by Risk Preference was useful in defining the notion of “changes in stakes” but it is also useful in identifying positive and negative rewards that have the same “absolute value”. Formally, we say that $c^*$ is an absolute value of $c$ if\footnote{For simplicity, we use this terminology rather than the more accurate one that $c^*$ has the same absolute value as $c$. This terminology anticipates the fact that in terms of the representation $c, c^*$ will satisfy $u(c^*) = |u(c)|$.} $c$ is positive and satisfies $c^* \sim c$, or $c$ is negative and satisfies

$$\frac{1}{2}c + \frac{1}{2}c^* \sim 0.$$

23
That is, the 50-50 lottery over \( c \) and \( c^* \) is as good as receiving 0. Roughly speaking, \( c^* \) is equally distant from 0 in terms of preference as is \( c \). For instance, a risk averse agent may find that losing \$m \) has the same absolute value as gaining \$2m \).

For any stream \( x \), define \( x^* \) by the stream that replaces each outcome \( x_t \) with an absolute value \( (x_t)^* \), that is,

\[
x_t^* = (x_t)^* \quad \text{for all } t.
\]

Note that the absolute value \( c^* \) is not unique, since anything in its indiﬀerence class will also be an absolute value for \( c \).

Our earlier axioms below were formulated for positive streams only. We deﬁne a negative stream as one that delivers negative outcomes and we impose a Symmetry axiom so that axioms on positive streams translate into restrictions on negative ones.

Our main axiom in this section states that:

**Axiom 9 (Symmetry)** If \( x \) is a negative stream then

\[
(c_x)^* \sim c_{x^*}.
\]

Consider a negative stream \( x \) and its present equivalent \( c_x \). The axiom states that the absolute value \( (c_x)^* \) of this present equivalent is as good as the present equivalent \( c_{x^*} \) of the stream’s absolute value \( x^* \). For instance, to take our earlier example, if the agent considers \(-20 \) today as good as \(-5 \) in every period, then for him \( 40 \) today is as good as \( 10 \) in every period. That is, if a positive and negative stream give outcomes with identical absolute value, then their present equivalents have the same absolute value as well.

As noted earlier, present equivalents carry information about the agent’s assessment of the outcomes and his impatience towards them. So the axiom suggests that the agent’s impatience towards two streams is identical when the streams give outcomes that have identical absolute values. This suggests that if impatience is not constant, it can only change with the absolute value of outcomes.

Finally, we re-deﬁne the set of small streams \( X^*_s \):

\[
X^*_s = \{ x \in X : \alpha c_{x^*} > \alpha x^* \text{ for all } \alpha \in (0, 1) \text{ and } x^* \}.
\]

That is, an arbitrary stream \( x \) is small if its absolute value \( x^* \) is such that small changes in scale leads to changes in impatience (as with Symmetry, this
presumes that impatience depends only on the absolute value of the stakes). Note that
\[ x \in X^*_s \iff x^* \in X_s, \ \forall x. \]

This impacts Restricted Boundary and \( X_s \)-Separability axioms, which henceforth are re-defined as follows:

**Axiom 10 (Restricted Boundary\(^*\))**  There is some \( c \succ 0 \) such that
\[ x \in X^*_s \iff (0, x^*_{-0}) \preccurlyeq c. \]

**Axiom 11 (\( X^*_s \)-Separability)**  For all \( x \in X^*_s \) and all \( t, t^\prime \),
\[ \frac{1}{2}c_{x_{t0}} + \frac{1}{2}c_{x_{t\prime}} \sim \frac{1}{2}c_x + \frac{1}{2}c_0. \]

Conspicuously missing is an axiom describing streams that offer both positive and negative consumption. Because of \( X_s \)-Separability, it turns out to be unnecessary to formulate such an axiom for small streams, since we can restrict attention to dated rewards (streams \( c^t \) that pay \( c \) at \( t \) and 0 otherwise) for much of our analysis, and such streams are either positive or negative.

For large streams, we impose the following axiom:

**Axiom 12 (Large Homotheticity)**  For all non-positive and non-negative streams \( x \notin X^*_s \) and \( \alpha \in (0, 1) \),
\[ \alpha x \notin X^*_s \implies \alpha c_x \sim \alpha x. \]

Note that Weak Homotheticity implies this property for large positive streams. Together with Symmetry, the same holds also for large negative streams. The axiom requires the same property for the other large streams. On \( X^*_s \), \( \alpha c_x \succ \alpha x \) holds for positive streams by Weak Homotheticity. Since small losses in the future are more discounted than losses in the present, presumably \( \alpha x \succ \alpha c_x \) for negative streams. The ranking between \( \alpha c_x \) and \( \alpha x \) is in general ambiguous for other small streams.
5.2 Representation

We establish that:

**Theorem 6** A preference $\succapprox$ on $X$ satisfies Non-triviality, Regularity, $X^*_s$-Separability, Weak Homotheticity, Restricted Boundary*, Symmetry, and Large Homotheticity if and only if it admits a CE Representation.

Moreover, if there are two CE representations $(u^i, \phi^i, K^i), i = 1, 2$ of the same preference $\succapprox$, then there exists $\alpha > 0$ such that (i) $u^2 = \alpha u^1$, (ii) $\phi^2 = \alpha \phi^1$, and (iii) $K^2 = \alpha K^1$.

6 Application: Accommodating Evidence

6.1 Magnitude Effect

Loewenstein and Prelec [22] and subsequent work use the curvature of utility for money to incorporate the magnitude effect: since

$$ (x, 0) \sim (y, t) \implies u(x) = D(t)u(y) \implies \frac{x}{y} = \frac{u^{-1}(D(t)u(y))}{y}, $$

the computed discount factor $\frac{x}{y} = \frac{u^{-1}(D(t)u(y))}{y}$ can exhibit the magnitude effect. Noor [27] provides a calibration theorem to show that the curvature of utility is not an adequate explanation.\(^{16}\) Benhabib et al [5] show that preference reversals and the magnitude effect can be explained simultaneously by a fixed cost of waiting, which inhibits waiting for smaller delayed rewards more substantially than larger delayed rewards.\(^{17}\) This, however, does not explain magnitude effects that exist when all rewards used to compute the discount factor (including the money-equivalent for the noted delayed rewards) are delayed (Ericson and Noor [12]).

\(^{16}\)The calibration theorem implies that for an arbitrary discount function, concave utility, and arbitrary background stream of wealth, if the agent exhibits, say, $\$15$ now $\sim \$60$ in a year then the following must hold: give $\$60$ in a year unconditionally to the agent, and she will subsequently reject $\$4x = \$60 \times \frac{15}{x}$ return in a year at costs of $\$x$ today, for any value of $x$.

\(^{17}\)To illustrate, assume linear $u$ and suppose that waiting for any reward involves a constant fixed utility cost $c$. Then an indifference between (money, time) pairs imply the following: $(x, 0) \sim (y, t) \implies x = y - c \implies \frac{c}{y} = 1 - \frac{c}{y}$, where $\delta_y := \frac{c}{y}$ is increasing in $y$. Note that the extension of the model $U(x, t) = u(x) - c(t)$ with $c$ increasing is indistinguishable from $V(x, t) = e^{U(x, t)} = e^{u(x) - c(t)} := D(t)v(x)$.

26
In our model, the discount function for a small reward $c$ is given by

$$D_c(t) = \gamma(t) |u(c)|^{\frac{1}{m-1}},$$

which is increasing in $u(c)$. This expresses itself behaviorally as in Thaler (1981) under the assumption that the utility for money is approximately linear. This assumption can be justified by the calibration theorem in Noor (2011).

While not as well established as the magnitude effect for gains, Hardistry et al [19] show that subjects tend to exhibit a reverse magnitude effect for losses. In fact, they find that subjects would prefer incur extra loss immediately, thereby exhibiting negative discount rates. For instance, they may prefer losing $9.10 now over losing $9 in a week, while preferring a loss of $110 in a week to losing $100 now, exhibiting high patience for small losses and low impatience for large ones. The natural interpretation is that the agent is more willing to take an immediate small loss because it is not so costly to avoid having a future loss looming over her head. We believe this behavior is best understood in terms of anticipatory emotions such as dread (Loewenstein and Prelec [23]) rather than empathy, and therefore do not try to accommodate it in the CE model, which we have formulated so that a magnitude effect exists for losses as well.

6.2 Preference Reversal

A notable finding in the experimental literature is that of preference reversals (Fredrick et al [13]). A preference reversal exists if there is $c < \hat{c}$ and some $d > 0$ such that

$$c \succ (\hat{c})^d \quad \text{and} \quad (c)^t \succ (\hat{c})^{t+d} \quad \text{for some } t > 0,$$

and such behavior is often viewed as evidence of present bias. Our model explains such behavior as follows.\(^{18}\) If empathy is sufficiently costly, the current self will be selfish and may well prefer the immediate option as in the

\[^{18}\text{When } (\hat{c})^d \text{ is small, the two comparisons are } u(c) \text{ vs } \gamma(d)u(\hat{c})^{\frac{1}{m-1}} \text{ and } u(c) \text{ vs } \left(\frac{\gamma(t+d)}{\gamma(t)}\right)^{\frac{m-1}{m}} u(\hat{c}). \text{ If } (\hat{c})^d \text{ is large then they are } u(c) \text{ vs } (mK)^\frac{1}{m} \gamma(d)^{\frac{m-1}{m}} u(\hat{c}) \text{ and } u(c) \text{ vs } \left(\frac{\gamma(t+d)}{\gamma(t)}\right)^{\frac{m-1}{m}} u(\hat{c}). \text{ Depending on parameter values the agent can exhibit preference reversals or their converse.}\]
first comparison. If the cost function for empathizing with self \( t \) is not much different from that for self \( t + d \), the agent empathizes with self \( t \) and self \( t + d \) to a similar degree, and therefore prefers the higher outcome, thereby generating the reversal.

The attention received by preference reversals in the literature notwithstanding, it should be noted that there is also substantial evidence of the reverse (Fredrick et al \[13\]):

\[
c \prec (\hat{c})^d \quad \text{and} \quad (c)^t \succ (\hat{c})^{t+d} \text{ for some } t > 0.
\]

This behavior is perfectly natural from the perspective of our model. Impatience is sufficient to explain the choice of sooner smaller delayed reward, \((c)^t \succ (\hat{c})^{t+d}\), and this impatience arises because her empathy for future selves is decaying at some rate. On the other hand, if the period 0 self has high empathy for the first \( d \) selves, then intuitively speaking, the agent’s current self “extends” that many periods into the future, and therefore a large reward \((\hat{c})^d\) may be chosen over a smaller immediate reward \(c\).

While such behavior is frequently observed in experiments, it is all but ignored in theory and applications. This is presumably because this behavior exhibits “future bias” or “increasing impatience”, which appears highly anomalous from the lens of present bias, the dominant way of thinking about time preference in the literature. We note that even the beta-delta can give rise to such behavior: the model is silent on how long period 0 is, and if it is long enough, then the agent would exhibit “future bias”.

### 6.3 Preference for Increasing Sequences

Loewenstein and Prelec \[23\] demonstrate that subjects prefer increasing sequences of consumption to constant or decreasing sequences with the same present value. Limiting attention only to consumption levels that yield positive utility for simplicity, and assuming three periods, our next proposition shows that if self 2 is better off than self 1, under certain conditions the agent may be willing to reduce self 1’s welfare to improve self 2’s further, suggesting a preference for increasing sequences. Intuitively, the convex transformation \(m_m\) can cause the marginal utility at time \( t \) to be increasing.\(^{19}\)

---

\(^{19}\)Note that whether marginal utility from consumption increases or not depends on both \(\frac{m}{m-1}\) and the curvature of vNM function \(u\) over consumption. In the following proposition, we assume that \(u\) is linear over consumption.
Proposition 4  Suppose $C = \mathbb{R}_+$ and $u$ is linear. If $c_1 < c_2$ and if $a_2/a_1$ is sufficiently close to one, then for all $\varepsilon$ in some positive interval,\(^{20}\)

$$(c_0, c_1 + \varepsilon, c_2 - \varepsilon) \prec (c_0, c_1 - \varepsilon, c_2 + \varepsilon).$$

6.4 Non-Separability: Preference for Spread

Suppose there are only three time periods and that consumption in the final period $2$ is fixed at $c_2$. Consider a prospect of consuming $c_1$ in period $1$ and let the present equivalent $p(c_1; c_2)$ denote the amount received today that would make him indifferent to it:

$$(p(c_1; c_2), 0, c_2) \sim (0, c_1, c_2).$$

We show how the present equivalent for $c_1$ can depend on the value of $u(c_2)$.

Proposition 5  If $(0, c_1, c_2)$ is small (resp large), then $p(c_1; c_2)$ is constant (resp. decreasing) in $u(c_2)$.\(^{21}\)

The intuition is simply that if the agent’s empathy constraint is binding, then higher values of $c_2$ causes a reallocation of the limited empathy away

---

**Proof.** The desired preference obtains if:

$$\gamma(1)[(u(c_1) + \varepsilon)^{\frac{m}{m-1}} - (u(c_1) - \varepsilon)^{\frac{m}{m-1}}] < \gamma(2)[(u(c_2) + \varepsilon)^{\frac{m}{m-1}} - (u(c_2) - \varepsilon)^{\frac{m}{m-1}}].$$

Due to convexity $\frac{m}{m-1} > 1$, there is some $e > 0$ s.t. the inequality holds for all $e < \varepsilon < u(c_1)$. \(\blacksquare\)

**Proof.** If $(0, c_1, c_2)$ is small, then we are in the additively separable part of the model and so $p(c_1; c_2)$ is constant for small changes in $c_2$. Suppose $(0, c_1, c_2)$ is large. Using the non-additive part of the representation, we get that

$$u(p) = (mK)^{\frac{1}{m}} \left( (\gamma(1)u(c_1)^{\frac{m}{m-1}} + \gamma(2)u(c_2)^{\frac{m}{m-1}})^{\frac{m-1}{m}} - \gamma(2)^{\frac{m-1}{m}}u(c_2) \right).$$

Letting $a = \gamma(1)u(c_1)^{\frac{m}{m-1}} > 0$ and $b = \gamma(2)u(c_2)^{\frac{m}{m-1}} > 0$, we see that

$$u(p) = (mK)^{\frac{1}{m}} \left( (a + b)^{\frac{m-1}{m-1}} - b^{\frac{m-1}{m-1}} \right),$$

which is decreasing and convex in $b$. In particular, $p$ is decreasing in $u(c_2)$. \(\blacksquare\)
from self 1 and towards self 2, making her care less about (more impatient towards) receiving $c_1$. We are not aware of any direct evidence of such an effect, but we note next that it is consistent with the experimental finding of “preference for spread” documented in Loewenstein and Prelec [23]. In particular, there can exist $u(c) < u(c')$ (such that $(0, c', c')$ is a large stream) and

$$(c, 0, 0) \prec (0, c', 0).$$

$$(c, 0, c') \succ (0, c', c').$$

Increasing final period consumption from 0 to $c'$ caused the agent to become more impatient, as in the above proposition. The preference pattern suggests a preference for spread of the type discussed in Loewenstein and Prelec [23] since the agent appears to reverse her initial preference for $c'$ tomorrow over $c$ today in order to spread out good consumption.22

6.5 Other

The sign effect refers to the finding that people are more patient when dealing with a loss of $c$ than a gain of $c$ (see Hardistry et al [19] for a recent reference). If $|u(-c)| > u(c)$ then we obtain

$$D_{-c}(t) = \gamma(t)|u(-c)|^{\frac{1}{m-1}} > \gamma(t)u(c)^{\frac{1}{m-1}} = D_c(t),$$

Dohmen et al [8] show that people with lower cognitive abilities are more impatient. In our model, higher cognitive costs $\varphi$ or lower cognitive capacity for empathy $K$ correspond to greater impatience. See Section 4.5.2 for related comparative statics.

Finally, there is evidence of negative discount rates for both positive and negative outcomes. For instance, subjects in Loewenstein and Prelec [23] would rather have a fancy french dinner later than sooner, presumably because delay comes with the pleasure of anticipation. In Hardistry et al [19], subjects would pay a little extra to incur a small loss immediately rather than with delay, presumably so as to avoid having it loom over their heads.

---

22It does not exactly match Loewenstein and Prelec [23], who ask how subjects would spread opportunities to have dinner at a fancy french restaurant $F$ rather than at home $H$ and finds that a majority exhibit $(F, H, H) \prec (H, F, H)$ and $(F, H, F) \succ (H, F, F)$. Each of these preferences require $D > 1$, which can be interpreted as anticipation. While we can allow this in our model, we do not do so for reasons specified in Sections 2.1 and 6.5.
We can readily allow for $D > 1$ in our model (the Impatience axiom does not play a significant role in our axiomatization) in which case an exceedingly high degree of empathy can be achieved with sufficient cognitive effort. However, we feel that negative discount rates are better understood in terms of anticipatory emotions like dread (as in Loewenstein and Prelec [23]) rather than high empathy.\footnote{See Galperti and Strulovici [16] for a recent model of anticipation, where self 0 cares directly for self 1 and self 2 but also cares for self 2 indirectly through self 1’s care for self 2. In our model there are no indirect connections.}

7 Application: Procrastination

While our model is static, the question of how to extend it to a dynamic choice setting is the same as for the multiple selves model: one may view the dynamic model as a game between selves (Laibson [21]), or one may turn to menus (Gul and Pesendorfer [18]). We leave a dynamic extension of the model to future research. We illustrate here that magnitude-dependence suggests a new avenue for an ex ante to influence future selves: by increasing the magnitude of the stakes (such as by bundling choices together), the future self is incentivized to behave more patiently.

Suppose the time horizon consists of period 0, 1 and 2, and there are two tasks that can be completed only in periods 0 and 1. Each task takes one period to complete, and two tasks can be completed in one period as well. Each requires effort $e < 0$ and returns a benefit $R > 0$ in period 2, regardless of when it was completed. A period in which no effort is exerted nor a benefit received yields the agent utility $u(0) = 0$. The utility cost and benefit of $n$ tasks, for $n = 1, 2$, is $u(ne) < 0$ and $u(nB) > 0$, where $u$ is weakly concave. We study the dynamic behavior of DU agents and CE agents in this context with assuming that they are either naive or sophisticated.

Consider a DU agent, with possibly time inconsistent preferences. Suppose that self 1 would not exert the effort to complete any one task when two tasks are remained to be completed:

$$V_1(1 \text{ task}) = u(e) + D(1)u(R) < 0 = V_1(0 \text{ task}).$$

Given weak concavity of $u$, $u(2R) \leq 2u(R)$ and $u(2e) \leq 2u(e)$. Thus, self 1
will not do both tasks together, since
\[ u(e) + D(1)u(R) < 0 \]
\[ \implies 2u(e) + D(1)2u(R) < 0 \]
\[ \implies V_1(2 \text{ tasks}) = u(2e) + D(1)u(2R) < 0. \]

If one task has been completed by self 0, self 1 compares to complete one
more task (yielding \( u(e) + D(1)u(2R) \)) with not to do (yielding \( D(1)u(R) \)).
Since \( u(2R) \leq 2u(R) \), we have
\[ u(e) + D(1)(u(2R) - u(R)) \leq u(e) + D(1)u(R) < 0, \]
which implies \( u(e) + D(1)u(2R) < D(1)u(R) \). Again, self 1 does not under-
take any task.

Self 0 would not do any of the tasks either since, given \( D(2) \leq D(1), \)
\[ V_0(1 \text{ task}) = u(e) + D(2)u(R) \leq u(e) + D(1)u(R) < 0 \]
and \( V_0(2 \text{ tasks}) = u(2e) + D(2)u(2R) \leq u(2e) + D(1)u(2R) < 0. \)

Therefore, we see that in the DU model, if self 1 is not willing to complete
one task, then the agent will not undertake any of the tasks.

This conclusion does not change even when the agent is sophisticated. If
self 0 postpones 2 tasks to period 1, self 1 does not undertake any task as
shown above. If self 0 completes only one task and postpones the other task
to self 1, the above observation implies that self 1 does not undertake the
other task. Since a sophisticated self 0 anticipates these behaviors of self 1,
his optimal choice is not to complete any task.

The behavior of the CE agent can differ substantially. On the small
streams, the agent has a discount factor \( D_{\text{ue}}(t) = \gamma(t)|u(e)|^{-\frac{1}{\alpha}} \). For sim-
plicity, let \( \gamma(t) = \delta^t \) where \( 0 < \delta \leq 1 \). When two tasks are remained to be
completed, it is possible that the CE self 1 would not exert effort to do only 1
task, but would exert the necessary effort to complete two simultaneously:\footnote{For instance if \( u \) is linear then, because \( 2 \frac{m}{m-\alpha} > 2 \), \( u(2e) + \delta^2 u(2R) \frac{m}{m-\alpha} = 2u(e) + 2 \frac{m}{m-\alpha} \delta^2 u(R) \frac{m}{m-\alpha} \) can be positive even if \( u(e) + \delta u(R) \frac{m}{m-\alpha} < 0. \)}
\[ U_1(1 \text{ task}) = u(e) + \delta u(R) \frac{m}{m-\alpha} < 0 = U_1(0 \text{ tasks}) \]
\[ U_1(2 \text{ tasks}) = u(2e) + \delta u(2R) \frac{m}{m-\alpha} \geq u(2e) + \delta^2 u(2R) \frac{m}{m-\alpha} > 0. \]
The intuition for this comes from the magnitude effect: the agent is more patient when there are larger outcomes for future selves. We show that a sophisticated CE self 0 would procrastinate even if she was willing to do both tasks by herself.

If the agent is naive, self 0’s utilities from doing one or both tasks are

\[
U_0(1 \text{ task}) = u(e) + \delta^2 u(R) \frac{m}{m-1} < 0 \quad \text{and} \quad U_0(2 \text{ tasks}) = u(2e) + \delta^2 u(2R) \frac{m}{m-1} > 0,
\]

where the cost is paid immediately and the benefit received in 2 periods. As with self 1, the magnitude effect can lead self 0 to be willing to do both tasks even if she is unwilling to do one, i.e., it is possible that

\[
U_0(2 \text{ tasks}) > 0 = U_0(0 \text{ tasks}) > U_0(1 \text{ task}).
\]

But even if she would be willing to do both tasks, when she is sophisticated, she may procrastinate in order to exploit self 1’s willingness to complete both tasks, in which case she receives

\[
U_0(0 \text{ tasks}) = u(0) + \delta |u(2e)| \frac{1}{m-1} u(2e) + \delta^2 u(2R) \frac{m}{m-1} \\
= \delta |u(2e)| \frac{1}{m-1} u(2e) + \delta^2 u(2R) \frac{m}{m-1} \\
> u(2e) + \delta^2 u(2R) \frac{m}{m-1} \quad \text{(because } \delta |u(2e)| \frac{1}{m-1} \leq \frac{1}{\delta_1} < 1) \\
= U_0(2 \text{ tasks}).
\]

It follows that the agent would procrastinate even if \(U_0(2 \text{ tasks}) > 0\).\(^{25}\)

\[\]

A Appendix: Implications of the CE Representation

A.1 Proof of Proposition 1

First, we solve the cognitive optimization problem for each \(x\). Let \(\varphi_t(d) = a_t d^m\) on \(d \in [0, \tilde{d}_t]\). As explained in Section 2.1, the boundary constraint

\[\]

25If self 0 completes only one task and postpones the other task to self 1, self 1 may or may not complete the task. In the former case, \(U_0(1 \text{ task}) = u(e) + \delta |u(e)| \frac{1}{m-1} u(e) + \delta^2 u(2R) \frac{m}{m-1}\). Since \(u_{\frac{m}{m-1}}\) is not necessarily weakly concave, we may have \(U_0(1 \text{ task}) > U_0(0 \text{ tasks})\), in which case the agent postpones one task to period 1.
$D(t) \leq \bar{d}_t$ is effectively ignored by condition (iii) of regularity. For each $x$, we know $D_x(0) = 1$ because $\varphi_0(d) = 0$ for all $d \in [0,1]$. For each $x$, an optimal discount function $\{D_x(t)\}_{t \geq 1}$ is determined by

$$\max_{D \geq 0} \sum_{t \geq 1} D(t)u(x_t) - \sum_{t \geq 1} \varphi_t(D(t)),$$

subject to $\sum_{t \geq 1} \varphi_t(D(t)) \leq K$.

The F.O.C. of the above maximization problem is obtained as the F.O.C. of the following Lagrangian:

$$L = \sum_{t \geq 1} D(t)u(x_t) - \sum_{t \geq 1} a_tD(t)^m + \lambda(K - \sum_{t \geq 1} a_tD(t)^m),$$

where $\lambda \geq 0$ is a Lagrange multiplier for the capacity constraint. By differentiating $L$ with respect to $D(t)$, we have

$$D_x(t) = \left( \frac{u(x_t)}{(1 + \lambda)ma_t} \right)^{\frac{1}{m-1}}, \quad (9)$$

for all $t = 1, \cdots, T$.

Suppose $x$ is small. Since the capacity constraint is not binding, we have $\lambda = 0$. Thus,

$$D_x(t) = \left( \frac{u(x_t)}{ma_t} \right)^{\frac{1}{m-1}}$$

and

$$U(x) = u(x_0) + \sum_{t=1}^T D_x(t)u(x_t) = u(x_0) + \sum_{t=1}^T \gamma(t)u(x_t)^{\frac{m}{m-1}},$$

where $\gamma(t) = (ma_t)^{-\frac{1}{m-1}}$.

Next, suppose $x$ is large. Then, the capacity constraint is binding. By substituting (9) into the capacity constraint,

$$\sum_{t=1}^T a_t \left( \frac{u(x_t)}{(1 + \lambda)ma_t} \right)^{\frac{m}{m-1}} = K.$$
By rearrangement,

\[
\frac{1}{(1 + \lambda)^{\frac{m-1}{m}}} = \frac{K^\frac{1}{m}}{\left\{ \sum_{t=1}^{T} a_t \left( \frac{u(x_t)}{m \alpha_t} \right)^{\frac{m}{m-1}} \right\}^\frac{1}{m}}.
\]

By substituting it into (9),

\[
D_x(t) = \frac{K^\frac{1}{m} \left( \frac{u(x_t)}{m \alpha_t} \right)^{\frac{1}{m-1}}}{\left\{ \sum_{t=1}^{T} a_t \left( \frac{u(x_t)}{m \alpha_t} \right)^{\frac{m}{m-1}} \right\}^\frac{1}{m}} = \frac{(mK)^{\frac{1}{m}}}{\left\{ \sum_{t=1}^{T} \gamma(t) u(x_t)^{\frac{m}{m-1}} \right\}^\frac{1}{m}}.
\]

Therefore,

\[
U(x) = u(x_0) + \sum_{t=1}^{T} D_x(t) u(x_t)
\]

\[
= u(x_0) + (mK)^{\frac{1}{m}} \left\{ \sum_{t=1}^{T} \gamma(t) u(x_t)^{\frac{m}{m-1}} \right\}^{\frac{m-1}{m}}.
\]

Finally, we derive a threshold where small and large streams are distinguished. At this boundary of consumption streams,

\[
\sum_{t=1}^{T} \varphi_t(D_x(t)) = \sum_{t=1}^{T} a_t \left( \frac{u(x_t)}{m \alpha_t} \right)^{\frac{m}{m-1}} = K.
\]

Equivalently,

\[
\sum_{t=1}^{T} \gamma(t) u(x_t)^{\frac{m}{m-1}} = mK.
\]

Therefore, at the boundary,

\[
U(x) = u(x_0) + \sum_{t=1}^{T} \gamma(t) u(x_t)^{\frac{m}{m-1}} = u(x_0) + mK.
\]
A.2 Necessary Axioms for the CE Representation

For each components \((u, \{\varphi_t\}, K)\), a CE representation is given as in Definition 1. As shown in Appendix A.1, its reduced form is obtained as

\[
U(x) = u(x_0) + \sum_{t \geq 1} D_x(t) \cdot u(x_t),
\]

where

\[
D_x(t) = \gamma(t)\left|u(x_t)\right|^\frac{1}{m-1} \tag{10}
\]

if \(U(x^*) - |u(x_0)| \leq mK\), and

\[
D_x(t) = \frac{(mK)^\frac{1}{m} \gamma(t)\left|u(x_t)\right|^{\frac{1}{m-1}}}{\left\{\sum_{\tau=1}^{T} \gamma(\tau)\left|u(x_\tau)\right|^{\frac{m}{m-1}}\right\}^{\frac{1}{m}}} \tag{11}
\]

if \(U(x^*) - |u(x_0)| > mK\).

By definition, \(x \in X^*_s\) if and only if \(x^* \in X_s\). Since \(u(x^*_t) = |u(x_t)|\), for all \(x\), by Proposition 1, if \(U(x^*) - |u(x_0)| > mK\), \(D_{\alpha x}(t)\) is constant for all \(\alpha\) sufficiently close to one, and hence \(\alpha c_x \sim \alpha x\) for such \(\alpha\). If \(U(x^*) - |u(x_0)| \leq mK\), \(D_x(t) = D_{|u(x_t)|}(t)\) is strictly increasing in \(|u(x_t)|\), and hence, \(\alpha c_x \succ \alpha x\) for all \(\alpha \in (0, 1)\). Therefore,

\[
X^*_s = \{x \in X : \sum_{t > 0} \gamma(t)\left|u(x_t)\right|^{\frac{m}{m-1}} \leq mK\}, \tag{12}
\]

which implies Restricted Boundary*.

Note that \(D_x(t)\) is continuous in \(x\). By (10) and (11), \(D_x(t)\) is strictly increasing in \(|u(x_t)|\) in \(X^*_s\); and it is constant on a ray in \(X \setminus X^*_s\). It is obvious to see that \(\succ\) that \(U\) represents satisfies Order, Continuity, Baseline, Impatience, Present Equivalents, Risk Preference, Symmetry, and Large Homotheticity. Since \(D_x(t)\) depends only on \(u(x_t)\) on \(X^*_s\), \(U\) is additively separable on this subdomain, which implies \(X^*_s\)-Separability.

**Lemma 1** \(\succ\) satisfies Monotonicity.

**Proof.** Take any \(x, y\) such that \(u(x_t) \geq u(y_t)\) for all \(t \geq 0\). Since \(U(x)\) is additively separable between \(x_0\) and everything else, it is enough to show Monotonicity for streams \(x, y\) with \(u(x_0) = u(y_0) = 0\). From now on, we consider such streams only.
Step 1:
\[ \sum \gamma(t)|u(x_i)|^{\frac{1}{m-1}} u(x_i) \geq \sum \gamma(t)|u(y_i)|^{\frac{1}{m-1}} u(y_i). \]

For each \( t \), there are three cases: (1) \( u(x_i) \geq u(y_i) \geq 0 \), (2) \( u(x_i) \geq 0 \geq u(y_i) \), and (3) \( 0 \geq u(x_i) \geq u(y_i) \). In any case, \( |u(x_i)|^{\frac{1}{m-1}} u(x_i) \geq |u(y_i)|^{\frac{1}{m-1}} u(y_i) \). By adding up these inequalities across \( t \), we have the desired result.

By Step 1, if \( x \) and \( y \) are small, we have the desired result. From now on, suppose that either \( x \) or \( y \) is large.

Step 2: If
\[ \sum \gamma(t)|u(x_i)|^{\frac{1}{m-1}} u(x_i) \geq 0 \geq \sum \gamma(t)|u(y_i)|^{\frac{1}{m-1}} u(y_i), \quad (13) \]
then \( U(x) \geq U(y) \). Suppose that \( x \) is large. Then, \( U(x) \) is obtained by multiplying the expression in (13) by a positive multiplier, that is,
\[ U(x) = \left( \frac{mK}{\sum \gamma(t)|u(x_i)|^{\frac{1}{m-1}}} \right)^{\frac{1}{m}} \sum \gamma(t)|u(x_i)|^{\frac{1}{m-1}} u(x_i). \]

Since this operation does not change the sign, we have the desired result. The same argument is applicable also when \( y \) is large.

Now assume that
\[ \sum \gamma(t)|u(x_i)|^{\frac{1}{m-1}} u(x_i) \geq \sum \gamma(t)|u(y_i)|^{\frac{1}{m-1}} u(y_i) \geq 0. \quad (14) \]

Step 3: Let \( S = \{ t \geq 1 \mid u(y_t) \leq 0 \} \). For any \( z \in X \) satisfying \( 0 \geq u(z_t) \geq u(y_t) \) for \( t \in S \) and \( u(z_t) = u(y_t) \) elsewhere, \( U(z) \geq U(y) \).

By assumption, for all \( t \in S \), \( 0 \geq u(z_t) \geq u(y_t) \), which implies \( |u(z_t)| \leq |u(y_t)| \). Thus, \( \sum \gamma(t)|u(z_t)|^{\frac{1}{m-1}} \leq \sum \gamma(t)|u(y_t)|^{\frac{1}{m-1}} \). If \( y \) is small, \( z \) must be small as well. Hence, assume that \( y \) is large.

Suppose that \( z \) is large. By representation,
\[ U(z) = \left( \frac{mK}{\sum \gamma(t)|u(z_t)|^{\frac{1}{m-1}}} \right)^{\frac{1}{m}} \sum \gamma(t)|u(z_t)|^{\frac{1}{m-1}} u(z_t). \]

If a consumption stream changes from \( z \) to \( y \), the numerator is decreasing because (14) holds for \( z \) and \( y \), while the denominator is increasing by assumption. Thus, \( U(z) \geq U(y) \).
Next suppose that $z$ is small. Since $y$ is large, by Proposition 1, $mK \leq \sum \gamma(t)|u(y_t)|^{\frac{1}{m-1}}$. Since (14) holds for $z$ and $y$,

$$U(z) = \sum \gamma(t)|z_t|^{\frac{1}{m-1}} u(z_t) \geq \left( \frac{mK}{\sum \gamma(t)|u(y_t)|^{\frac{m}{m-1}}} \right) \sum \gamma(t)|u(y_t)|^{\frac{1}{m-1}} u(y_t) = U(y),$$

as desired.

Now turn to the comparison between $x$ and $y$. Define $S_N = \{ t \geq 1 \mid u(x_t) < 0 \}$ and $S_P = \{ t \geq 1 \mid u(y_t) > 0 \}$. Let $y^d$ be the stream such that $y^d_t = x_t$ on $t \in S_N$, $y^d_t = y_t$ on $t \in S_P$, and $y^d_t = 0$ otherwise. Note that $u(y^d_t) \geq u(y_t)$ for all $t$. Moreover, $0 \geq u(y^d_t) \geq u(y_t)$ for $t \notin S_P$ and $u(y^d_t) = u(y_t)$ elsewhere. By Step 3, $U(y^d) \geq U(y)$.

It is enough to show that $U(x) \geq U(y^d)$. Note that $u(x_t) \geq u(y^d_t) \geq 0$ for all $t \notin S_N$ and $u(x_t) = u(y^d_t)$ elsewhere. From the representation,

$$U(x) = \sum_{t \in S_N} D_x(t)u(x_t) + \sum_{t \notin S_N} D_x(t)u(x_t)$$

and

$$U(y^d) = \sum_{t \in S_N} D_{y^d}(t)u(y^d_t) + \sum_{t \notin S_N} D_{y^d}(t)u(y^d_t).$$

Note that for all $t \in S_N$, $u(x_t) = u(y^d_t) < 0$. We show $U(x) \geq U(y^d)$ by the following two steps.

Step 4: For all $t \in S_N$, $D_x(t) \leq D_{y^d}(t)$.

Since $\sum \gamma(t)|x_t|^{\frac{m}{m-1}} \geq \sum \gamma(t)|y^d_t|^{\frac{m}{m-1}}$, we have either (a) $x$ and $y^d$ are large, or (b) $x$ is large and $y^d$ is small. If (a) holds, for all $t \in S_N$,

$$D_{y^d}(t) = \left( \frac{mK}{\sum_{\tau=1}^{T} \gamma(\tau)|y^d_\tau|^{\frac{m}{m-1}}} \right) \gamma(t)|y^d_t|^{\frac{1}{m-1}} \geq \left( \frac{mK}{\sum_{\tau=1}^{T} \gamma(\tau)|x_\tau|^{\frac{m}{m-1}}} \right) \gamma(t)|x_t|^{\frac{1}{m-1}} = D_x(t).$$

If (b) holds, $D_{y^d}(t) = \gamma(t)|y^d_t|^{\frac{1}{m-1}}$. Since $x$ is large, the multiplier for $D_x(t)$ is smaller than one. Thus we have the desired result.
Step 5: $\sum_{t \in S_N} D_z(t)u(x_t) \geq \sum_{t \in S_N} D_{y^d}(t)u(y^d_t)$.

Since $\sum_{t} \gamma(t)|u(x_t)|^{\frac{m}{m-1}} \geq \sum_{t} \gamma(t)|u(y^d_t)|^{\frac{m}{m-1}}$, we have either (a) $x$ and $y^d$ are large, or (b) $x$ is large and $y^d$ is small. If we define a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ by

$$f(\theta) = \left(\frac{a}{b + \theta}\right)^\frac{1}{m} \theta$$

for some $a, b > 0$, it is easy to verify that $f'(\theta) > 0$, that is, $f$ is a strictly increasing function because $m > 1$. Given this observation, if (a) holds,

$$\sum_{t \in S_N} D_z(t)u(x_t)$$

$$\geq \left(\sum_{t \in S_N} |u(x_t)|^{\frac{m}{m-1}} \right)^\frac{1}{m} \sum_{t \in S_N} \gamma(t)u(x_t)^{\frac{m}{m-1}}$$

$$= \sum_{t \in S_N} D_{y^d}(t)u(y^d_t),$$

as desired.

Suppose that (b) holds. Define $x(\alpha) = \alpha x + (1 - \alpha)y^d$ for all $\alpha \in (0, 1)$. Note that $u(x_t) = u(x_t(\alpha)) = u(y^d_t)$ for all $t \in S_N$ and $u(x_t(\alpha)) \geq u(y^d_t) \geq 0$ otherwise. By continuity of the representation, there exists $\alpha^* \in (0, 1)$ such that $x(\alpha)$ is large if $\alpha > \alpha^*$ and $x(\alpha)$ is small if $\alpha \leq \alpha^*$. By the same argument as above,

$$\sum_{t \in S_N} D_z(t)u(x_t)$$

$$\geq \left(\sum_{t \in S_N} |u(x_t(\alpha^*))|^{\frac{m}{m-1}} \right)^\frac{1}{m} \sum_{t \notin S_N} \gamma(t)u(x_t(\alpha^*))^{\frac{m}{m-1}}$$

$$= \left(\frac{mK}{mK}\right)^\frac{1}{m} \sum_{t \notin S_N} \gamma(t)u(x_t(\alpha^*))^{\frac{m}{m-1}}$$

$$= \sum_{t \notin S_N} \gamma(t)u(y^d_t)^{\frac{m}{m-1}} = \sum_{t \notin S_N} D_{y^d}(t)u(y^d_t).$$
Thus, (b) holds by the same argument as above.\[\]which implies part (a).

From (10) and (11),\[\]

Proof.\[\]

Lemma 2 \(\preceq\) satisfies Weak Homotheticity.

Proof. From (10) and (11), \(D_x(t) \geq D_{\alpha x}(t)\) for all \(t \geq 1\) and \(\alpha \in (0, 1)\). Thus, \(\alpha c_x \preceq \alpha x\) if and only if \(\alpha U(x) \geq U(\alpha x)\) if and only if

\[
\alpha(u(x_0) + \sum_{t>0} D_x(t)u(x_t)) \geq \alpha(u(x_0) + \sum_{t>0} D_{\alpha x}(t)u(x_t)),
\]

which implies part (a).

For all \(x \not\in X_s\), if \(\alpha x \not\in X_s\), then (11) implies that \(D_x(t) = D_{\alpha x}(t)\). Part (b) holds by the same argument as above.

To show part (c), take any \(x \in X_s\) with \(u(x_0) = 0\). Then, \(\beta c_x \sim \alpha x\) if and only if

\[
\beta \sum_{t>0} D_{x_t}(t) \cdot u(x_t) = \sum_{t>0} D_{\alpha x_t}(t) \cdot u(\alpha x_t)
\]

\[
\iff \beta \sum_{t>0} \left(\frac{u(x_t)}{\text{ma}_t}\right)^{\frac{1}{m-1}} \cdot u(x_t) = \sum_{t>0} \left(\frac{u(\alpha x_t)}{\text{ma}_t}\right)^{\frac{1}{m-1}} \cdot u(\alpha x_t)
\]

\[
\iff \beta \sum_{t>0} \left(\frac{u(x_t)}{\text{ma}_t}\right)^{\frac{1}{m-1}} \cdot u(x_t) = \alpha \sum_{t>0} \left(\frac{u(x_t)}{\text{ma}_t}\right)^{\frac{1}{m-1}} \cdot u(x_t).
\]

40
Thus, $\beta = \alpha \frac{m}{t}$. Since $\beta$ is independent of $x$, we show part (c). ■

B Appendix: Proof of Theorem 1

Necessity is obvious. We show sufficiency. By Lemma 3 of Appendix D, there exists a continuous utility representation $U : X \to \mathbb{R}$ of $\succeq$ such that its restriction $u$ on $C$ is continuous, mixture linear, and homogeneous (that is, $u(\alpha c) = \alpha u(c)$ for all $\alpha \geq 0$), and satisfies $u(0) = 0$. Moreover, by $X$-Separability and a similar argument as in Lemma 4 without Small Domi-
nance, $U$ can be written as an additively separable utility form, that is, for all $x \in X$,

$$U(x) = u(x_0) + \sum_{t>0} U_t(x_t),$$

where $U_t : C \to \mathbb{R}$ is continuous with $U_t(0) = 0$ for each $t$. For all $x$, recall that $(x_t)^t$ denotes the dated reward that pays $x_t$ in period $t$ and pays 0 otherwise. Since $U(c^t) = U_t(c)$ for all dated rewards $c^t$,

$$U(x) = u(x_0) + \sum_{t>0} U((x_t)^t). \quad (16)$$

For all dated rewards $c^t$ with $c \neq 0$, define $D_c(t) := U_t(c)/u(c)$. Then, $U(c^t) = U_t(c) = D_c(t)u(c)$ for all $x$ and $\alpha \in [0,1]$, by Homotheticity, $\alpha c_x \sim \alpha x$. Since $u$ is homogeneous,

$$\alpha U(x) = \alpha u(c_x) = u(\alpha c_x) = U(\alpha x).$$

In particular, for dated rewards,

$$\alpha D_c(t)u(c) = \alpha U(c^t) = U(\alpha c^t) = D_{\alpha c}(t)u(\alpha c).$$

By linearity of $u$, we have $D_c(t) = D_{\alpha c}(t)$ for all $\alpha$. That is, $D_c(t)$ is independent of $c$ and can be written as $D(t)$. From (16), for all $x$,

$$U(x) = u(x_0) + \sum_{t>0} U((x_t)^t) = u(x_0) + \sum_{t>0} D(t)u(x_t).$$

Finally, Impatience implies that $1 \geq D(t) \geq D(t + 1)$ as desired.
First, we show sufficiency. Since $\succsim$ satisfies $X$-Separability, we can use a similar argument as in Lemma 4 without Small Dominance, and show that $\succsim$ is represented by

$$U(x) = u(x_0) + \sum_{t>0} D_{x_t}(t)u(x_t),$$

where $u$ is continuous, mixture linear, and homogeneous, and satisfies $u(0) = 0$. Moreover, by Impatience, $1 \geq D_c(t) \geq D_c(t + 1)$ for all $c$.

By the same argument as in Lemma 7 of Appendix D, $D_{x_t}(t)$ can be written as $D_{u(x_t)}(t)$ and $D_{u(c)}(t)$ is continuous and strictly increasing in $u(c)$. By the same argument as in Lemma 8, we can construct an increasing convex cost function $\varphi_t : [0, 1] \to \mathbb{R}_+ \cup \{\infty\}$ from the FOC. It is shown that $\varphi_t(d) = 0$ for all $d \in [0, \underline{d}_t)$, $\varphi_t(d)$ is differentiable, strictly increasing, and strictly convex on $d \in (\underline{d}_t, \overline{d}_t]$, and $\varphi_t(d) = \infty$ for all $d \in (\overline{d}_t, 1]$. By construction of $\varphi_t$, for all $x$, we have

$$u(x_t) = \varphi_t'(D_{u(x_t)}(t)),$$

that is,

$$D_x = \arg \max_{D \in [0, 1]^T} \left\{ \sum_{t>0} (D(t) \cdot u(x_t) - \varphi_t(D(t))) \right\}. \quad (17)$$

By Weak Homotheticity (c), by adopting the same argument as in Lemma 20, for all dated reward $c^t$ in the effective domain, $D_{\alpha c^t} = \alpha D_{c^t}$ for all $\alpha \in (0, 1)$. Thus, $D_{\alpha c^t} \to 0$ as $\alpha \to 0$, which implies $\underline{d}_c = 0$. Moreover, as shown in Lemmas 21 and 22, $\varphi_t(d)$ can be written as $a_t d^m$ for some $a_t > 0$ and $m > 1$.

We turn to necessity. By the same argument as in Appendix A.1, the reduced form of the unconstrained CE representation is given as

$$U(x) = u(x_0) + \sum_{t>0} D_{x_t}(t) \ast u(x_t),$$

where

$$D_{x_t}(t) = \begin{cases} \left( \frac{u(x_t)}{m a_t} \right)^{\frac{1}{m-1}} & \text{if } u(x_t) \leq \overline{d}_t^{m-1} m a_t, \\
\overline{d}_t & \text{if } u(x_t) > \overline{d}_t^{m-1} m a_t. \end{cases} \quad (18)$$

Since $D_{x_t}(t)$ is continuous and weakly increasing, the preference that $U$ represents satisfies Regularity. Since $D_{x_t}(t)$ depends only on $x_t$, $U$ is additively separable across $t$, which implies $X$-Separability.
Since $D_{x_1}(t)$ is weakly increasing, for all $x \in X_+$ and $\alpha \in (0, 1)$, $\alpha c_x \succeq \alpha x$ if and only if $\alpha U(x) \geq U(\alpha x)$, or

$$u(x_0) + \sum_{t>0} D_{x_1}(t) \cdot u(x_t) \geq u(x_0) + \sum_{t>0} D_{\alpha x_1}(t) \cdot u(x_t),$$

which implies part (a) of Weak Homotheticity. It is clear from (19) that $\alpha c_x \succ \alpha x$ if and only if there exists some $t \geq 1$ such that $u(x_t) > 0$ and $D_{x_1}(t) > D_{\alpha x_1}(t)$. Therefore,

$$X_s = \{x \in X_+ : \text{there exists some } t \geq 1 \text{ such that } 0 < u(x_t) \leq \overline{d}_t^{m-1} m a_t\}.$$

If $x \notin X_s$, for all $t \geq 1$, we have either $u(x_t) = 0$ or $u(x_t) > \overline{d}_t^{m-1} m a_t$, that is, either $u(x_t) = 0$ or $D_{x_1}(t) = \overline{d}_t$. Therefore, on the subdomain of $X \setminus X_s$, $U$ is identified with a DU model with a fixed discount function $(\overline{d}_t)_{t \geq 1}$. Consequently, if $\alpha x \notin X_s$, we have Weak Homotheticity (b), that is, $\alpha c_x \sim \alpha x$. Weak Homotheticity (c) is implied by the same argument as in Lemma 2.

It is easy to see that for all $x$, $\alpha x \in X_s$ for sufficiently small $\alpha > 0$, which implies part (a) of Absorbing. As shown above, $\alpha c_x \succ \alpha x$ if and only if $0 < u(x_t) \leq \overline{d}_t^{m-1} m a_t$ for some $t \geq 1$. Since $0 < u(\alpha x_t) \leq \overline{d}_t^{m-1} m a_t$ for all $\alpha \in (0, 1)$, we have $\alpha x \in X_s$, that is, part (b) of Absorbing holds.

\section{Appendix: The General CE Model}

As an intermediate result for the CE representation, we prove a sufficiency for the General CE representation as given by Definition 3.

Consider the following axioms.

**Axiom 13 (Time-Invariance)** For all $c, \tilde{c} \in C$ and $t$, if $c^t, \tilde{c}^t \in X_s$, then

$$c \succeq \tilde{c} \iff c^t \succeq \tilde{c}^t.$$ 

Time-Invariance requires that preferences over time $t$ instantaneous consumption are independent of $t$. This is implied by Monotonicity.

**Axiom 14 (Small Dominance)** For any positive stream $x$ and any $S \subset \{1, \cdots, T\}$ such that $x_t > 0$ for some $t \in S$,

$$x \in X_s \implies x S 0 \in X_s.$$
This says that if there is a small stream \( x \) paying positive outcomes at some periods within \( S \), then the stream that is identical on \( S \) and paying nothing elsewhere is also small. This is implied by Restricted Boundary and Monotonicity.

The Regularity axiom is weakened as follows:

**Axiom 15 (Weak Regularity)** \( \succsim \) satisfies Order, Continuity, Base-line, Impatience, Present Equivalents, and Risk Preference.

In this Appendix, we show the following:

**Theorem 7** If a preference \( \succsim \) on \( X \) satisfies Non-Triviality, Weak Regularity, \( X^*_s \)-Separability, Weak Homotheticity (a), Absorbing, Time-Invariance, Small Dominance, Symmetry, and Large Homotheticity, then it admits a General CE representation as given by Definition 3.

Note that the above theorem does not require part (b) and part (c) of Weak Homotheticity.

A necessary and sufficient condition for the General CE representation is provided by the supplementary appendix (Noor and Takeoka [28]).

**D.1 Representation on \( X^*_s \)**

**Lemma 3** The preference \( \succsim |_C \) is represented by a utility function \( u : C \to \mathbb{R} \) with \( u(0) = 0 \) which is continuous, mixture linear, and homogeneous (that is, \( u(\alpha c) = \alpha u(c) \) for all \( \alpha \geq 0 \)). Moreover, the preference \( \succsim \) on \( X \) is represented by a continuous utility function \( U : X \to \mathbb{R} \) such that \( U(c) = u(c) \) for all \( c \in C \).

**Proof.** By Weak Regularity, \( \succsim |_C \) satisfies the vNM axioms. There exists a continuous mixture linear function \( u : C \to \mathbb{R} \) which represents \( \succsim |_C \) and which can be chosen so that \( u(0) = 0 \).

Establish homogeneity of \( u \) next. If \( \alpha \in [0, 1] \), by mixture linearity of \( u \), together with identifying \( \alpha c \) with \( \alpha c + (1 - \alpha)0 \),

\[
u(\alpha c) = u(\alpha c + (1 - \alpha)0) = \alpha u(c) + (1 - \alpha)u(0) = \alpha u(c).
\]

If \( \alpha > 1 \), we identify \( \alpha c \) with \( c' \in C \) satisfying \( c = \frac{1}{\alpha}c' + \frac{\alpha - 1}{\alpha}0 \). Then, mixture linearity of \( u \) implies that \( u(c) = \frac{1}{\alpha}u(c') \), that is, \( u(\alpha c) = u(c') = \alpha u(c) \), as desired.
For any \( x \in X \), the Present Equivalents axiom ensures that there exists \( c_x \in C \) such that \( c_x \sim x \). Define \( U(x) = u(c_x) \). By construction, \( U \) represents \( x \). Moreover, for all \( c \in C \), \( U(c) = u(c) \). In particular, we have \( U(0) = u(0) = 0 \).

To show the continuity of \( U \), take any sequence \( x^n \rightarrow \overline{x} \). There exists a corresponding present equivalent \( c_{x^n} \sim x^n \). We want to show that \( U(x^n) = u(c_{x^n}) \) converges to \( U(\overline{x}) = u(c_{\overline{x}}) \). First, assume that \( \overline{x}, x^n \gtrsim 0 \) for all \( n \). Fix \( c^* \in C \) with \( c^* > 0 \) arbitrarily. Since \( u \) is continuous and homogeneous, there exists a unique \( \lambda(x^n) \geq 0 \) such that \( u(c_{x^n}) = \lambda(x^n)u(c^*) = u(\lambda(x^n)c^*) \). For \( \overline{x} > \lambda(\overline{x}) \), \( \overline{x} \) belongs to the set \( W = \{ x \in X | \lambda c^* > x \gtrsim 0 \} \). By Continuity, we can assume \( x^n \in W \) for all \( n \) without loss of generality.

Since \( U(x^n) = \lambda(x^n)u(c^*) \) and \( U(\overline{x}) = \lambda(\overline{x})u(c^*) \), it is enough to show that \( \lambda(x^n) \rightarrow \lambda(\overline{x}) \). Seeking a contradiction, suppose otherwise. Then, there exists a neighborhood of \( \lambda(\overline{x}) \), denoted by \( B(\lambda(\overline{x})) \), such that \( \lambda(x^n) \notin B(\lambda(\overline{x})) \) for infinitely many \( m \). Let \( \{x^m\} \) denote the corresponding subsequence of \( \{x^n\} \). Since \( x^n \rightarrow \overline{x} \), \( \{x^m\} \) also converges to \( \overline{x} \). Since \( \{\lambda(x^m)\} \) is a sequence in \([0, \overline{x}]\), there exists a convergent subsequence \( \{\lambda(x^m)\} \) with a limit \( \overline{\lambda} \neq \lambda(\overline{x}) \). On the other hand, since \( x^\ell \rightarrow \overline{x} \) and \( x^\ell \sim \lambda(x^\ell)c^* \), Continuity implies \( \overline{x} \sim \overline{\lambda}c^* \). Since \( \lambda(\overline{x}) \) is unique, \( \lambda(\overline{x}) = \overline{\lambda} \), which is a contradiction.

The symmetric argument can be applied for the case that \( 0 \gtrsim \overline{x}, x^n \) for all \( n \). Finally, suppose that \( \overline{x} \sim 0 \). If \( x^n \sim 0 \) for some \( n \), \( U(x^n) = 0 = U(\overline{x}) \) for such \( n \). Thus, we can assume without loss of generality that \( x^n \not\sim 0 \) for all \( n \). For the subsequence \( \{x^n\} \) of \( \{x^n\} \) satisfying \( x^n \sim 0 \), we have \( x^n \rightarrow \overline{x} \). By the above argument, \( U(x^n) \rightarrow U(\overline{x}) \). Similarly, for the subsequence \( \{x^m\} \) of \( \{x^n\} \) satisfying \( 0 \gtrsim x^m \), we have \( U(x^m) \rightarrow U(\overline{x}) \). Therefore, \( U(x^n) \rightarrow U(\overline{x}) \), as desired.

Prior to the next lemma, we verify the following: For all \( x \) and all \( S \subset \{1, \cdots, T\} \) such that \( x_t \not\sim 0 \) for some \( t \in S \),

\[ x \in X^*_s \implies xS0 \in X^*_s. \tag{20} \]

Note that \( x \in X^*_s \) if and only if \( x^* \in X_s \). Since \( x^* \) is a positive stream, Small Dominance implies that \( x^*S0 \in X_s \). Thus, we have \( xS0 \in X^*_s \).

The property (20) is referred to as Small Dominance*.

**Lemma 4.** On the subdomain \( X^*_s \cup C \subset X \), \( U \) can be written as an additively separable utility form, i.e. \( U : X^*_s \cup C \rightarrow \mathbb{R} \) s.t. for all \( x \in X^*_s \cup C \),

\[ U(x) = u(x_0) + \sum_{t>0} U_t(x_t), \]

\[ 45 \]
where $u$ is given as in Lemma 3 and $U_t : C \to \mathbb{R}$ are continuous with $U_t(0) = 0$ for each $t$.

**Proof.** Take any $x \in X^*_t$, which is denoted by $x = (x_0, x_1, \cdots, x_T)$. There exists some $t > 0$ with $x_t \neq 0$. We start with the case where there are two $x_t, x_s \neq 0$. By notational convenience, denote such a stream by $(x_t, x_s, 0, \cdots, 0)$. By $X^*_s$-Separability,

$$1/2c_{(0, 0, \cdots, 0)} + 1/2c_{(x_t, 0, \cdots, 0)} \sim 1/2c_{(x_t, x_s, 0, \cdots, 0)} + 1/2.$$ 

Since $u$ is mixture linear,

$$u(c_{(0, 0, \cdots, 0)}) + u(c_{(x_t, 0, \cdots, 0)}) = u(c_{(x_t, x_s, 0, \cdots, 0)}) + u(0)$$

$$\iff U(0, x_s, x_t, 0, \cdots, 0) + U(x_t, x_s, 0, \cdots, 0) = U(x_t, x_s, 0, \cdots, 0).$$

Define $U_t(x_t) = U(x_t, 0, \cdots, 0)$ and $U_s(x_s) = U(0, x_s, 0, \cdots, 0)$. Then, we have

$$U(x_t, x_s, 0, \cdots, 0) = U_t(x_t) + U_s(x_s).$$

In particular, if $t = 0$, $U_s(x_s) = u(x_t)$.

If a stream has three outcomes $x_t, x_s, x_r \neq 0$, denote it by $(x_t, x_s, x_r, 0, \cdots, 0)$. By Small dominance*, $(x_t, x_s, 0, \cdots, 0)$ is small. From the above argument, we have (21). By $X^*_s$-Separability,

$$1/2c_{(0, 0, \cdots, 0)} + 1/2c_{(x_t, x_s, 0, \cdots, 0)} \sim 1/2c_{(x_t, x_s, x_r, 0, \cdots, 0)} + 1/2.$$ 

Since $u$ is mixture linear,

$$u(c_{(0, 0, \cdots, 0)}) + u(c_{(x_t, x_s, 0, \cdots, 0)}) = u(c_{(x_t, x_s, x_r, 0, \cdots, 0)}) + u(0)$$

$$\iff U(0, 0, x_r, 0, \cdots, 0) + U(x_t, x_s, 0, \cdots, 0) = U(x_t, x_s, x_r, 0, \cdots, 0).$$

Define $U_r(x_r) = U(0, 0, x_r, 0, \cdots, 0)$. Then, we have

$$U(x_t, x_s, x_r, 0, \cdots, 0) = U_r(x_r) + U(x_t, x_s, 0, \cdots, 0)$$

$$= U_t(x_t) + U_s(x_s) + U_r(x_r).$$

By repeating the same argument finitely many times, we have

$$U(x) = u(x_0) + \sum_{t>0} U_t(x_t).$$

46
where $U_t(x_t)$ is defined as $U_t(x_t) = U(0, \cdots, 0, x_t, 0, \cdots, 0)$. By definition, $U_t(0) = 0$. Since $U$ is continuous, $U_t$ is also continuous. ■

We show several properties of an absolute value of streams.

**Lemma 5**  
(1) For all negative outcomes $c \in C$, $u(c) = -u(c^*)$.

(2) For any dated reward $c^f$ with a negative outcome, $(c^f)^* \sim (c^*)^f$.

(3) For all $x \in X$, if $x^*$ is an absolute value of $x$, $\alpha x^*$ is an absolute value of $\alpha x$ for all $\alpha > 0$.

**Proof.** (1) If $0 \succ c$, by definition, its absolute value $c^* \in C$ satisfies

$$\frac{1}{2}c + \frac{1}{2}c^* \sim 0.$$ 

Since $u$ is mixture linear, $u(c) = -u(c^*)$.

(2) Since the dated reward $c^f$ with this negative outcome is a negative stream, by Symmetry, $(c^f)^* \sim c(c^f)^*$. Since $(c^f)^* = (c^*)^f$ by definition, we have a desired result.

(3) By part (1), for any negative outcome $c \in C$, its absolute value $c^* \in C$ satisfies $u(c) = -u(c^*)$. Since $u$ is homogeneous, for all negative outcomes $c \in C$ and $\alpha > 0$, we have

$$u(\alpha c) = \alpha u(c) = -\alpha u(c^*) = -u(\alpha c^*),$$

that is, $\alpha c^*$ is an absolute value of $\alpha c$. Thus, the claim holds by definition. ■

If the condition of Time-Invariance holds by replacing $X_s$ with $X_s^*$, it is referred to as Time-Invariance*.

**Lemma 6** $\succsim$ satisfies Time-Invariance*

**Proof.** By Time-Invariance, the statement holds for all positive outcomes. Thus, take any non-positive outcomes $c, \hat{c}$. If $c \succsim 0 \succsim \hat{c}$, $c^f$ is a positive stream and $\hat{c}^f$ is a negative stream. So, we have $c^f \succsim 0 \succsim \hat{c}^f$. Thus, for negative outcomes $c, \hat{c}$ with $c \succsim \hat{c}$, assume $c^f, \hat{c}^f \in X_s^*$. By definition of $X_s^*$, $(c^f)^* = (c^*)^f \in X_s$ and $(\hat{c}^f)^* = (\hat{c}^*)^f \in X_s$. Moreover, by Lemma 5 (1), $\hat{c}^* \succsim c^*$. Time-Invariance implies that $(\hat{c}^*)^f \succsim (c^*)^f$. By Lemma 5 (2), $(c^*)^f \succsim (c^f)^*$. Since $c^f$ and $\hat{c}^f$ are negative outcomes, again by Lemma 5 (1), $c^f \succsim \hat{c}^f$, or equivalently, $c^f \succsim \hat{c}^f$, as desired. ■
Lemma 7 The function $U : X_s \cup C \to \mathbb{R}$ defined as in Lemma 4 can be written as follows:

$$U(x) = u(x_0) + \sum_{t > 0} D_{|u(x_t)|}(t) \cdot u(x_t),$$

where for all $t > 0$, $D_{|u(c)|}(t) \in [0, 1]$ and $D_{|u(c)|}(t)$ is continuous and strictly increasing in $|u(c)|$.

Proof. Taking the additive representation from Lemma 4, by Time-Invariance*, we have that $U_t(x_t)$ can be written as an increasing transformation of $u(x_t)$.

So we can write $U_t(x_t)$ as $U_t(u(x_t))$. Define $D_x$ by $D_{u(x)}(t) = \frac{U_t(u(x_t))}{u(x_t)} > 0$ for any $x_t \in C$ with $x_t \not\approx 0$. Define $d_x = \inf \{ D_{u(c)}(t) : 0 \not\approx c \in C \}$. Then

$$U(x) = u(x_0) + \sum_{t > 0} D_{u(x_t)}(t) \cdot u(x_t), \text{ for all } x \in X_s \cup C.$$

By Lemma 5, the representation implies

$$D_{u(c)}(t) \cdot u(c) = U(c') = u(c) = -u((c_c)^*) = -u(c_c^*) = -U((c_c^*)^*) = -D_{u(c^*)}(t) \cdot u(c^*) = D_{|u(c)|}(t) \cdot u(c),$$

and hence, $D_{u(c)}(t) = D_{|u(c)|}(t)$, as desired.

To see that $D_{|u(c)|}(t)$ is strictly increasing in $|u(c)|$, it suffices to show by the above observation that $D_{u(c)}(t)$ is strictly increasing in $u(c) > 0$. Note that for any positive stream $x \in X_s$ and its present equivalent $c_x$, by definition of $X_s$, $\alpha U(c_x) > U(\alpha x)$ for all $\alpha \in (0, 1)$ and thus $\alpha U(x) > U(\alpha x)$.

Applying this more specifically to a dated reward $c$ with $u(c) > 0$ and exploiting mixture linearity of $u$, we obtain $\alpha D_{u(c)}(t) \cdot u(c) > D_{u(c)}(t) u(\alpha c) = \alpha D_{\alpha u(c)}(t) u(c)$ and thus

$$D_{u(c)}(t) > D_{\alpha u(c)}(t), \text{ for all } \alpha \in (0, 1),$$

as desired.

Since $u$ and $U_t$ are continuous, so is $D_{u(c)}(t)$ in $u(c)$ on the domain of $u(c) \neq 0$. Moreover, since $|u(c)|$ is continuous, so is $D_{|u(c)|}(t)$ for all $|u(c)| \neq 0$. Since $d_x$ is defined as $\inf \{ D_{u(c)}(t) : 0 \not\approx c \in C \}$ and $D_{u(c)}(t)$ is strictly increasing in $|u(c)|$, $D_{|u(c)|}(t)$ is indeed continuous for all $|u(c)|$.

By Impatience, for all positive $c$ and $t \geq 1$, $u(c) = U(c^t) \geq U(c^t) = D_{|u(c)|}(t) \cdot u(c)$, which implies $D_{|u(c)|}(t) \leq 1$. ■
Lemma 8  The function $U : X^*_s \cup C \to \mathbb{R}$ appeared in Lemma 7 can be written as follows:

$$U(x) = u(x_0) + \sum_{t>0} D_x(t) \cdot u(x_t),$$

s.t. $D_x = \arg \max_D \{ \sum (D(t) \cdot |u(x_t)| - \varphi_t(D(t))) \}$

where for each $t > 0$, $\varphi_t : [0, 1] \to \mathbb{R}_+ \cup \{\infty\}$ is an increasing convex function that is strictly increasing, strictly convex, and differentiable on $\{d \mid 0 < \varphi_t(d) < \infty\}$, and satisfies $\varphi_t(d) = 0$ and $\varphi_t'(d) = 0$. Moreover, $\varphi_t(d) \leq \varphi_{t+1}(d)$ for all $t < T$ and $d$.

Proof. By Small Dominance*, if $x \in X^*_s$, then $x0 \in X^*_s$, that is, $x^*0 \in X_s$. Thus, $\varphi_t$ can be derived from the positive dated rewards at $t$ as follows. Define

$$S_t = \{d \in [0, 1] \mid d = D_{[u(c)]}(t) \text{ for some } c' \in X_s\}.$$ 

By Absorbing (b), if $c' \in X_s$, then $\alpha c' \in X_s$ for all $\alpha \in (0, 1)$. Thus, $S_t$ is an interval. Note $d_t = \inf S_t$. Denote $\bar{d}_t = \sup S_t$. Define $I_t = S_t \cup \{\bar{d}_t, d_t\}$. The cost function $\varphi_t$ on $I_t$ is implicitly defined by the first order condition

$$|u(c)| = \varphi_t'(D_{[u(c)]}(t)), \tag{22}$$

along with the assumption that $\varphi_t(d_t) = 0$. Moreover, the continuity of $D_{[u(c)]}(t)$ wrt $|u(c)|$ requires that $0 = \varphi_t'(d_t)$. The function is by construction once differentiable and has a positive slope. Since $D_{[u(c)]}(t)$ is strictly increasing in $|u(c)|$, (22) implies that $\varphi_t'$ is strictly increasing, and hence, $\varphi_t$ is strictly convex.

By construction, the set $\arg \max_D \{ \sum (D(t) \cdot |u(x_t)| - \varphi_t(D(t))) \}$ is nonempty and moreover, it is a singleton since $\sum (D(t) \cdot |u(x_t)| - \varphi_t(D(t)))$ is a strictly concave function of $D$. Thus $D_x$ is a unique solution.

The cost function can be extended to $[0, 1]$ by

$$\varphi_t(d) = \begin{cases} 
0 & \text{if } d \in [0, d_t) \\
\varphi_t(d) & \text{if } d \in I_t \\
\infty & \text{if } d \in (\bar{d}_t, 1]
\end{cases}.$$ 

Then, $\varphi$ is increasing and convex on $[0, 1]$.

By Impatience, for all positive $c$ and for all $t < T$, $D_{[u(c)]}(t) \cdot u(c) = U(c') \geq U(c'+1) = D_{[u(c)]}(t+1) \cdot u(c)$. Thus, $D_{[u(c)]}(t)$ is weakly decreasing wrt $t$. This
observation implies that the effective domain \( \text{eff}(\varphi_t) \) of \( \varphi_t \) includes that of \( \varphi_{t+1} \). For any \( d := D_{u(c)}(t + 1) \) in the effective domain of \( \varphi_{t+1} \), it follows from the F.O.C. that

\[
\varphi_t'(d) \leq \varphi_t'(D_{u(c)}(t)) = u(c) = \varphi_{t+1}'(D_{u(c)}(t + 1)) = \varphi_{t+1}'(d),
\]

that is, \( \varphi_t'(d) \leq \varphi_{t+1}'(d) \) for all \( d \in \text{eff}(\varphi_{t+1}) \). By integrating both functions we obtain \( \varphi_t(d) \leq \varphi_{t+1}(d) \) for all \( d \in \text{eff}(\varphi_{t+1}) \). Consequently, \( \varphi_t(d) \leq \varphi_{t+1}(d) \) for all \( d \in [0, 1] \).

D.2 Extending the Representation to \( X_+ \)

Recall that \( X_+ \) is the set of positive streams. We extend the representation on \( X_+ \). The first lemma states that \( X_+ \) has a boundary point on a ray.

**Lemma 9** For any positive \( x \not\in C \), there exists a unique \( \alpha_x \in (0, 1] \) such that

\[
\left\{ \begin{array}{c}
\alpha < \alpha_x \implies \alpha c_x > \alpha x, \\
\alpha \geq \alpha_x \implies \alpha c_x \sim \alpha x.
\end{array} \right.
\]

**Proof.** Step 1: For all positive \( x \not\in C \), \( \alpha c_x > \alpha x \) implies \( \beta c_x > \beta x \) for all \( \beta \in (0, \alpha] \). By definition, a present equivalent of \( \alpha x \), denoted by \( c_{\alpha x} \), satisfies \( \alpha c_x \sim \alpha x \). For any \( \gamma \in (0, 1) \), let \( \beta = \gamma \alpha \in (0, \alpha) \). By Weak Homotheticity (a) and Risk Preference,

\[
\beta c_x = \gamma \alpha c_x \succ \gamma \alpha x \succsim \gamma \alpha x = \beta x,
\]

as desired.

Step 2: The result. If \( x \in X_+ \), \( \alpha_x = 1 \) satisfies this condition. Thus, assume \( x \not\in X_+ \). Let \( A = \{ \alpha \in (0, 1] \mid \alpha c_x \succ \alpha x \} \). By Absorbing (a), \( A \) is non-empty. Moreover, by Step 1, \( A \) is an interval with \( \inf A = 0 \). Let \( \alpha_x \) be a supremum of \( A \). If \( A = (0, 1) \), \( \alpha_x = 1 \) and this \( \alpha_x \) satisfies the desired property. If \( A \) is a proper subset of \( (0, 1) \), \( \alpha_x < 1 \). Then, there exists a sequence \( \alpha^n \rightarrow \alpha_x \) with \( \alpha^n > \alpha_x \). Since \( \alpha^n c_x \sim \alpha^n x \), by Continuity, \( \alpha_x c_x \sim \alpha_x x \), as desired.

**Lemma 10** For any positive \( x \not\in C \), take \( \alpha_x \in (0, 1] \) which is defined as in Lemma 9. Then,

\[
\left\{ \begin{array}{c}
\alpha \leq \alpha_x \implies \alpha x \in X_+, \\
\alpha > \alpha_x \implies \alpha x \not\in X_+.
\end{array} \right.
\]
Proof. If $\alpha < \alpha_x$, $\alpha c_x \succ \alpha x$. By Absorbing (b), for all $\beta \in (0, 1)$, $\beta c_{\alpha x} \succ \beta \alpha x$. Thus, $\alpha x \in X_s$. If $\alpha \geq \alpha_x$, by Lemma 9, $\alpha c_x \sim \alpha x$, that is, $\alpha c_x$ is a present equivalent of $\alpha x$. If $\alpha = \alpha_x$ in particular, for all $\beta \in (0, 1)$, Lemma 9 implies that $\beta \alpha_x c_x \succ \beta \alpha_x x$ because $\beta \alpha_x x < \alpha_x x$. Thus, $\alpha x \in X_s$. If $\alpha > \alpha_x$, for all $\beta$ sufficiently close to one, we have $\beta \alpha > \alpha_x$, and hence $\beta \alpha c_x \sim \beta \alpha x$. Thus, $\alpha x \notin X_s$. ■

Lemma 11 For all positive streams $x, y \notin C$, take $\alpha_x, \alpha_y \in (0, 1]$ which are defined as in Lemma 9. If $x_t \sim y_t$ for all $t$, then $\alpha_x = \alpha_y$.

Proof. By Lemma 8, the representation depends only on utility streams $(u(x_t))_{t=0}^\infty$. Since $u(x_t) = u(y_t)$ for all $t$, $x$ is small if and only if $y$ is small. In this case, $\alpha_x = \alpha_y = 1$. Assume that $x$ and $y$ are large streams. Seeking a contradiction, suppose that $\alpha_x \neq \alpha_y$. Without loss of generality, let $\alpha_x > \alpha_y$. For any $\alpha \in (\alpha_y, \alpha_x)$, by Lemma 10, $\alpha x$ is small and $\alpha y$ is large. Since $u(\alpha x_t) = u(\alpha y_t)$ for all $t$, this contradicts to the above argument. Thus, $\alpha_x = \alpha_y$, as desired. ■

By Lemma 10, for any $x \in X_+ \setminus C$, note that

$$\alpha_x = \sup_{\alpha \leq 1} \{ \alpha x \in X_s \}.$$ 

Moreover, $\alpha_x x \in X_s$.

Lemma 12 The function $U : X_+ \to \mathbb{R}$ appeared in Lemma 3 can be written as

$$U(x) = u(x_0) + \sum_{t>0} D_x(t) \cdot u(x_t),$$

s.t.

$$D_x = \begin{cases} 
\arg \max_D \{ \sum D(t) \cdot |u(x_t)| - \varphi_t(D(t)) \} & \text{if } x \in X_s \cup C, \\
D_{\alpha_x x} & \text{if } x \notin X_s \cup C.
\end{cases}$$

Proof. By Lemma 8, $U$ has the desired form on $X_s \cup C$. Consider the case of $x \notin X_s \cup C$. Lemma 9 implies $\alpha_x c_x \sim \alpha_x x$. Note that $U : X \to \mathbb{R}$ represents $\succ$ by the utility $u(c_x)$ of the present equivalent $c_x$ of each $x \in X$. Since $u(\alpha_x c_x) = U(\alpha_x x)$,

$$U(x) = u(c_x) = \frac{1}{\alpha_x} U(\alpha_x x).$$

51
Since $\alpha x \in X_s$, Lemma 8 implies
\[
U(x) = \frac{1}{\alpha_x} U(\alpha_x x) = \frac{1}{\alpha_x} \left( u(\alpha_x x_0) + \sum_{t > 0} D_{\alpha_x x}(t) \cdot u(\alpha_x x_t) \right) \\
= u(x_0) + \sum_{t > 0} D_{\alpha_x x}(t) \cdot u(x_t),
\]
as desired. ■

**Lemma 13** There is a function $K : X_+ \setminus C \to \mathbb{R}_+ \cup \{\infty\}$ such that $\succeq$ is represented by
\[
U(x) = u(x_0) + \sum_{t > 0} D_x(t) \cdot u(x_t),
\]
s.t. \[D_x = \text{arg max}_{D \in \Lambda_x} \{ \sum_{t > 0} D(t) \cdot |u(x_t)| - \varphi_t(D(t)) \} \]
\[
\Lambda_x := \{ D \in [0,1]^T \mid \varphi(D) := \sum_{t > 0} \varphi_t(D(t)) \leq K_x \}.
\]
Moreover, (1) the function $K_x$ satisfies $K_x = K_{\lambda x}$ for any $x$ and $\lambda$, and (2) for all positive streams $x, y$, if $u(x_t) = u(y_t)$ for all $t$, then $K_x = K_y$.

**Proof.** Since $U(c) = u(c)$ for all $c \in C$, $K$ does not play any role for consumption stream on $C$. Take any $x \in X_+ \setminus C$. If $\lambda x \in X_s$ for all $\lambda > 0$, define $K_x = K_{\lambda x} = \infty$ for all $\lambda > 0$. Otherwise, we can find another $x$ on the same ray with $x \notin X_s$. For such $x$, define
\[
K_x := \varphi(D_{\alpha_x x}).
\]
Extend to $X_s$ by requiring $K_x = K_{\lambda x}$ for any $\lambda > 0$.

For all $x \in X_+ \setminus C$, by Lemma 10, there exists $\alpha_x > 0$ such that $\alpha_x x \in X_s$. For any $\beta \in (0, \alpha_x)$, since $\varphi$ is strictly increasing and $D_{u(c)}(t)$ is strictly increasing in $u(c)$, $K_x = \varphi(D_{\alpha_x x}) = \varphi(D_{\alpha_x x}) > \varphi(D_{\beta x}) \geq 0$. Hence, $K_x > 0$.

For any $x \in X_+ \setminus C$, define
\[
\Lambda_x := \{ D \in [0,1]^T \mid \varphi(D) \leq K_x \}.
\]
From Lemma 12, for any $x \in X_s$ we have
\[
U(x) = u(x_0) + \sum_{t > 0} D_x(t) \cdot u(x_t),
\]
52
s.t. \[ D_x = \arg\max_D \{ \sum D(t) \cdot |u(x_t)| - \varphi_t(D(t)) \}. \]

There exists \( x' \notin X_s \cup C \) such that \( x = \alpha x' \) for some \( \alpha \in (0, 1) \). Since \( \varphi \) is strictly increasing and \( D_{\alpha x'} \) is increasing in \( \alpha \) up to \( \alpha x' \), \( \varphi(D_x) \leq \varphi(D_{\alpha x'}) = K_x \), that is, we have \( D_x \in \Lambda_x \). Thus, \( D_x \) is also the unique maximizer in the constrained problem:

\[ D_x = \arg\max_{D \in \Lambda_x} \{ \sum D(t) \cdot |u(x_t)| - \varphi_t(D(t)) \}, \]

thereby establishing the result for \( x \in X_s \).

Next consider \( x \notin X_s \cup C \), and take \( \alpha x \in X_s \). By definition, note that \( K_x < \infty \). By the preceding,

\[ D_{\alpha x} = \arg\max_{D \in \Lambda_x} \{ \sum D(t) \cdot |u(\alpha x_t)| - \varphi_t(D(t)) \}. \]

We first prove that

\[ D_{\alpha x} \in \arg\max_{D \in \Lambda_x} D \cdot |u(x)|. \tag{23} \]

To see this, suppose by way of contradiction that there is \( D \in \Lambda_x \) s.t. \( D \cdot |u(x)| > D_{\alpha x} \cdot |u(x)| \). Since \( D_{\alpha x} \) is on the boundary of \( \Lambda_x \) and \( D \in \Lambda_x \), we have \( \varphi(D_{\alpha x}) = K_x \geq \varphi(D) \). But these inequalities imply that

\[ D \cdot |u(\alpha x)| - \varphi(D) > D_{\alpha x} \cdot |u(\alpha x)| - \varphi(D_{\alpha x}) \]

contradicting the optimality of \( D_{\alpha x} \) for \( \alpha x \), as desired.

To conclude the proof of the lemma, observe that for any \( D \in \Lambda_x \) with \( D \neq D_{\alpha x} \),

\[ D_{\alpha x} \cdot |u(\alpha x)| - \varphi(D_{\alpha x}) > D \cdot |u(\alpha x)| - \varphi(D) \]

\[ \Rightarrow D_{\alpha x} \cdot |u(\alpha x)| - D \cdot |u(\alpha x)| > \varphi(D_{\alpha x}) - \varphi(D) \]

\[ \Rightarrow \alpha x \cdot D_{\alpha x} \cdot |u(x)| - D \cdot |u(x)| > \varphi(D_{\alpha x}) - \varphi(D) \]

\[ \Rightarrow D_{\alpha x} \cdot |u(x)| - D \cdot |u(x)| > \varphi(D_{\alpha x}) - \varphi(D) \]

(since \( D_{\alpha x} \cdot |u(x)| \geq D \cdot |u(x)| \), by (23))

\[ \Rightarrow D_{\alpha x} \cdot |u(x)| - \varphi(D_{\alpha x}) > D \cdot |u(x)| - \varphi(D). \]

Thus,

\[ D_{\alpha x} = \arg\max_{D \in \Lambda_x} \{ \sum D(t) \cdot |u(x_t)| - \varphi_t(D(t)) \}, \]

as desired.

By Lemma 11, if \( u(x_t) = u(y_t) \) for all \( t \), \( \alpha_x = \alpha_y \). Thus \( K_x \) is finite if and only if \( K_y \) is finite. If \( K_x \) is finite, it is obvious from the definition that \( K_x \) depends only on the utility stream \( (u(x_t))_{t=1}^T \). Thus, we have \( K_x = K_y \).
D.3 Extending the Representation to $X$

Finally, we extend the representation on $X^*_s \cup C$ to all of $X$.

**Lemma 14** For any $x \notin C$ and any its absolute value $x^*$, there exists a unique $\alpha_{x^*} \in (0, 1]$ such that

\[
\begin{align*}
\alpha < \alpha_{x^*} & \implies \alpha c_{x^*} \succ \alpha x^*, \\
\alpha \geq \alpha_{x^*} & \implies \alpha c_{x^*} \sim \alpha x^*.
\end{align*}
\]

**Proof.** Since $x^*$ is a positive stream, such an $x^*$ exists by Lemma 9. ■

**Lemma 15** Take any $x \notin C$ and any its absolute values $x^*$ and $x^{**}$. Then,

(1) $\alpha_{x^*} = \alpha_{x^{**}}$, where $\alpha_{x^*}$ and $\alpha_{x^{**}}$ are defined as in Lemma 14.

(2) If a positive stream $y$ satisfies $x^*_t \sim y_t$ for all $t$, $\alpha_{x^*} = \alpha_y$.

**Proof.** (1) By definition of absolute values, $x^*_t \sim x^{**}_t$ for all $t$. By Lemma 11, $\alpha_{x^*} = \alpha_{x^{**}}$.

(2) Immediate from Lemma 11. ■

For all $x \notin C$, by Lemma 14, define

\[\alpha_x := \alpha_{x^*}.\]

By Lemma 15, $\alpha_x$ is well-defined.

**Lemma 16** For any $x \notin C$,

\[
\begin{align*}
\alpha \leq \alpha_x & \implies \alpha x \in X^*_s, \\
\alpha > \alpha_x & \implies \alpha x \notin X^*_s.
\end{align*}
\]

**Proof.** The same proof as in Lemma 10 goes through by replacing $X_s$ with $X^*_s$. ■

Note that by Lemma 16, for all $x \notin C$, $\alpha_x x \in X^*_s$.

**Lemma 17** The function $U : X \rightarrow \mathbb{R}$ appeared in Lemma 3 can be written as

\[U(x) = u(x_0) + \sum_{t>0} D_x(t) \cdot u(x_t),\]

s.t. \[D_x = \begin{cases} \arg \max_D \left\{ \sum D(t) \cdot |u(x_t)| - \varphi_t(D(t)) \right\} & \text{if } x \in X^*_s \cup C, \\
D_{\alpha_x} & \text{if } x \notin X^*_s \cup C. \end{cases}\]
Proof. By Lemma 8, \(U\) has the desired form on \(X^*_s \cup C\). Consider the case of \(x \notin X^*_s \cup C\).

We will claim that \(\alpha_x c_x \sim \alpha_x x\) for all \(x \notin X^*_s\). If \(\alpha_x = 1\), the claim holds by definition of \(c_x\). So, assume \(\alpha_x < 1\). If \(x\) is a positive stream, \(x^* = x\). Thus, Lemma 9 implies \(\alpha_x c_x \sim \alpha_x x\), as desired. Next assume that \(x\) is a negative stream. By Lemma 14, \(\alpha_x c_x^* \sim \alpha_x x^*\). By Symmetry, \(\alpha_x c_x^* \sim \alpha_x x^*\). By Lemma 5 (3), \((\alpha_x c_x)^* \sim (\alpha_x x)^*\). Since \(\alpha_x x \in X^*_s\), together with Lemma 5 (1),

\[
\begin{align*}
\begin{align*}
\text{Proof. By Lemma 8, } U \text{ has the desired form on } X^*_s \cup C. \text{ Consider the case of } x \notin X^*_s \cup C.
\quad & \quad \quad \text{We will claim that } \alpha_x c_x \sim \alpha_x x \text{ for all } x \notin X^*_s. \text{ If } \alpha_x = 1, \text{ the claim holds by definition of } c_x. \text{ So, assume } \alpha_x < 1. \text{ If } x \text{ is a positive stream, } x^* = x. \text{ Thus, Lemma 9 implies } \alpha_x c_x \sim \alpha_x x, \text{ as desired. Next assume that } x \text{ is a negative stream. By Lemma 14, } \alpha_x c_x^* \sim \alpha_x x^*. \text{ By Symmetry, } \alpha_x c_x^* \sim \alpha_x x^*. \text{ By Lemma 5 (3), } (\alpha_x c_x)^* \sim (\alpha_x x)^*. \text{ Since } \alpha_x x \in X^*_s, \text{ together with Lemma 5 (1),}
\end{align*}
\end{align*}
\]

\[
\begin{align*}
\begin{align*}
\begin{align*}
\text{Thus, } \alpha_x c_x \sim \alpha_x x, \text{ as desired. Finally assume that } x \text{ is a non-positive and non-negative stream. By Lemma 16, a sequence } \alpha^n \to \alpha_x \text{ with } \alpha^n > \alpha_x \text{ satisfies } \alpha^n x \notin X^*_s. \text{ By Large Homotheticity, } \alpha^n c_x \sim \alpha^n x. \text{ By Continuity, } \alpha_x c_x \sim \alpha_x x \text{ as } n \to \infty.
\end{align*}
\end{align*}
\end{align*}
\]

We have the desired result by the same argument as in Lemma 12. \(\blacksquare\)

**Lemma 18** There is a function \(K : X \setminus C \to \mathbb{R}_{++} \cup \{\infty\}\) such that \(\succeq\) is represented by

\[
U(x) = u(x_0) + \sum_{t>0} D_x(t) \cdot u(x_t),
\]

\[\text{s.t. } D_x = \arg \max_{D \in \Lambda_x} \left\{ \sum_{t>0} D(t) \cdot |u(x_t)| - \varphi_t(D(t)) \right\}\]

\[\Lambda_x := \{ D \in [0,1]^T \mid \varphi(D) := \sum_{t>0} \varphi_t(D(t)) \leq K_x \}.\]

Moreover, the function \(K_x\) satisfies \(K_x = K_{x\lambda}\) for any \(x\) and \(\lambda\).

**Proof.** By Lemma 13, there exists such a function \(K\) on \(X_+ \setminus C\). For any \(x \in X_+ \setminus C\), since its absolute value \(x^*\) is a positive stream, define

\[K_x := K_{x^*} > 0\]

for some absolute value \(x^*\) of \(x\). Note that for all absolute values \(x^*\) and \(x^{**}\) of \(x\), \(x^*_1 \sim x^{**}_1\). By the property of \(K\) stated in Lemma 13, \(K_{x^*} = K_{x^{**}}\). Thus,
$K_x$ is well-defined. Moreover, by definition, for all $\lambda > 0$, $K_{\lambda x} = K_{\lambda x^*} = K_{x^*} = K_x$.

Note that $\lambda x \in X_s^*$ for all $\lambda > 0$ if and only if $\lambda x^* \in X_s$ for all $\lambda > 0$. Thus, $K_x = K_{x^*} = \infty$. Otherwise, we can find another $x$ on the same ray with $x \notin X_s^*$. For such an $x$,

$$K_x = K_{x^*} = \varphi(D_{\alpha_x x^*}).$$

By Lemma 16, there exists $\alpha_x > 0$ such that $\alpha_x x \in X_s^*$. Moreover, by Lemma 15 (2), $\alpha_x = \alpha_x^*$. Since $D_{\alpha_x x}$ depends only on an absolute value of $\alpha_x x$, $\varphi(D_{\alpha_x x}) = \varphi(D_{\alpha_x x^*})$. That is, $K_x = \varphi(D_{\alpha_x x})$.

For any $x \in X \setminus C$, define

$$\Lambda_x := \{ D \in [0,1]^T | \varphi(D) \leq K_x \}.$$

By replacing $X_s$ with $X_s^*$, the subsequent argument is the same as in Lemma 13. 

E Appendix: Proof of Theorem 4

Necessity is proved in Appendix A.2. We show sufficiency.

As stated at the beginning of Appendix D, the set of axioms for the CE representation implies that of the General CE representation except for Absorbing.

**Lemma 19** $\succsim$ satisfies Absorbing.

**Proof.** Given Restricted Boundary and Continuity, for any $x \notin C$, $\beta \to 0$ implies $\beta x \succ c$ for some $\beta > 0$, and so $\beta x \in X_s$. For all $\gamma \in (0,1)$, $\gamma c_{\beta x} > \gamma \beta x$. Let $\alpha = \gamma \beta$. Since $\beta c_{\alpha x} \succ \beta x \sim c_{\beta x}$, Risk Preference implies that

$$\alpha c_x = \gamma \beta c_x \succ \gamma c_{\beta x} > \gamma \beta x = \alpha x.$$

That is, part (a) holds.

To show part (b), let $\alpha c_x \succ \alpha x$. First consider $x \in X_s$. Then, by Monotonicity, $\alpha x \succ x \succ c$, and hence, $\alpha x \in X_s$. Next assume $x \notin X_s$. Seeking a contradiction, suppose $\alpha x \notin X_s$. By Weak Homotheticity (b), $\alpha c_x \sim \alpha x$, which is a contradiction. 

56
Thus, by the result of Appendix D, $≿$ admits a General CE representation $(u, \{\varphi_t\}, K)$. We pay attention to positive streams $X_+$.

We first show that the cost function takes the power form for some constants $m > 1$ and $a_t > 0$,

$$\varphi_t(d) = a_t d^m. \quad (24)$$

**Lemma 20** For any $d$ in the effective domain of $\varphi_t^\prime$,

$$\varphi_t^\prime(\alpha r d) = \alpha \varphi_t^\prime(d).$$

**Proof.** Let $c_{ax}$ denote the present equivalent of dated reward $ax$. Define $\beta_\alpha^r$ by $\beta_\alpha^r ax \sim ax$. For any $x$, define $h_x(\alpha)$ by $D_{ax} = h_x(\alpha) D_x$. Thus $D_{ax} \cdot u(\alpha x) = \alpha h_x(\alpha) D_x \cdot u(x)$, i.e $U(\alpha x) = \alpha h_x(\alpha) U(x)$. It follows that $\beta_\alpha^r = h_x(\alpha)$. Weak Hometheticity (c) implies that $\beta_\alpha^r$ is independent of $x$, and thus we can write $h(\alpha)$, and in particular we have

$$D_{ax} = h(\alpha) D_x \text{ for any dated reward } x.$$

Given that $D_{ax}$ is continuous and increasing wrt $\alpha$, so is $h(\alpha)$.

From the above displayed equality, it follows that $h(\alpha \gamma) D_x = D_{ax \gamma} = h(\alpha) D_{\gamma x} = h(\alpha) h(\gamma) D_x$. Indeed, $h$ satisfies the multiplicative Cauchy equation,

$$h(\alpha \gamma) = h(\alpha) h(\gamma), \quad \alpha, \gamma \in [0,1].$$

The general continuous solution is $h(\alpha) = \alpha^r$ (apply Aczel to the transformation defined by $f(r) = h(\frac{1}{r})^{-1}$ for any $r \in \mathbb{R}_+$. Since $h$ is increasing, $r > 0$.

By these definitions and the FOC,

$$\varphi_t^\prime(\alpha^r D_x(t)) = \varphi_t^\prime(h(\alpha) D_x(t)) = \varphi_t^\prime(D_{ax}(t)) = \alpha u(x_t) = \alpha \varphi_t^\prime(D_x(t)).$$

Writing $d = D_x(t)$, this yields $\varphi_t^\prime(\alpha^r d) = \alpha \varphi_t^\prime(d)$, as desired. 

**Lemma 21** There exist $a_t > 0$ and $m_t > 1$ such that

$$\varphi_t(d) = a_t d^{m_t}.$$
Proof. Let $\tau$ be the largest dated reward at $t$ that will lie on the boundary of $X_s$ and $d_\tau$ be the associated discount factor. By Lemma 20,

$$\varphi_t'(\alpha^r d) = \alpha \varphi_t'(d)$$

$$\implies \varphi_t'(\alpha d) = \alpha^{\frac{1}{r}} \varphi_t'(d)$$

$$\implies \varphi_t'(d) = \varphi_t'\left(\frac{d}{d_\tau}\right) = \left(\frac{d}{d_\tau}\right)^{\frac{1}{r}} \varphi_t'(d_\tau) = d^{\frac{1}{r}} k$$

for some constant $k > 0$. Integrating $\varphi_t'$ wrt to $d$ and using the fact that $\varphi_t(0) = 0$, it follows that

$$\varphi_t(d) = a_t d^{m_t}$$

for some $m_t > 1$ and $a_t > 0$, as was to be shown. ■

Lemma 22 $m_t$ is independent of $t$.

Proof. The FOC yields that for any $c$, $u(c) = \varphi_t'(d_c) = m_t a_t d_c^{m_t-1}$ and thus $d_c = \left(\frac{u(c)}{m_t a_t}\right)^{\frac{1}{m_t-1}}$. In particular

$$\varphi_t(d_c) = a_t \left(\frac{u(c)}{m_t a_t}\right)^{\frac{m_t}{m_t-1}}. \quad (25)$$

We use this to show that for a dated reward $x$ that pays $c \neq 0$ at $t > 0$:

$$U(c^t) = d_c u(c) = \left(\frac{u(c)}{m_t a_t}\right)^{\frac{1}{m_t-1}} u(c) = \left(\frac{1}{m_t a_t}\right)^{\frac{1}{m_t-1}} u(c)^{\frac{1}{m_t-1} + 1}$$

$$= m_t a_t \left(\frac{1}{m_t a_t}\right)^{\frac{1}{m_t-1}} u(c)^{\frac{1}{m_t-1}} = m_t a_t \left(\frac{u(c)}{m_t a_t}\right)^{\frac{m_t}{m_t-1}}$$

That is,

$$U(c^t) = m_t \varphi_t(d_c). \quad (26)$$

Finally, to show that $m_t$ is independent of $t$, take for any $c \neq 0$, and for any $t > 0$. Consider any $\alpha$ and define $\beta$ by $\beta U(c^t) = U(\alpha c^t)$. By (25) and (26),

$$\beta U(c^t) = U(\alpha c^t) = m_t \varphi_t(d_{ac}) = \alpha^{m_t} \left[ m_t \varphi_t(d_c) \right] = \alpha^{m_t} U(c^t).$$
That is, 
\[ \beta = \alpha^{-\frac{m}{m-1}}. \]

However, Weak Homotheticity (c) requires that \( \beta \) depends only on \( \alpha \). Therefore, \( m_t \) must be invariant wrt \( t \), as desired. This completes the proof. ■

By Restricted Boundary and Monotonicity, for all \( x \), there exists a sufficiently large \( \alpha > 0 \) such that \( \alpha x \notin X_s \). Together with the construction of \( K \) in Lemma 13, this implies \( K_x < \infty \) for all \( x \).

Now, we show that \( K : X \setminus C \to \mathbb{R}_{++} \) is constant. Define
\[ bd(X_s) = \{ x \in X_s \mid \lambda x \notin X_s \text{ for all } \lambda > 1 \}. \]

Since \( K_x = K_{\lambda x} \) for all \( \lambda > 0 \), it suffices to show that \( K_x = K_y \) for all \( x, y \in bd(X_s) \). We prepare two lemmas as below.

**Lemma 23** For all \( x \in X_s \),
\[ \varphi(D_x) = \frac{1}{m}[U(x) - u(x_0)]. \]

**Proof.** By (24), \( \varphi_t(d) = a_t d^m \). For all \( x \in X_s \), the FOC implies \( ma_t(D_{x_t}(t))^{m-1} = u(x_t) \). Thus,
\[
\varphi(D_x) = \sum_{t>0} \varphi_t(D_{x_t}(t)) = \sum_{t>0} a_t \left( \frac{u(x_t)}{ma_t} \right)^{\frac{m}{m-1}} = \frac{1}{m} \sum_{t>0} \left( \frac{u(x_t)}{ma_t} \right)^{\frac{1}{m-1}} u(x_t)
\]
\[
= \frac{1}{m} \sum_{t>0} D_{x_t}(t)u(x_t) = \frac{1}{m}[U(x) - u(x_0)],
\]
as desired. ■

By Restricted Boundary, there exists \( \tau \in C \) with \( \tau \succ 0 \) such that
\[(x_0, x_1, \cdots, x_T) \in X_s \iff (0, x_1, \cdots, x_T) \prec \tau. \]

**Lemma 24** \( bd(X_s) = \{ x \in X \mid (0, x_1, \cdots, x_T) \sim \tau \} \).

**Proof.** Take any \( x \in bd(X_s) \). Since \( x \in X_s \), Restricted Boundary implies \( (0, x_1, \cdots, x_T) \prec \tau \). Seeking a contradiction, suppose \( (0, x_1, \cdots, x_T) \sim \tau \). By Continuity, there exists some \( \lambda > 1 \) such that \( (0, \lambda x_1, \cdots, \lambda x_T) \prec \tau \), which
implies \( \lambda x \in X_s \) by Restricted Boundary. However, by definition of \( bd(X_s) \), \( \lambda x \notin X_s \). This is a contradiction.

Conversely, take any \( x \in X \) satisfying \((0, x_1, \cdots, x_T) \sim \tau\). By Restricted Boundary, \( x \in X_s \). By seeking a contradiction, suppose \( x \notin bd(X_s) \). Then, there exists some \( \lambda > 1 \) with \( \lambda x \in X_s \). By Monotonicity, \((0, \lambda x_1, \cdots, \lambda x_T) \succ (0, x_1, \cdots, x_T) \sim \tau\). Restricted Boundary implies \( \lambda x \notin X_s \), a contradiction.

Lemma 25  For all \( x, y \in bd(X_s) \), \( K_x = K_y \).

**Proof.** By Lemmas 23 and 24,

\[
\begin{align*}
&x, y \in bd(X_s) \\
\implies & (0, x_1, \cdots, x_T) \sim (0, y_1, \cdots, y_T) \\
\implies & \sum_{t>0} D_{x_t}(t)u(x_t) = \sum_{t>0} D_{y_t}(t)u(y_t) \\
\implies & U(x) - u(x_0) = U(y) - u(y_0) \\
\implies & \varphi(D_x) = \varphi(D_y) \\
\implies & K_x = K_y.
\end{align*}
\]

The last implication comes from the definition of \( K_x \), which says that \( K_x = \varphi(D_x) \) for all \( x \in bd(X_s) \).\(^{26}\)

From now on, let \( K > 0 \) be the constant number implied by Lemma 25.

Lemma 26  For all \( t \geq 1 \), \( a_t \tilde{d}^m_t = K \).

**Proof.** Consider a dated reward \( c^t \in X_s \). By the FOC, \( u(c) = ma_t D_{c^t}(t)^{m-1} \). Since \( c^t \) is small,

\[
D_{c^t}(t) = \left( \frac{u(c)}{ma_t} \right)^{\frac{1}{m-1}} \leq \tilde{d}_t.
\]

Let \( \tilde{c}^t \) be a dated reward which attains a supremum of \( \{u(c)\mid c^t \in X_s\} \). Since the capacity constraint will be binding at \( \tilde{c}^t \),

\[
K = a_t \left( \frac{u(\tilde{c})}{ma_t} \right)^{\frac{1}{m-1}} = a_t \tilde{d}^m_t,
\]

as desired.\(^{26}\)

---

\(^{26}\)See the proof of Lemma 13.
F Appendix: Proof of Theorem 5

For any dated reward \( x = c^t \) with \( u(c) > 0 \), the discount function (which requires \( D_x(t) > 0 \) and \( D_x(\tau) = 0 \) for \( \tau \neq t \)) is determined by preference: if \( \gamma \in [0, 1] \) is such that \( \gamma c \sim x \), then \( D_x(t) = \gamma \). Thus the discount functions for dated rewards are uniquely pinned down by preference. Moreover, the set \( \{D_x(t) \in [0, 1] : c \succsim 0\} \) defines the effective domain of the cost function \( \varphi_t \) in any representation. We make use of these observations below.

Take two CE representations for the preference. Since \( u_1 \) and \( u_2 \) are linear and represent the same preference, there exists \( \alpha > 0 \) such that \( u_2 = \alpha u_1 \). Take a positive dated reward \( x = c^t \). By the first order condition,

\[ (\varphi^2_t)'(D_x(t)) = |u^2(c)| = \alpha |u^1(c)| = \alpha (\varphi^1_t)'(D_x(t)), \]

which implies \( \varphi^2_t = \alpha \varphi^1_t \). In particular, \( m_1 = m_2 \) and \( a_1^2 = \alpha a_1^1 \) for all \( t \).

From (12) in Appendix A.2, for all \( i = 1, 2 \),

\[ X^*_i = \{ x \in X \mid \sum_{t > 0} \gamma^i(t)|u^i(x_t)|^m = mK^i \}. \]

Since

\[ \gamma^2(t)|u^2(x_t)|^m = (ma^2_t)^{-1/m}|u^2(x_t)|^m = (m\alpha a^1_t)^{-1/m}|\alpha u^1(x_t)|^m \]

\[ = \alpha \gamma^1(t)|u^1(x_t)|^m, \]

we have \( K^2 = \alpha K^1 \), as desired.

G Proofs for Section 4.5

G.1 Proof of Proposition 2

If either (a) or (b) holds, we have \( \varphi_t(d) \to 0 \) for all \( d \). Indeed, if (a) holds, \( \varphi_t(d) \to 0 \) as \( a_t \to 0 \). If (b) holds, for all \( d \leq \tilde{d}_t < 1 \), \( \varphi_t(d) = a_t d^m \to 0 \) as \( m \to \infty \). In an unconstrained CE representation, an optimal discount function \( D_x(t) \) is obtained as (18) of Appendix C. Note that \( \tilde{d}^{m-1}_t ma_t \to 0 \) as \( a_t \to 0 \) or \( m \to 0 \) because by L'Hôpital’s rule,

\[ \lim_{m \to \infty} \tilde{d}^{m-1}_t m = \lim_{m \to \infty} \frac{m}{(\ln \tilde{d}_t)^m} = \lim_{m \to \infty} \ln \tilde{d}_t = 0. \]

Therefore, as \( a_t \to 0 \) or \( m \to \infty \), \( D_{x_t}(t) = \tilde{d}_t \) for all positive \( x \), as desired.
G.2 Proof of Proposition 3

Note that a cost function $\varphi_t(d)$ can be written as $\varphi_t(d) = K\left(\frac{d}{d_t}\right)^m$. On the other hand, the capacity constraint is equivalent to

$$\varphi(D) = \sum_{t>0} \left(D(t)\frac{1}{d_t}\right)^m \leq 1,$$

which is independent of $K$. For all discount functions $D$ with $D(t) < \bar{d}_t \leq 1$, $\left(D(t)\frac{1}{d_t}\right)^m \to 0$ as $m \to \infty$. Thus, all $D \in \text{eff}(\varphi) \setminus \{\bar{d}_t^T\}$ eventually satisfies the capacity constraint.

Note that from Proposition 1, $X_s$ approaches to $X$ and all positive streams become small streams as $m \to \infty$. Proposition 1 implies that for all positive streams $x$,

$$D_{u(x_t)}(t) = \left(\frac{\bar{d}_t^m u(x_t)}{mK}\right)^{\frac{1}{m-1}} = \left(\frac{u(x_t)}{K}\right)^{\frac{1}{m-1}} \bar{d}_t^{\frac{m}{m-1}} = \left(\frac{u(x_t)}{K}\right)^{\frac{1}{m-1}} m^{-\frac{1}{m-1}} \bar{d}_t^{\frac{m}{m-1}}.$$

Since $\frac{1}{m-1} \to 0$ and $\frac{m}{m-1} \to 1$, $\left(\frac{u(x_t)}{K}\right)^{\frac{1}{m-1}} \to 1$ and $\bar{d}_t^{\frac{m}{m-1}} \to \bar{d}_t$. Let $f(m) = m^{-\frac{1}{m-1}}$. By L’Hôpital’s rule,

$$\lim_{m \to \infty} \ln f(m) = \lim_{m \to \infty} -\frac{1}{\frac{1}{m}} = 0,$$

which implies $\lim_{m \to \infty} f(m) = 1$. Thus, $D_{u(x_t)}(t) \to \bar{d}_t$ as desired.

References


