Constrained Optimal Discounting*

Jawwad Noor Norio Takeoka

February 28, 2023

Abstract

Noor and Takeoka [19] model an agent whose impatience is determined by an (unconstrained) cognitive optimization problem. This paper presumes instead a limited stock of cognitive resources. The key behavioral implications are the absence of magnitude effects and violations of Separability for large rewards. The model endogenously produces lower impatience with age, providing a cognitive account for anomalous life-cycle saving behavior. The consumption-savings profile cannot be replicated by beta-delta discounting. In a task-completion setting, the model can exhibit cycles of activity and inactivity in completing tasks where standard models would predict the completion of all tasks in consecutive periods.

1 Introduction

Macroeconomic studies have documented a variety of anomalies for the Life-Cycle Consumption model. One such anomaly is that saving rates are "too low" among young cohorts and "too high" among older cohorts (Browning and Crossley [4]). A plausible explanation is that discount factors are age-dependent, and in particular increasing with age (Kureishi et al [13]). It may be natural to understand such age-dependence in terms of an exogenous evolution of preference stemming from maturity. However, it is possible to provide a more nuanced account that generates age-dependence endogenously. If agents require cognitive resources to generate patience (by generating empathy for future selves, or enhancing attention towards future consumption), and if they have limited cognitive resources, then they will tend to be impatient in their youth because they have to spread these resources

*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. Takeoka gratefully acknowledges financial support from JSPS KAKENHI Grant Number JP19KK0308. Part of this research was conducted while Norio Takeoka was visiting the Department of Economics, Boston University, whose hospitality is gratefully acknowledged. We thank Jiaqi Yang for expert research assistance. The usual disclaimer applies.

over a long horizon. As they age, the same amount of cognitive resources are expended across a shorter horizon, and therefore they are able to achieve greater patience with age.

More direct evidence of a relationship between cognitive abilities and time preference is provided by Dohmen et al [6], who show that people with higher cognitive abilities are less impatient. There is, however, a dearth of research seeking to understand the precise mechanism by which the two are related. Noor and Takeoka [19] (henceforth NT) hypothesize that time preferences are determined by a cognitive optimization problem which involves, for any given consumption stream, optimally spending cognitive resources on enhancing patience in order to better appreciate the stream – by engaging in the cognitively costly activity of developing empathy for future selves, she enhances the discount factor $D(t)$ used for different t. This can be viewed as a subjective version of Becker and Mulligan [3]'s model where the agent spends physical resources to enhance patience (such as by obtaining education). NT's conceptual contribution is to show that the main behavioral content of such a theory is a *magnitude effect*: the agent should be less impatient when dealing with larger rewards. The fact that the magnitude effect is a robust finding in the experimental literature (Fredrick et al [9]) gives credibility to NT's cognitive optimization hypothesis. In this paper, we continue the exploration of NT and study the behavioral meaning of *constrained* cognitive optimization. Cognitive costs capture one aspect of cognitive ability, and the size of cognitive resource capacity captures another, and we seek to understand how the latter manifests in behavior. NT show how their model unifies a range of Life Cycle anomalies, and our extension extends the range to include age-dependent impatience. In our concluding section, we discuss how this paper potentially speaks to a range of choice contexts outside intertemporal choice.

NT's primitive is a (static) preference $\succsim$ over the set X of finite horizon consumption streams, and they provide behavioral foundations for the *Costly Empathy (CE)* representation, which is described by an instantaneous consumption utility u and an increasing and convex cognitive cost function φ_t for each t such that $\succsim$ is represented by the function $U : X \to \mathbb{R}_+$ defined by

$$U(x) = u(x_0) + \sum_{t\geq 1} D_x(t)u(x_t), \qquad x \in X, \tag{1}$$

$$\text{where } D_x = \arg\max_{D\in[0,1]^T} \{\sum_{t\geq 1} D(t)u(x_t) - \sum_{t\geq 1} \varphi_t(D(t))\}. \tag{2}$$

Thus, the period 0 self evaluates a consumption stream $x = (x_0, \cdots, x_T)$ via the discounted utility formula (1) where the discount function D_x is the result of a cognitive optimization problem (2). Intuitively, the discount function D is the distribution of empathy across future selves, and it is chosen so as to balance discounted future utility $\sum_{t\geq 1} D(t)u(x_t)$ against the cognitive cost $\sum_{t\geq 1} \varphi_t(D(t))$ of generating empathy D. In the present paper, we are interested in the behavioral content of the *Constrained Costly Empathy (CCE)* representation which extends the CE model to allow for limited cognitive capacity, in which case the cognitive optimization problem (2) is subject to the capacity constraint:

$$\sum_{t\geq 1} \varphi_t(D(t)) \leq K,$$

for some $K \leq \infty$. That is, the agent cannot produce empathy D that costs more than K.

The main theoretical contribution of the paper lies in the study of the behavioral meaning of constrained cognitive optimization. We make two key observations:

1. The first relates to the magnitude effects. In NT, the magnitude effect arises because a larger reward incentivizes higher cognitive investment into reducing impatience. Since there is no constraint on cognitive capacity in their model, scaling up rewards should always give rise to a reduction in impatience, that is, a magnitude effect. However, if the agent has limited cognitive capacity, then at some point the capacity constraint will bind, and any further scaling up rewards would not lead to any magnitude effect.

2. The second relates to the violation of Separability. [1]To illustrate, imagine that an agent exhibits the patient preference $(0, 0, 300) \succ (100, 0, 0)$. Separability requires that she should continue to exhibit the same preference when, say, \$200 at time 1 is added to both streams: $(0, 200, 300) \succ (100, 200, 0)$. This is satisfied in NT's model because the unlimited stock of empathy allows $D(t)$ to be optimized separately across all t, that is, the \$200 reward does not affect how she allocates cognitive resources to any of the other rewards offered by the streams. However, when there are cognitive constraints, then the agent may become more impatient and exhibit $(0, 200, 300) \prec (100, 200, 0)$, thereby violating Separability: when evaluating the stream $(0, 200, 300)$, some of the limited cognitive resources may be taken away from \$300 and spent on \$200, making the highest reward look less attractive than it did before, while no such tradeoff occurs when evaluating $(100, 200, 0)$.

Putting these observations together, we identify two key behavioral predictions of constrained cognitive optimization: (a) "small" streams should exhibit a magnitude effect as well as satisfy Separability (as both arise when the cognitive constraint is slack), and (b) "large" streams should exhibit no magnitude effect, and should typically violate Separability. These predictions are the content of our "Preference-Based Threshold" and "Magnitude-Sensitive Separability" axioms and constitute our assessment of the main empirical content of constrained cognitive optimization.

The CCE model additionally places stringent requirements across streams that lie on the threshold where the cognitive capacity begins to bind. These restrictions are very technical and devoid of economic meaning, implying that a meaningful axiomatization of the CCE model may in fact be elusive. Against this backdrop, we nevertheless provide two results. The first is to provide a transparent axiomatization of a special case of the model where the cost function is of the homogeneous form $\varphi_t(d) = a_t d^m$. An interesting behavioral property of this special case is that the threshold where the capacity constraint binds has a very simple characterization: it can be described using a particular indifference curve. This

[1]The Separability axiom states that if two streams contain a common outcome x_t at time t, then the ranking of the streams do not change if this common outcome is replaced with any other y_t. Separability of preference is necessary for the existence of an additively separable representation for preference.

also makes the model analytically tractable since a single parameter determines whether the agent's cognitive capacity is binding at a given stream. Our second result drops the homogeneity property of the cost function, yielding a version of the CCE model where the cognitive constraint K_x can vary across streams in a limited way.

We provide several simple applications of the CCE model. In a dynamic consumption-savings context, one might expect that higher cognitive capacity would lead to higher saving by all selves. Interestingly, this intuition is not precise: it relies crucially on the assumption of naivete. We demonstrate in a 3-period context that a sophisticated period 0 self may in fact begin to save less, since a higher cognitive capacity for all selves may exacerbate dynamic inconsistency. In another application, we show that the consumption-savings profile (as a function of the interest rate) of the (sophisticated) agent cannot be replicated by standard exponential discounting or by the beta-delta model. We also formalize the sense in which the dynamic model gives rise to decreasing impatience with respect to age, as in the opening paragraph above. Another set of applications study the task completion problem, uncovering patterns of behavior that are inconsistent with models that satisfy Separability.

There are several models in the literature that incorporate subjective optimization, such as those of optimal expectations (Brunnermeier and Parker [5]), optimal contemplation (Ergin and Sarver [8]) and optimal attention (Ellis [7], Gabaix [10]). To our knowledge, constrained subjective optimization is considered only in the literature on willpower, where the decision maker is assumed to resist to temptation within the constraint of limited willpower. Specifically, Ozdenoren, Salant, and Silverman [24] consider the cake-eating problem with a fixed initial stock of willpower, which is depleted over time with exercising self-control. In a discrete setting, Masatlioglu, Nakajima, and Ozdenoren [18] axiomatize a limited willpower model by using the pair of ex ante preference over menus and ex post choice from menus. These papers presume that self-control is not costly. Liang, Grant and Hsieh [15] consider a menus of lotteries setting and incorporate the cost of self-control. They show that the content of limited will-power in that setting lies in the violation of the vNM Independence axiom for menus of lotteries. The main take-away of the present paper is that, in any context, binding cognitive constraints express themselves behaviorally in the absence of a magnitude effect for large rewards. Moreover, in contexts where separability may hold (such as intertemporal choice), there may be violations of separability for large rewards. Of the two, the former is fundamental – it is an extension of NT's strategy to identify cognitive optimization by viewing how behavior changes with the size of the rewards.

The remainder of the paper proceeds as follows. Section 2 describes our basic framework and derives some key properties of the CCE representation. Section 3 formulates behavioral counterparts of these properties and show that together they characterize a version of the CCE representation (which we refer to as the CCE* representation). Section 4 shows that adding a homogeneity restriction on preferences characterizes the CCE representation with homogeneous cost functions. Section 5 provides applications of the model. All proofs are relegated to the appendices and the supplementary appendix (Noor and Takeoka [21]).

2 CCE Framework

2.1 Primitives

There are $T+1<\infty$ periods, starting with period 0. The space C of outcomes is assumed to be $C=\mathbb{R}_+$. Let Δ denote the set of simple lotteries over C, with generic elements $p,q,\cdots$. We will refer to p as consumption. Consider the space of consumption streams $X=\Delta^{T+1}$, endowed with the product topology. A typical element in X is denoted by $x=(x_0,x_1,\cdots,x_T)$. The primitive of our model is a preference $\succsim$ over X.

Let $\Delta_0\subset X$ denote the set of streams $x=(p,0,\cdots,0)$ that offer consumption p immediately and 0 in every subsequent period. Abusing notation, we often use p to denote both a lottery $p\in\Delta$ and a stream $(p,0,\cdots,0)\in\Delta_0$. Thus, 0 also denotes the stream $(0,\cdots,0)$. An element of Δ that is a mixture between two consumption alternatives $p,q\in\Delta$ is denoted $\alpha\circ p+(1-\alpha)\circ q$ for any $\alpha\in[0,1]$. Streams are mixed pointwise: $\alpha\circ x+(1-\alpha)\circ y=(\alpha_0\circ x+(1-\alpha)\circ y_0,\cdots,\alpha\circ x_T+(1-\alpha)\circ y_T)$.

As a benchmark, we define the *Discounted Utility (DU)* representation for a preference over X by

$$U(x)=u(x_0)+\sum_{t>0}D(t)u(x_t),\qquad x\in X,$$

where $D(t)$ is weakly decreasing in t, and u is a utility index. A defining feature of the DU model is that the discount function evaluates time independently of the stream of rewards being evaluated. The CE and CCE models relax such magnitude-independent discounting.

2.2 Representation

Say that a tuple $(u,\{\varphi_t\}_{t\geq 1})$ is *regular* if

(i) $u:\Delta\to\mathbb{R}_+$ is continuous and mixture linear with (a) $u(0)=0$, (b) u is strictly increasing on C and (c) $u(\Delta)=\mathbb{R}_+$,

(ii) $\varphi_t:[0,1]\to\mathbb{R}_+$ is an increasing convex function that is strictly increasing, strictly convex and continuously differentiable on $\{d:\varphi_t(d)>0\}$, and satisfies $\varphi_t(0)=0$, $\varphi_t'(0)=0$ and $\varphi_t'(1)<\infty$,

(iii) For all $t<T$, the cost functions satisfy $\varphi_t'\leq\varphi_{t+1}'$.

Condition (i) imposes familiar properties on the utility from consumption. Condition (i)(a), while natural, is more than a mere normalization of u since the sign of u matters for the cognitive optimization problem defined below. The unboundedness of u imposed by condition (i)(c) ensures the existence of a present equivalent of each stream $x\in X$ (see Section 3.1). Condition (ii) requires $\{\varphi_t\}$ to be a family of convex functions. Interpreting the period t discount factor $d\in[0,1]$ as the degree of appreciation of consumption at time t, we interpret $\varphi_t(d)$ as the cognitive cost of achieving d. Each φ_t can take a value of 0 on some interval $[0,\underline{d}]$ but must have standard properties of a cost function on $[\underline{d},1]$, and a bounded slope. Condition (iii) requires that the marginal cost $\varphi_t'(d)$ of producing discount factor d is increasing in t. Integrating and applying the restriction $\varphi_t(0) = 0$ implies that the cost functions are increasing: $\varphi_t \leq \varphi_{t+1}$ for all $t < T$.

Definition 1 (CCE Representation) *A Constrained Costly Empathy (CCE) representation is a tuple $(u, \{\varphi_t\}, K)$, where $(u, \{\varphi_t\})$ is regular and $0 < K \leq \varphi_1(1)$, and $\succsim$ is represented by the function $U : X \rightarrow \mathbb{R}_+$ defined by*

$$U(x) = u(x_0) + \sum_{t\geq 1} D_x(t)u(x_t), \qquad x \in X,$$

where D_x solves

$$D_x = \arg\max_{D\in[0,1]^T} \{\sum_{t\geq 1} D(t)u(x_t) - \sum_{t\geq 1} \varphi_t(D(t))\} \tag{3}$$

subject to

$$\sum_{t\geq 1} \varphi_t(D(t)) \leq K. \tag{4}$$

The unconstrained version of the model defined by setting $K = \infty$ and dropping the restriction $0 < K \leq \varphi_1(1)$ corresponds to NT's *Costly Empathy (CE) representation*. The CCE representation was interpreted in the Introduction, except that we need to intepret the condition $0 < K \leq \varphi_1(1)$. Since the cost functions are increasing in t it follows that in fact we must have

$$0 < K \leq \varphi_t(1), \text{ for all } t > 0.$$

This states that achieving the maximal possible discount factor $D(t) = 1$ for any period $t > 0$ costs at least K. That is, K is exhausted by the time that perfect patience is achieved. Other than being intuitive, the condition also makes the model more tractable, for the following reason. Note that the maximization problem (3) is in fact subject to *two* distinct constraints. The first is the obvious one where D is feasible only if it costs at most K, which we will refer to as the *capacity constraint*: $\sum_{t\geq 1} \varphi_t(D(t)) \leq K$. The second is that D satisfies the *boundary constraint*: $D(t) \leq 1$ for each $t > 0$. The requirement that $\varphi_t(1) \geq K$ for all t guarantees the substantive computational simplication that solving the maximization problem subject to the capacity constraint alone is *sufficient*, in that it satisfies the boundary constraint automatically.

Regarding the interpretation of the capacity constraint K, the CCE model should be viewed as in NT where there exists a cap K on the cognitive resources that can be used for *each* stream. One could also envision a model where K is used for a menu of streams, and can give rise to menu-dependent impatience. Such a model could violate the Weak Axiom of Revealed Preference. We opt to focus on stream-dependent impatience since it is closer to the standard model and more tractable. We leave it to future research to analyze a menu-dependent version of the model.

2.3 Analyzing the CCE Optimization Problem

We make some observations about the utility representation and in particular to the solution D_x to the cognitive optimization problem (3)-(4) in the CCE model. These observations will guide our subsequent behavioral analysis. Consider the set of streams $\{x \in X : \sum_{t\geq 1} \varphi_t(D_x(t)) < K\}$ where the capacity constraint (4) is slack. For such streams, at the solution D_x, the following first order condition holds for each $t > 0$:

$$u(x_t) = \varphi_t'(D_x(t)).$$

By the continuity properties of the representation, this first order condition must hold also for the closure of the noted set. The set of streams where the first order condition holds for all t is therefore defined by:

$$X_s = cl\{x \in X : \sum_{t\geq 1} \varphi_t(D_x(t)) < K\}.$$

Observe that for any stream $x \in X$, current consumption x_0 does not play any role in determining whether x is in X_s or not, since the first order condition depends only on future consumption $x_t, t > 0$. Say that a stream $x \in X$ is *nontrivial* if $u(x_t) > 0$ for some t.

We explore some properties of the representation in relation to X_s and its complement. The first proposition notes a distinguishing feature between streams in X_s and those outside of it: it states that any stream $x \in X_s$ must exhibit a strict magnitude effect when scaled down to αx for *any* $\alpha < 1$, whereas those not in X_s must exhibit no magnitude effect upto some scaling down of the stream.

Proposition 1 (Magnitude Sensitivity) *In the CCE model, if $x \in X_s$ is nontrivial, then impatience is magnitude-sensitive:,*

$$D_{\alpha x} \lneq D_x \text{ for all } \alpha \in (0, 1).$$

If $x \notin X_s$ then impatience is not magnitude-sensitive:

$$D_{\alpha x} = D_x \text{ for all } \alpha \text{ sufficiently close to } 1.$$

Intuitively, if the capacity constraint is not binding, then discounting will be sensitive to incentives provided by the consumption in the stream, whereas such sensitivity will not exist when the capacity constraint binds. We will exploit this in the sequel to behaviorally identify whether the (unobserved) capacity constraint is binding at a stream.[2]

[2]It is worth noting that the property that D_x may not change if x is scaled up can arise in the CE model as well although there is no capacity constraint in it. In NT, φ_t is not necessarily real-valued. Let the supremum discount factor be given by $\overline{d}_t := \sup\{d : \varphi_t(d) < \infty\}$. In NT, once x is large enough that the supremum discount function $t \mapsto D(t) = \overline{d}_t$ is achieved, scaling up will not incentivize higher patience since the cost of doing so is infinite.

Next we present two conditions that will be used in the sequel to formulate behavioral postulates. The first is that the utility preferences is additively separable only on X_s. For any streams $x = (x_1, \cdots, x_T)$ and $y = (y_1, \cdots, y_T)$ denote by (x_t, y_{-t}) the stream that pays x_t at time t and pays $y_{t'}$ at all other $t' \neq t$. So, for instance, if we take $x = (10, 10, 10)$ and $y = (0, 0, 0)$ then $(y_2, x_{-2}) = (10, 10, 0)$ and $(x_2, y_{-2}) = (0, 0, 10)$.

Proposition 2 (Separability) *In the CCE model,*

$$x \in X_s \implies U(x) = U(0_t, x_{-t}) + U(x_t, 0_{-t}) \text{ for all } t > 0.$$

$$x \notin X_s \implies U(x) \leq U(0_t, x_{-t}) + U(x_t, 0_{-t}) \text{ for all } t > 0.$$

The first claim is that when the cognitive constraint does not bind, then the utility of, say, stream $(10, 10, 10)$ is the same as the total utility of say, streams $(10, 10, 0)$ and $(0, 0, 10)$ which sum to the original stream. Intuitively, when the capacity constraint is not binding, the cognitive process can optimally tailor the discount factor $D_x(t)$ for each t separately, and so the discounted utility of a stream is the same as the sum of the discounted utility of sections of the stream. Outside of X_s, such separability will generically be violated due to the cognitive constraint. To see this, consider stream $x = (10, 10, 10)$ with corresponding discount function D_x that is determined with a binding cognitive constraint. Since no cognitive resources are allocated to periods with 0 consumption, and since there are more zero's in the streams $(10, 10, 0)$ and $(0, 0, 10)$, there are more resources available to appreciate the periods with nonzero 0 consumption in these streams compared to $(10, 10, 10)$. Consequently, the discount factors $D_{(10,10,0)}(1)$ and $D_{(0,0,10)}(2)$ must respectively be higher than $D_x(1)$ and $D_x(2)$. The discounted utility form of U then implies the noted inequality.

Finally we observe that in the CCE model, starting with a nontrivial $x \notin X_s$ it is always possible to decreasing the future consumption offered by it until one produces a stream y in X_s. Doing so for any $x \in X_s$ will always just produce another $y \in X_s$. For any x, let $x_{-0} = (0, x_1, \cdots, x_T)$.

Proposition 3 (Threshold) *The following hold for the CCE model:*
(a) If $x \notin X_s$ is nontrivial, then there exists $y_{-0} \lneq x_{-0}$ such that $y \in X_s$.
(b) If $x \in X_s$ is nontrivial and y is such that $y_{-0} \lneq x_{-0}$ then $y \in X_s$.

The intuition is that reducing future consumption reduces the discount function that would optimally arise without cognitive constraints. Thus if the cognitive constraint is slack at x then it must be slack at y as well if $y_{-0} \lneq x_{-0}$. If it is not slack at x, then it can be made slack by taking some sufficiently small y with $y_{-0} \lneq x_{-0}$.

Finally, we argue that the model must have a specific restriction on the boundary of X_s. This corresponds to the set of streams where the FOC holds but the capacity constraint binds:

$$bd(X_s) = \{x : \sum_{t \geq 1} \varphi_t(D_x(t)) = K \text{ and } u(x_t) = \varphi'_t(D_x(t)) \text{ for each } t > 0\}.$$

For any $x \in bd(X_s)$, the FOC implies that the optimal D_x must be given by $D_x(t) = (\varphi'_t)^{-1}(u(x_t))$ for each $t > 0$, and it must cost K. Therefore the boundary can be characterized as follows:

$$bd(X_s) = \{x : \sum_{t\geq 1} \varphi_t((\varphi'_t)^{-1}u(x_t)) = K\}.$$

Observe that $\sum_{t\geq 1} \varphi_t((\varphi'_t)^{-1}u(x_t)) = \sum_{t\geq 1} \varphi_t((\varphi'_t)^{-1}u(y_t))$ must hold for all $x, y \in bd(X_{ms})$, and therefore the CCE model implies some relationship across $bd(X_s)$. Identifying this relationship is *necessary* to axiomatize the CCE model. From Proposition 3(b), we can infer that $bd(X_s)$ has a "negative slope" in that it is never possible to have nontrivial $x, y \in bd(X_s)$ such that $y \lneq x$. But it is unclear whether there are any other behaviorally meaningful properties possessed by $bd(X_s)$. In fact our analysis suggests that a meaningful behavioral characterization of the model may be elusive or even infeasible (see Section 4.2 for the unappealingly technical, albeit exhaustive, restriction on $bd(X_{ms})$ required by the CCE model).

It is natural to relax the problem by searching for a meaningful behavioral characterization within a wider class than the CCE class. Define a "General CCE" representation by requiring that, instead of being a constant, the cognitive capacity is a function $x \mapsto K_x \in (0, \infty]$.[3] The difficulty that arises in following this direction is that the GCCE class generically violates the standard Monotonicity property which requires that, across deterministic streams, $x \geq y \implies x \succsim y$ (larger streams are better).[4]

These issues not withstanding, in the sequel we will identify a GCCE subclass that satisfies both Monotonicity and the main behaviors associated with the CCE model. Our solution to the above hurdles will be to effectively assume that the boundary $bd(X_s)$ is related to an indifference curve: streams in $bd(X_s)$ must be such that they are indifferent in terms of future consumption (if $x, y \in bd(X_s)$ then $(0, x_{-0}) \sim (0, y_{-0})$). Then, the cognitive constraint must be slack for streams that are "sufficiently small" in the sense of their future utility. To the extent that Propositions 1-3 capture the essential content of the CCE model, our main theoretical contribution is to obtain a class of models that have this same essential content (and preserve Monotonicity), while differing from the CCE model only on technical details regarding the structure of $bd(X_s)$. Moreover, the class has the benefit of added tractability: in the model, a single parameter $\overline{v} > 0$ determines whether or not the capacity constraint binds at a given stream x.

[3] Future research might explore how a stream-dependent capacity K_x could be used to capture *salience*: it may be that some streams grab the agent's attention more than others, which result in an expanded capacity to appreciate the stream. For instance, one might hypothesize that moving consumption to earlier rewards may increase the salience of the future consumption offered by the stream.

[4] To see this, define X_s by $cl\{x \in X : \sum_{t\geq 1} \varphi_t(D_x(t)) < K_x\}$ and consider any two deterministic streams $x, y \in bd(X_s)$ that offer strictly positive consumption in every period. There always exists some scaling up of y by some $\lambda^* > 0$ such that $\lambda^* y \gg x$. Note that λ^* is defined by the consumption levels offered by x and y, and the only way the preference plays a role is through the fact that both x, y are in $bd(X_s)$. In order for Monotonicity to hold, it must be that $bd(X_s)$ possesses some property that ensures $\lambda^* y \succ x$. It is hardly clear what this property must be.

3 Behavioral Postulates

For any stream x, we refer to $c_x \in C$ as its *present equivalent* if it satisfies:

$$c_x \sim x.$$

Denote by $p^t \in X$ the stream that pays $p \in \Delta$ at time t and 0 in all other periods. Such a stream is called a *dated reward*.

3.1 Basic Axioms

The following axiom is the same as in NT.

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) (Impatience) For any $p \in \Delta$ *and* $t < t'$,

$$p^t \succsim p^{t'}.$$

(d) (C-Monotonicity) for all $c, c' \in C$,

$$c \geq c' \Longleftrightarrow c \succsim c'.$$

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

$$(x_t, 0, \cdots, 0) \succsim (y_t, 0, \cdots, 0) \text{ for all } t \Longrightarrow x \succsim y.$$

Moreover, if $(x_t, 0, \cdots, 0) \succ (y_t, 0, \cdots, 0)$ *for some* t*, then* $x \succ y$*.*
(f) (Risk Preference) For any $p, p', p'' \in \Delta$ *and* $\alpha \in (0, 1]$,

$$p \succ p' \Longrightarrow \alpha \circ p + (1 - \alpha) \circ p'' \succ \alpha \circ p' + (1 - \alpha) \circ p''.$$

(g) (Present Equivalents) For any stream x there exists $c_x \in C$ s.t.

$$c_x \succsim x.$$

Order and Continuity are standard. Impatience states that consumption is better when received sooner. C-Monotonicity states that more consumption is better than less. While C-Monotonicity applies only to immediate consumption, Monotonicity is a property on arbitrary streams: it requires that point-wise preferred streams are preferred. Present Equivalents states that for any stream, there are immediate consumption levels that are better than x. Given Order and Continuity, this ensures that each stream x has a present equivalent $c_x \in C$. Notably, each x has a unique present equivalent c_x (by C-Monotonicity, $x \sim c_x > c_y \sim y$ implies $c_x \succ c_y$ and therefore $x \succ y$). Risk Preference imposes vNM Independence only on immediate consumption.

3.2 Magnitude-Sensitivity

To identify whether an agent exhibits a magnitude effect (that is, greater patience towards larger rewards), we follow the "lottery approach" considered by NT, which we describe now.[5] For any $p \in \Delta$ and $\alpha \in [0,1]$ define the mixture $\alpha \circ p := \alpha \circ p + (1-\alpha) \circ 0$. In particular, for $c \in C$, we write this lottery as $\alpha \circ c$ in order to distinguish $\alpha \circ c \in \Delta$ with a deterministic consumption $\alpha c \in C$. For any stream $x = (x_0, \cdots, x_T)$ define $\alpha \circ x := (\alpha \circ x_0, \cdots, \alpha \circ x_T)$. We will say that "$\alpha \circ x$ scales down stream x by α" to mean that it scales down the probability of receiving x by α. Abusing notation, write $\alpha \circ p$ for the stream $(\alpha \circ p, 0, \cdots, 0)$. Consider a stream x and its present equivalent $c_x \sim x$. Note that the agent's evaluation of immediate consumption c_x does not rely on impatience whereas that of a stream x does. If impatience does *not* change in response to scaling down x by α, then it must be that:

$$\alpha \circ c_x \sim \alpha \circ x,$$

since the scaling down affects the evaluation of consumption equally for the immediate reward and the stream. On the other hand, if scaling down x by α increases the agent's impatience, then $\alpha \circ x$ must lose its desirability faster than the immediate reward $\alpha \circ c_x$ (for which impatience is irrelevant):

$$\alpha \circ c_x \succsim \alpha \circ x.$$

Accordingly, NT behaviorally define an agent who's impatience is decreasing in the magnitude of rewards as one who exhibits:

Axiom 2 (Weak Homotheticity) *For any $x \in X$ and any $\alpha \in (0,1)$,*

$$c_x \sim x \Longrightarrow \alpha \circ c_x \succsim \alpha \circ x.$$

The special case where $\alpha \circ c_x \sim \alpha \circ x$ always holds is termed *Homotheticity* and is used, along with Regularity and Separability (defined below) to characterize the DU representation. Since we saw in Proposition 1 that cognitive optimization implies magnitude-decreasing impatience, we maintain Weak Homotheticity. For later reference, we say that a stream $x \in X$ is *magnitude-sensitive* if the agent's impatience strictly reduces *whenever* the stream is made less desirable:

Definition 2 (Magnitude-Sensitivity) *A stream $x \in X$ is magnitude sensitive if*

$$c_x \sim x \Longrightarrow \alpha \circ c_x \succ \alpha \circ x \textit{ for all } \alpha \in (0,1).$$

By vNM Independence, it is clear that immediate rewards are not magnitude sensitive. Thus any magnitude-sensitive stream must have $x_t \succ 0$ for some $t > 0$ (which corresponds to "nontriviality" as defined in Section 2.3).

[5] Ideally, a temporal property like impatience should be behaviorally defined without reference to risk preferences. NT also study an alternative approach based on the marginal rate of intertemporal substitution (MRS). We take the lottery approach here since it communicates the main ideas more easily. We expect that the ideas are straightforward to translate into the MRS approach.

3.3 Non-Separability

Proposition 2 suggests that a non-binding cognitive constraint will give rise to both magnitude-sensitivity and a separability property, whereas both properties are violated if the cognitive constraint binds. We formulate Separability as in NT, and accordingly define Non-Separability.

Suppose we measure the desirability of a stream $(10, 10, 10)$ by its certainty equivalent $c_{(10,10,10)}$. Pick any period, say period 2, and consider "breaking" the stream into two streams like so: $(10, 10, 0)$ and $(0, 0, 10)$. Measure the desirability of these by the certainty equivalents $c_{(10,10,0)}$ and $c_{(0,0,10)}$. Intuitively, Separability should require that the "total desirability" of the streams $(10, 10, 0)$ and $(0, 0, 10)$ must in some sense be the same as that of $(10, 10, 10)$ Formally, NT's *Separability* axiom requires the the $\frac{1}{2}$- mixtures of present equivalents of $(10, 10, 0)$ and $(0, 0, 10)$ must be indifferent to that of $(10, 10, 10)$ and $(0, 0, 0)$:[6] For all $x \in X$ and all $t > 0$,

$$\frac{1}{2} \circ c_x + \frac{1}{2} \circ c_0 \sim \frac{1}{2} \circ c_{(x_t, 0_{-t})} + \frac{1}{2} \circ c_{(0_t, x_{-t})}.$$

As suggested in Section 5, cognitive constraints give rise to violations of Separability. Intuitively, if the agent is cognitively constrained then she has a limited amount of cognitive resources at her disposal to generate patience. Since, compared to $(10, 10, 10)$, there are fewer periods to spend cognitive resources on in the streams $(10, 10, 0)$ and $(0, 0, 10)$, the agent should exhibit more patience at these streams compared to $(10, 10, 10)$. Consequently, these streams should jointly be more attractive than $(10, 10, 10)$. The direction of the violation of Separability must therefore be as follows:

Axiom 3 (Weak Separability) *For all $x \in X$ and all $t > 0$,*

$$\frac{1}{2} \circ c_x + \frac{1}{2} \circ c_0 \precsim \frac{1}{2} \circ c_{(x_t, 0_{-t})} + \frac{1}{2} \circ c_{(0_t, x_{-t})}.$$

For later reference, we define separability of a stream $x \in X$:

Definition 3 (Separable Streams) *A stream $x \in X$ is separable* if for *all $t > 0$,*

$$\frac{1}{2} \circ c_x + \frac{1}{2} \circ c_0 \sim \frac{1}{2} \circ c_{(x_t, 0_{-t})} + \frac{1}{2} \circ c_{(0_t, x_{-t})}.$$

[6] In a deterministic setting, the Separability condition in Koopmans [12] is strong enough to guarantee additive separability of a representation only when it is defined on product domains, a feature that is not satisfied in our model, where Separability must hold only on the set of magnitude-sensitive streams. If we enrich the domain to include lotteries, then additive separability can be imposed via an vNM Independence condition. In our context, we could simply use such an Independence condition if we had the richer domain of *lotteries over streams.* However, in order to avoid adding a second layer of lotteries to their domain, NT formulate a Separability condition by exploiting mixtures of certainty equivalence.

Observe that dated rewards are trivially separable. Therefore non-separability can be exhibited only by streams that have at least two non-zero future rewards.

We have seen in Section 2.3 (Propositions 1 and 2) that streams where the cognitive constraint binds must not be magnitude-sensitive and may fail separability, while streams where the cognitive constraint does not bind must be separable. These motivate the following key behavioral condition: For any $x \in X$,

$$x \text{ is not separable} \Longrightarrow x \text{ is not magnitude-sensitive.}$$

In order to avoid stating negatives, we write this key condition in its contrapositive form:

Axiom 4 (Magnitude-Sensitive (MS) Separability) *For any $x \in X$, if x is magnitude-sensitive then it is separable.*

Denote by X_{ms} the set of all streams that are magnitude-sensitive and separable:

$$X_{ms} = \{x \in X : x \text{ is magnitude-sensitive and separable}\}.$$

Since immediate rewards cannot be magnitude-sensitive, Δ_0 lies outside of X_{ms}. The streams outside of $X_{ms} \cup \Delta_0$ consists only of magnitude-insensitive dated rewards (which are trivially separable) and non-separable magnitude-insensitive streams. It is natural to intepret streams in X_{ms} as those at which the agent's cognitive constraint is not binding, and streams outside of $X_{ms} \cup \Delta_0$ as those at which the constraint is binding.[7]

3.4 Threshold Condition

Proposition 3 suggests that when there exist cognitive capacity constraints, a "sufficient reduction" in any stream will relax the cognitive constraint. Moreover, if the cognitive constrain is already slack at a stream, then it will be slack for any reduction of the stream. Given our behavioral definition of X_{ms}, this feature can be readily formulated in an axiom that mirrors Proposition 3 in that it replaces X_s with X_{ms}. However the subsequent discussion in Section 2.3 argued that some structure needs to be imposed on the threshold between constrained and unconstrained streams, that is, on the boundary of X_{ms}. Based on that discussion, we will effectively presume that the boundary of X_{ms} is determined by an indifference condition. Our next axiom mirrors the claims in Proposition 3 but replaces the $\geq$-dominance requirement with a $\succsim$-dominance requirement. In light of our Monotonicity axiom, this amounts to a strengthening of the claims in that proposition.

For any $x \in X$, recall that x_{-0} denotes the stream that pays 0 in period 0 and pays according to x from period 1 onward. That is, $x_{-0} = (0, x_1, \cdots, x_T)$. We impose

[7] Cognitive constraints are not relevant for immediate streams since we are interested in cognitive resources that are used to appreciate future rewards only. For the sake of comparison with Section 2.3 we note that for a given CCE model, X_{ms} is a strict subset of the set X_s, since X_{ms} excludes Δ_0 while X_s does not. The relevance of magnitude-sensitivity is suggested by the first claim in Proposition 1. Accordingly, we interpret the magnitude-sensitivity of the streams in X_{ms} as *revealing* that the agent's cognitive constraint is not binding. The behavioral counterpart of the second claim in Proposition 1 will be implied by our axioms.

Axiom 5 (Preference-Based Threshold) *For all $x \in X$, the following hold.*
(i) if $x \notin X_{ms} \cup \Delta_0$, then there exists $y \in X_{ms}$ such that $x_{-0} \succsim y_{-0}$.
(ii) if $x \in X_{ms}$ and y satisfy $x_{-0} \succsim y_{-0}$, then $y \in X_{ms}$.

This axiom requires that the magnitude sensitivity of a stream x should be associated with the "future utility" of the steam. Lower (resp. higher) future utility arises simultaneously with lower (resp. higher) consumption levels and therefore slack (resp. binding) cognitive constraints. This property is reminiscent of Becker and Mulligan [3], where there exists complementarity between time preference and future utilities. In their model, higher physical wealth leads the agent to invest more resources in both increasing future consumption as well as the future oriented capital. This leads to decreasing impatience for large future utilities.

4 Representation Result: CCE* Model

Say that a tuple $(u, \{\varphi_t\}, \overline{v})$ is *regular** if $(u, \{\varphi_t\})$ is regular and $0 < \overline{v} \leq \varphi_1'(1)$. We will point out the relevance of the last condition after interpreting the following definition.

It will be useful to introduce some notation, specifically, the discount function D_x^{un} obtained by solving the unconstrained optimization problem. If a discount function D_x^{un} solves the *unconstrained* cognitive optimization problem, then it satisfies the first order condition $u(x_t) = \varphi_t'(D_x^{un}(t))$ for each $t > 0$, and consequently it must satisfy

$$D_x^{un}(t) := (\varphi_t')^{-1}(u(x_t)).$$

This equality defines D_x^{un} for *all* streams, including those streams at which the cognitive constraint binds when solving the constrained optimization problem.

Definition 4 (CCE* Representation) *A CCE* representation is a regular* tuple $(u, \{\varphi_t\}, \overline{v})$ such that $\succsim$ is represented by the function $U : X \to \mathbb{R}_+$ defined as follows:*

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

where D_x solves

$$D_x = \arg\max_{D\in[0,1]^T} \{\sum_{t>0} D(t)u(x_t) - \sum_{t>0} \varphi_t(D(t))\} \tag{5}$$

subject to

$$\sum_{t>0} \varphi_t(D(t)) \leq K_x \tag{6}$$

where $K_x := \sum_{t=1}^T \varphi_t(D_{\lambda_x \circ x}^{un}(t))$ is defined using the unique solution λ_x to

$$\sum_{t>0} D_{\lambda\circ x}^{un}(t)u(\lambda \circ x_t) = \overline{v}. \tag{7}$$

The difference from the CCE model is that the cognitive capacity K_x may vary with the stream x. The functional form requires that K_x for stream x equals the cost of the unconstrained discount function $D^{un}_{\lambda_x \circ x}$ for a scaled stream $\lambda_x \circ x$. The scaling λ_x is determined by solving (7), requiring that the unconstrained discounted utility for $\lambda_x \circ x$ must equal $\overline{v}$. Note that $D^{un}_{\lambda \circ x}$ and the left-hand side of (7) are increasing in λ. Therefore if $\lambda_x \geq 1$, that means that K_x is high enough that the agent's cognitive capacity does not bind at x. If $\lambda_x < 1$, then the cognitive capacity must be binding. Thus the model offers a simple tool for analysis: to determine whether the cognitive constraint binds or not at x, compute the unconstrained discounted utility of x and check if it is higher or lower than $\overline{v}$. The existence of a solution to (7) is guaranteed by the assumption that $0 < \overline{v} \leq \varphi_1'(1)$.[8]

4.1 Representation Theorem

If all $x \in X$ are separable, then we are in the context of the CE model. So our interest lies in preferences that are *nonseparable* in that there exists $x \in X$ that is not separable. Our general representation result is that:

Theorem 1 *A nonseparable preference $\succsim$ on X satisfies Regularity, Weak Homotheticity, MS Separability and Preference-Based Threshold if and only if it admits a CCE* representation.*

NT prove that a preference on X satisfies Regularity, Weak Homotheticity and Separability if and only if it admits an (unconstrained) CE representation. MS Separability relaxes Separability, while Preference-Based Threshold adds structure to the homotheticity violations permitted by Weak Homotheticity.

The CCE* model has strong uniqueness properties. For any CCE* representation $(u, \{\varphi_t\}, \overline{v})$, define the maximal discount factor $\overline{D}(t)$ by the equation $\overline{D}(t)\varphi_t'(\overline{D}(t)) = \overline{v}$. Denote by $\varphi_t|_{[0,\overline{D}(t)]}$ the restriction of φ_t to the subdomain $[0, \overline{D}(t)]$.

Theorem 2 *If there are two CCE* representations $(u^i, \{\varphi_t^i\}, \overline{v}^i)$, $i = 1, 2$ of the same preference $\succsim$, then (1) $\overline{D}^1(t) = \overline{D}^2(t) = \overline{D}(t)$ for all $t > 0$, and (2) there exists $\alpha > 0$ such that (i) $u^2 = \alpha u^1$, (ii) $\varphi_t^2|_{[0,\overline{D}(t)]} = \alpha\varphi_t^1|_{[0,\overline{D}(t)]}$ for all $t > 0$, and (iii) $\overline{v}^2 = \alpha\overline{v}^1$.*

Thus the tuple $(u, \{\varphi_t\}, \overline{v})$ is unique upto a common scalar multiple. The sharp uniqueness is obtained from the separability of the representation on the subdomain X_{ms} and the fact that $u(0) = 0$ is presumed in the representation

Although we did not impose Weak Separability in the main theorem, we verify that it is in fact implied by the other axioms:

[8] Since φ_t' is increasing in t (by definition of regularity of $(u, \{\varphi_t\})$), this assumption implies $0 < \overline{v} \leq \varphi_t'(1)$ for all $t > 0$. Note that the function $L(\lambda) := D^{un}_{\lambda x}(t)u(\lambda x_t) = \sum_{t=1}^{T}(\varphi_t')^{-1}(u(\lambda x_t))u(\lambda x_t)$ is increasing and continuous. Moreover, since $\varphi_t'(1) < \infty$ (given in the definition of regularity of $(u, \{\varphi_t\})$), there exists $\overline{\lambda}$ such that $u(\overline{\lambda}x) \geq \varphi_t'(1) \geq \overline{v}$ for all $t > 0$, and thus in particular $(\varphi_t')^{-1}(u(\overline{\lambda}x_t)) \geq 1$ and $u(\overline{\lambda}x) \geq \overline{v}$. It follows that $L(\overline{\lambda}) \geq T\overline{v} \geq \overline{v}$. At the same time $L(0) = 0$. Therefore the intermediate value theorem guarantees the existence of a solution to (7).

Proposition 4 *If $\succsim$ admits a CCE* representation then it must satisfy Weak Separability.*

We close this subsection with a reduced form of the model (with nontrivial constraints) that may be useful for applications. The proof for this result is in the supplementary appendix (Noor and Takeoka [21]).

Proposition 5 *A preference $\succsim$ on X admits a CCE* representation $(u, \{\varphi_t\}, \overline{v})$ if and only if it admits the following representation defined by the tuple $(u, \{\varphi_t\}, \overline{v})$:*

$$U(x) = u(x_0) + \sum_{t>0} (\varphi_t')^{-1}(u(\alpha_x x_t))u(x_t),$$

where

$$\alpha_x = \max \left\{ \alpha \leq 1 \;\middle|\; \sum_{t>0} \alpha u(x_t)(\varphi_t')^{-1}(\alpha u(x_t)) \leq \overline{v} \right\}.$$

It is evident that when $\alpha_x = 1$ the model is additively separable, whereas when $\alpha_x < 1$ then it can violate Separability by virtue of the fact that α_x depends on the stream (specifically on future consumption $x_1, \cdots, x_T$).

An equivalent way to write the reduced form representation is:

$$U(x) = \begin{cases} u(x_0) + \sum_{t>0} (\varphi_t')^{-1}(u(x_t))u(x_t) & \text{if } \sum_{t>0} (\varphi_t')^{-1}(u(x_t))u(x_t) \leq \overline{v}, \\ u(x_0) + \max_{D \in \mathcal{D}_x} \sum_{t>0} D(t)u(x_t) & \text{otherwise,} \end{cases}$$

where

$$\mathcal{D}_x := \{D \in [0,1]^T \;:\; \varphi(D) = K_x\}.$$

Moreover, the maximizer over $\mathcal{D}_x$ satisfies

$$D_x(t) = (\varphi_t')^{-1}(\alpha_x u(x_t)), \; \forall t > 0.$$

An interesting observation is that when restricted to streams that are "large" (in the sense that the future utility $\sum_{t>0} (\varphi_t')^{-1}(u(x_t))u(x_t)$ of the stream strictly exceeds $\overline{v}$), the CCE* representation is a maxmin-type (more precisely, maxmax-type) representation à la Gilboa and Schmeidler [11]. The max operator implies that the CCE* representation satisfies convexity ($x \sim y \implies \alpha \circ x + (1 - \alpha) \circ y \succsim y$ for all $\alpha \in [0, 1]$) for large streams. Since the model is Separable for small streams, it satisfies convexity on that subdomain as well. It may violate convexity when mixing a large and a small stream, however. The CCE* model satisfies Weak Homotheticity, which is a property weaker than convexity and corresponds to star-shapedness of the representation (that is, $\alpha U(x) \geq U(\alpha \circ x)$ for all $\alpha \in [0, 1]$). Since our model violates convexity, it goes beyond models of convex preferences in the literature (such as Maccheroni et al [16]).

4.2 Proof Outline and Foundations of CCE

A proof sketch of sufficiency of Theorem 1 is as follows. Regularity and the separability of the streams in X_{ms}, yield an additively separable representation $U(x) = \sum_{t\geq 0} U(x_t)$ on X_{ms}. This representation can be rewritten as a discounted utility with stream-dependent discount function simply by defining $u(p) = U_0(p)$ and $D_x(t) = \frac{U_t(x_t)}{u(x_t)}$. Given u and D_x, we can derive an additive cost function $\varphi = \sum \varphi_t$ for which D_x is optimal – the cost function is defined by the desired cognitive first order condition $u(x_t) = \varphi'_t(D_x(t))$ for each $t > 0$. Here, the cost function φ_t is increasing and convex. This establishes that the preference over X_{ms} has a CCE* representation without any constraint.

The second step is to extend this representation to the whole domain. For any $x \in X \setminus \Delta_0$, consider the ray from the origin passing through x, that is, $\{\alpha \circ x \,|\, \alpha > 0\}$. Weak Homotheticity and Preference-Based Threshold imply that there exists a unique α_x such that $\alpha \circ x \in X_{ms}$ if and only if $\alpha \leq \alpha_x$. The stream $\alpha_x \circ x$ can be viewed as lying on the "boundary" of X_{ms}. We show that $\succsim$ satisfies "Homotheticity outside X_{ms}" in the sense that, for any $x \notin X_{ms} \cup \Delta_0$, the preference satisfies $c_x \sim x \implies \alpha \circ c_x \sim \alpha \circ x$ for all $\alpha \in [\alpha_x, 1]$. From this condition, the representation on X_{ms} can be extended by defining $U(x) = u(c_x) = U(\alpha_x \circ x)/\alpha_x$. Moreover, since $\alpha_x \circ x \in X_{ms}$, the utility $U(\alpha_x \circ x)$ admits an additively separable representation by the first step. Thus, $U(x)$ is more explicitly written as $U(x) = u(x_0) + \sum_{t\geq 1} D_{\alpha_x \circ x}(t) u(x_t)$.

The remaining problem is to infer a capacity constraint K_x and to show that $D_{\alpha_x \circ x}$ can be regarded as an optimal discount function for x under the constraint. As shown above, along a ray $\{\lambda \circ x \,|\, \lambda > 0\}$, as λ increases, $D_{\lambda \circ x}$ should first strictly increase (specifically, as long as $\lambda \circ x \in X_{ms}$) and eventually become constant once λx crosses the boundary of X_{ms}. The main step in proving the theorem is to find the set of discount functions, Λ_x, for which the optimal D computed subject to the constraint $D \in \Lambda_x$ is precisely $D_{\lambda_x \circ x}$. Define $K_x = \varphi(D_{\lambda_x \circ x})$ so that K_x is the total empathy cost of the unconstrained optimal discount function at the boundary of X_{ms}. The proof verifies that the constraint

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

does the job. The remaining step is to obtain the characterization for K_x using $\overline{v}$. By Preference-Based Threshold, all points on the boundary of X_{ms} have the same future utility $\overline{v} > 0$. Then the scalar λ_x used to compute K_x is determined by an equation requiring that the unconstrained utility of $\lambda_x \circ x$ to be $\overline{v}$.

We close by commenting on the characterization of the CCE model. For any stream $x \in bd(X_{ms})$, integration by parts implies that:

$$
\begin{aligned}
K_x = \varphi(D_x) &= \sum_{t\geq 1} \int_0^{u(x_t)} \varphi_t'(D_r(t)) \frac{\mathrm{d}D_r}{\mathrm{d}r}(t)\,\mathrm{d}r \\
&= \sum_{t\geq 1} \int_0^{u(x_t)} r \frac{\mathrm{d}D_r}{\mathrm{d}r}(t)\,\mathrm{d}r \\
&= \sum_{t\geq 1} \left[u(x_t) D_{u(x_t)}(t) - \int_0^{u(x_t)} D_r(t)\,\mathrm{d}r \right] \\
&= U(0, x_{-0}) - \sum_{t\geq 1} \int_0^{u(x_t)} D_r(t)\,\mathrm{d}r. \qquad (8)
\end{aligned}
$$

The technical condition that is needed to characterize the CCE model is that $K_x = K_y$ for all $x, y \in bd(X_{ms})$. This requires that for all $x, y \in bd(X_{ms})$,

$$
(U(x) - u(x_0)) - \sum_{t\geq 1} \int_0^{u(x_t)} D_r(t)\,\mathrm{d}r = (U(y) - u(y_0)) - \sum_{t\geq 1} \int_0^{u(y_t)} D_r(t)\,\mathrm{d}r.
$$

Although this technical restriction can in principle be expressed behaviorally[9], it does not translate into economically interesting behavior. This suggests that the "constant K" feature of CCE may not be of any particular economic interest. We propose the CCE* class since it contains all that is of economic interest in the CCE class, still intersects with it, and has a benefit of added tractability.

5 Representation Result: Homogeneous CCE Model

Most applications of the CCE* model will likely assume that costs have the power form $\varphi_t(d) = a_t \cdot d^m$. As it turns out, this defines a particularly interesting special case of our model.

The following restriction is taken from NT:

Axiom 6 (Magnitude-Sensitive (MS) Homogeneity) *For any magnitude-sensitive dated rewards $p^t, q^s \in X_{ms}$, their present equivalents $c_{p^t} \sim p^t$ and $c_{q^s} \sim q^s$, and any $\alpha, \beta \in (0, 1)$,*

$$
\beta \circ c_{p^t} \sim \alpha \circ p^t \Longrightarrow \beta \circ c_{q^s} \sim \alpha \circ q^s.
$$

[9] The expression (8) can be measured behaviorally as follows. The utility index u is an expected utility over lotteries $\Delta = \Delta(C)$, and so can be defined in the usual way: normalize $u(0) = 0$ and $u(c_1) = 1$ for some arbitrary $c_1 \in C$, and consequently for any c the utility $u(c)$ is given by $\theta \in \mathbb{R}_+$ s.t. $\frac{1}{\theta} \circ c \sim c_1$. With u defined behaviorally, we can take the present equivalent $c_{(0,x_{-0})} \sim (0, x_{-0})$ and behaviorally measure the term $U(0, x_{-0})$ by the utility $u(c_{(0,x_{-0})})$. For any c and t, the discount factor $D_{u(c)}(t)$ is defined by the $\theta \in [0, 1]$ s.t. $\theta \circ c \sim c^t$. Consequently the term $\sum_{t\geq 1} \int_0^{u(x_t)} D_r(t)\,\mathrm{d}r$ can be behaviorally measured.

MS Homogeneity states that if scaling down p^t by α is as good as scaling down its present-equivalent c_{p^t} by β, then β depends on α but not the dated reward. When imposed on the CCE* model, it yields the power form for cost functions, as intended. Surprisingly, it also forces K_x to be constant, thereby leading to a CCE model.

Theorem 3 *A preference $\succsim$ on X satisfies the CCE* axioms and MS Homogeneity if and only if it admits a Homogenous CCE representation, that is, a CCE representation $(u, \{\varphi_t\}, K)$ where for each $t \geq 1$, the cost function $\varphi_t : [0,1] \to \mathbb{R}_+$ takes the power form*

$$\varphi_t(d) = a_t \cdot d^m,$$

and a_t is increasing in t.

The uniqueness properties are given by:

Theorem 4 *If there are two Homogeneous CCE representations $(u^i, \{\varphi_t^i\}, K^i)$, where $\varphi_t^i(d) = a_t^i \cdot d^{m_i}$, $i = 1, 2$, of the same preference $\succsim$, then there exists $\alpha > 0$ such that (i) $u^2 = \alpha u^1$, (ii) $a_t^2 = \alpha a_t^1$ and $m_2 = m_1$, and (iii) $K^2 = \alpha K^1$.*

This is a corollary of Theorem 2. Part (i) is the same as the counterpart of that theorem. By part (ii) of Theorem 2, $\varphi_t^2|_{[0,\overline{[D(t)]}]} = \alpha \varphi_t^1|_{[0,\overline{D}(t)]}$, which pins down the curvature of the cost function, that is, $m_1 = m_2$, and hence, $a_t^2 = \alpha a_t^1$ follows. Finally, since $\overline{D}^1(t) = \overline{D}^2(t) = \overline{D}(t)$ as shown in Theorem 2, the regularity* condition implies $a_t^i(\overline{D}(t))^m = K^i$, $i = 1, 2$, which boils down to $K^2 = \alpha K^1$, as desired.

A proof sketch of sufficiency of Theorem 3 is as follows. Start with the CCE* representation. MS Homogeneity implies that φ_t satisfies the multiplicative Cauchy functional equation and so we conclude that it has the power form. To see why K_x is constant, recall from Section 4.2 that for any $x \in bd(X_{ms})$ the capacity K_x satisfies the expression (8). When φ_t has the power form, then a simple calculation (also implied by Euler's Theorem for homogeneous functions) yields that

$$K_x = \sum_{t \geq 1} \varphi_t(D_x(t)) \propto U(x) - u(x_0). \tag{9}$$

But Preference-Based Threshold already requires that $U(x) - u(x_0) = U(y) - u(y_0)$ for all streams $x, y \in bd(X_{ms})$. Therefore K_x is constant for all streams on the boundary of X_{ms}. This completes the proof because for any general stream x, the capacity K_x is constant along the ray, that is, $K_x = K_{\lambda \circ x}$ for all $\lambda > 0$, and thus is completely defined by the X_{ms}-boundary point $\lambda_x \circ x$.

5.1 Reduced Form

It is instructive to explicitly derive the reduced form of the Homogeneous CCE model. For any stream, the optimal discount function D_x satisfies the FOC of Lagrangian:

$$\mathcal{L} = \sum_{t\geq 1}(D(t)u(x_t) - \varphi_t(D(t))) + \xi(K - \sum_{t\geq 1}\varphi_t(D(t))),$$

where $\xi \geq 0$ is a Lagrange multiplier. Given the functional form of φ_t and the FOC with respect to $D(t)$ given by $u(x_t) = (1+\xi)\varphi_t'(D(t))$, we have

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

If the capacity constraint is slack, $\xi = 0$, and hence,

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

in which case $D_x(t)$ depends only on $u(x_t)$. If the capacity constraint is binding, then by definition, D_x satisfies

$$\sum_{t\geq 1}\varphi_t\left(\left(\frac{u(x_t)}{(1+\xi)ma_t}\right)^{\frac{1}{m-1}}\right) = K.$$

Solving for ξ and substituting it back into $D_x(t)$ yields

$$D_x(t) = \frac{(mK)^{\frac{1}{m}}\gamma(t)u(x_t)^{\frac{1}{m-1}}}{\left\{\sum_{\tau\geq 1}\gamma(\tau)u(x_\tau)^{\frac{m}{m-1}}\right\}^{\frac{1}{m}}},$$

which is not time-separable, that is, $D_x(t)$ depends on the entire stream x and not just payoff at time t. For large $\alpha > 1$, the capacity constraint binds for $\alpha \circ x$, and the optimal discount function $D_{\alpha\circ x}$ stops growing with α.

By virtue of being a special case of CCE*, this model admits a clean way of distinguishing magnitude sensitive and other streams in terms of the representation. For any stream x, the discounted "future utility" achieved from x is given by $\sum_{t\geq 1} D_x(t)u(x_t) = U(x) - u(x_0)$. In the Homogeneous CCE representation, a stream x is magnitude sensitive iff its future payoff $U(x) - u(x_0)$ is less than some threshold.[10] This is expressed in the next proposition.

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

[10]This property is reminiscent of Becker and Mulligan [3], that derive an observation about complementarity between time preference and future utilities. In our model, up to the threshold, impatience decreases in future payoffs, it achieves the minimum impatience at the threshold, and becomes invariant beyond that.

Proposition 6 *If $\succsim$ admits a CCE representation $(u, \{a_t d^m\}_{t=1}^T, K)$, then*

$$U(x) = \begin{cases} u(x_0) + \sum_{t>0} \gamma(t) u(x_t)^{\frac{m}{m-1}} & \text{if } \sum_{t>0} \gamma(t) u(x_t)^{\frac{m}{m-1}} \leq mK \\ u(x_0) + (mK)^{\frac{1}{m}} \left\{ \sum_{t>0} \gamma(t) u(x_t)^{\frac{m}{m-1}} \right\}^{\frac{m-1}{m}} & \text{if } \sum_{t>0} \gamma(t) u(x_t)^{\frac{m}{m-1}} > mK \end{cases}.$$

The proof is in the supplementary appendix (Noor and Takeoka [21]). According to the reduced form, when the future utility of a stream is "small", the utility function is additively separable (as in Proposition 2), and future utility index is effectively $u(x_t)^{\frac{m}{m-1}}$, a power transformation of immediate utility u. Since $u(p)$ is an expected utility, risk preferences are unchanged with t. However, the parameter m will affect intertemporal substitution. When the future utility of a stream is "large", the utility function is no longer additively separable: future utility is evaluated using a concave aggregator.

6 Applications

We study properties of the model in two settings: a consumption-savings problem and a task-completion problem. Below we use the terms "sophisticated" and "naive" in the sense of O'Donoghue and Rabin [23]. The proofs for the results in this section can be found in the supplementary appendix (Noor and Takeoka [21]).

6.1 Consumption-Savings Problem

Suppose that there are only 3 periods, $t = 0, 1, 2$, and consumption space is given by $C = \mathbb{R}_+$. Suppose that u is a power function $u(c) = c^\sigma$, and that the cost function is homogenous, $\varphi_t(d) = a_t d^m$ for some $m > 1$ and $0 < a_1 \leq a_2$. Self 0 is a CCE agent $(u, \{\varphi_1, \varphi_2\}, K_0)$ and Self 1 is a CCE agent $(u, \{\varphi_1\}, K_1)$ where $K_0, K_1 \leq a_1$, while Self 2 simply maximizes u. It will be natural to assume $K_0 = K_1$ but we leave the model more general for the sake of performing comparative statics.

Suppose that income in periods 0 and 1 are I_0 and I_1 respectively, and there is no income in period 2. The rate of interest is $r > 0$, and define $R = 1 + r$. The agent faces the usual intertemporal budget constraints:

$$c_0 + s_0 = I_0,\ c_1 + s_1 = I_1 + Rs_0,\ c_2 = Rs_1,$$

and we disallow borrowing, $s_t \geq 0$ for $t = 0, 1$.

6.1.1 Comparative Statics: Cognitive Capacity

Assume that both selves are cognitive constrained at the solution to the consumption-savings problem. Assume also that $0 < \sigma \frac{m}{m-1} < 1$, which says that the curvature of φ_t, parametrized by $m > 1$, is not too low. This is to ensure that the first order conditions are sufficient to establish a solution. Denoting the optimal saving rules by s_t^*, $t = 0, 1$, we consider impact on saving of an infinitesimal change in cognitive capacity.

Proposition 7 *Assume* $0 < \sigma\frac{m}{m-1} < 1$ *and suppose that both self 0 and self 1 are cognitively constrained at their respective optimal consumption path. Then the following hold for the sophisticated CCE model along the optimal consumption path:*

(i) s_1^* *is increasing in* K_1.

(ii) s_0^* *is increasing in* K_0.

(iii) s_0^* *is increasing in* K_1 *if and only if* $a_1/a_2 > (R^{-\sigma}(K_1/a_1)^{1-\sigma-\frac{1}{m}})^{\frac{1}{1-\sigma}}$.

(iv) Let $K_0 = K_1 = K$. *Then* s_0^* *is increasing in* K *if* $a_1/a_2 > (R^{-\sigma}(K/a_1)^{1-\sigma-\frac{1}{m}})^{\frac{1}{1-\sigma}}$.

The first two claims confirm the intuition that if the cognitive capacity of a given self increases then they become more patient. The third claim reveals a nuance. If self 1's cognitive capacity is increased, then self 0's response depends on parameters. The reason is that a change in self 1's preferences may or may not exacerbate the dynamic inconsistency anticipated by self 0. Because a_1, a_2 respectively determine the cost functions φ_1, φ_2, the ratio a_1/a_2 captures self 0's (magnitude-dependent) weighting of self 1 vs self 2. If a_1/a_2 is "low" then the agent does not care much for self 2 relative to self 1. However, if K_1 increases, then self 1 will increasingly care for self 2, thereby exacerbating dynamic inconsistency. In this case, self 0 will choose to reduce her savings.

The fourth claim in the proposition applies to the more natural specification of the model where both self 0 and self 1 have the same capacity K. While initial intuitions may have suggested that greater capacity would lead all selves to become more patient, the nuance uncovered in the third claim determines this intuition to be inaccurate. There is an interesting take-away from this. If field evidence in future research confirms the intuitive claim that savings are higher among less cognitively constrained agents, then such a finding would in fact be lending support to the hypothesis of naivete. Instead, if savings are found to be lower, then evidence for sophistication would be obtained.

6.1.2 Calibrating DU: Observational Non-Equivalence

Can the consumption-savings profile of a Sophisticated CCE agent be explained by a Sophisticated beta-delta model (Laibson [14])? It is readily shown that:

Proposition 8 *Suppose* $0 < \sigma\frac{m}{m-1} < 1$ *and consider the consumption profiles of a Sophisticated CCE agent with* $K_0 = K_1$ *that is constrained along the optimal consumption paths at any two distinct* $R, R' > 1$. *There does not exist a Sophisticated beta-delta model (with utility index* u*) that can simultaneously match these consumption profiles.*

Since the standard exponential discounting model is a special case of the Sophisticated beta-delta model, the result establishes observational non-equivalence with the standard model as well. The reason is that the nonseparability of the CCE model comes into play as R varies. In the proof we show that any calibrating beta-delta model must strictly change with R. While the algebra in the proof is greatly simplified by assuming that the beta-delta model has the same u as the CCE model, we expect the non-equivalence to hold even if we have freedom to choose the utility index for the calibrating model.

6.2 Age-Decreasing Impatience

As explained in the Introduction, the dynamic CCE model (where each self has the same K) should embody a notion of decreasing impatience with respect to age. Intuitive as this maybe, it is not straightforward to formalize this because it is not clear how to compare two discount functions that are magnitude-dependent and have different horizons. We establish two expressions of age-decreasing impatience.

In the consumption-savings context, restrict attention only to Naive CCE – this is so as to remove the impact on saving of the agent's dynamic inconsistency. We show that the marginal propensity to save (the derivative of the saving rule with respect to current wealth) of each self along the optimal consumption path is consistent with age-decreasing impatience.

Proposition 9 *Consider the Naive CCE model with $K_0 = K_1$, assume $0 < \sigma\frac{m}{m-1} < 1$ and suppose that both self 0 and self 1 are cognitively constrained at their respective optimal consumption path. Then self 1's marginal propensity to save is greater than self 0's marginal propensity to save along the optimal consumption path.*

That is, regardless of how I_0 compares with I_1, as long as the assumptions are satisfied, self 1 always has a higher marginal propensity to save than self 0. Because self 1 is constrained, and since she has only one future period to consider, she must be spending all her cognitive resources on that period and thus her optimal $D^{self\ 1}(1)$ must be maximal. Although Self 0 is constrained as well, her $D^{self\ 0}(1)$ may not be at this maximum since she has to spend resources on $D^{self\ 1}(2)$ as well. The net result is a greater impatience as expressed by marginal propensity to save.

For a second expression of age-decreasing impatience, consider two agents born in different periods – in what follows t, T represent delays from their respective period zeros. Compare the discount functions of an older and younger agent with respective horizons $T-1$ and T. In order to control for differences due to magnitude dependence, it is necessary to assume that both agents face streams that are identical upto period $T-1$. In order to avoid comparing discount functions with different horizons, we will in fact suppose that the older agent also has a horizon T but her consumption in the last period is fixed at 0. So, take any T-horizon stream x that pays 0 in the terminal period T, and consider a stream $x + \epsilon^T$ that pays the same as x upto period $T-1$ but pays $\epsilon > 0$ in period T. The proposition below allow us to compare the discount function D_x of the old agent who faces x and the discount function $D_{x+\epsilon^T}$ of the young agent who faces $x + \epsilon^T$, and establishes that

$$D_{x+\epsilon^T}(t) \leq D_x(t) \qquad \text{for all } 0 < t \leq T-1,$$

that is, the younger agent's discount function is dominated by the older agent's for all periods upto $T-1$. The proposition, however, is proved for more general pairs of streams – specifically, any $x \in X$ and any $\tau \leq T$.

Proposition 10 *For any static CCE model, stream $x \in X$ and ϵ^τ with $\tau > 0$,*

$$D_{x+\epsilon^\tau}(\tau) \geq D_x(\tau),$$

and

$$D_{x+\epsilon^\tau}(t) \leq D_x(t) \qquad \textit{for all } 0 < t \neq \tau.$$

The proposition states that improving a stream in one period τ increases impatience towards other periods. As shown in the proof, this increase leads to no change in the discount factor for other periods $t \neq \tau$ if the agent is unconstrained at $x + \epsilon^\tau$. But if she is constrained, then the resources used for $D_{x+\epsilon^\tau}(\tau)$ must be generated by diverting cognitive resources from $D_{x+\epsilon^\tau}(t)$ for $t \neq \tau$.

6.3 Task Completion Problem

Suppose that the horizon is $T + 1$ where $T + 1$ is an odd number. Suppose that decisions are to be made only in periods $t = 0, 2, 4, \cdots, T$. If the agent performs no task in period $t = 0, 2, \cdots, T$, then she receives $r > 0$ in that period and 0 in the following period $t + 1$. If she does a task in period $t = 0, 2, \cdots, T$, then she receives 0 in that period and $R > r$ in the following period $t + 1$.

Throughout this section we consider a Sophisticated CCE model with $K_0 = K_1$.

6.3.1 Too Many Desirable Tasks

Suppose there are n identical tasks available to the agent to be completed in any order. There is a single deadline – period T – such that only tasks completed by T yield any reward. At most one task can be completed at a time – therefore when there are $n > 0$ tasks, at least $2n - 2$ periods are required to complete them. It is possible to complete (and reap the rewards of) just a subset of them.

Proposition 11 *Suppose the agent would complete one task at $t = 0$ if there was one available to be done at $T = 0$. Then for any $n > 0$ and a deadline of $T = 2n - 2$, the CE agent would complete all n tasks. The CCE agent may complete less than n tasks, and in fact may cycle between activity and inactivity.*

The intuition is as follows. The CE agent has separable preferences. If a task is attractive today, then it remains attractive no matter how many tasks are completed in the future. Consequently, an induction argument allows us to show that the agent would do as many desirable tasks as are available. The CCE agent, on the other hand, violates Separability. In particular the attractiveness of the reward of today's task is impacted by how many rewards the agent is anticipating from future tasks - the presence of future rewards reduce the cognitive resources available to appreciate the reward of the current task. The more such future rewards the agent is expecting, the less cognitive capacity the agent has to allocate to appreciating the reward of today's task. This is where the induction argument breaks down.

In the proof we construct an example of a CCE agent with three tasks to be completed by period $T = 4$. By assumption, the task is desirable enough that it will be completed in period $T = 4$ if there is one to be done. In period $t = 2$, the agent knows that the period $T = 4$ self will complete the task, and as noted above, due to limited cognitive capacity, she may not be able to appreciate the rewards of doing a task today, and therefore does not do the task. One might expect the same consideration to arise in period $t = 0$, but interestingly, it may not. The reason is that the cost of thinking φ_5 about the reward R in period 5 (received due to self $T = 4$'s effort) are higher than the cost of thinking φ_3 and so the $t = 0$ self does not direct as many resources to period 5. That leaves more cognitive resources available to appreciate the reward from current effort.

6.3.2 Too Much Flexibility

Suppose now that there is only one task to be done and that the deadline is lax. We show that if the task is desirable, the CE agent would complete it immediately regardless of the deadline. However, for constrained agents, there is such a thing as giving them too much time to complete the task, in that beyond a certain point a longer deadline may prompt them to procrastinate on doing the task.

Proposition 12 *Suppose that there is only one task, and that the agent would complete it at $t = 0$ if if the deadline is $T = 2$. Then the CE agent would complete the task at $t = 0$ for any longer deadline $T \geq 2$. For the CCE agent there may exist $T > 2$ such that the agent would not do the task immediately.*

The intuition is similar to the case of too many tasks. We have assumed that within the deadline, not doing the task leads the agent to receive a small immediate benefit. When the deadline is long then there are more such small benefits to be accrued, and these detract from the benefit of doing the task.

7 Avenues for Future Research

This paper provides a first step towards understanding constrained cognitive optimization in time preference, and there is much room for future research to extend the model to study richer cognitive models of intertemporal choice. For instance, studies show that poverty is correlated with lower cognitive abilities (Mani et al [17]). This suggests that current consumption may impact cognitive capacity: tolerating low consumption may use up some of the resources that would otherwise be used to think about the future. This can be captured in an extension of the CCE model where K_{x_0} is increasing in current consumption. An interesting implication of such a model is that the poor may engage in myopic consumption (e.g. overspending on alcohol, festivals and underspending on food and education – Banerjee and Duflo [2]) *because* it helps bolster cognitive resources, and not because low cognitive resources makes them myopic. Other studies (such as Read and Scholten [25]) suggest that adding a future reward to a stream may reduce impatience by attracting attention towards the future. Attention can possibly be modelled as the enhancement of cognitive capacity available for a stream, something that can be captured by an extension of the CCE model where K_x depends on the stream in some appropriate manner.

Cognitive models may also be developed to relate impatience and self-awareness. It is very natural to expect that one's degree of sophistication about future behavior may be determined by costly cognitive effort. Thinking about future preferences may yield a belief over possible future preferences, and higher cognitive effort may sharpen these beliefs (see Noor and Takeoka [22] for a model of optimal cognitive uncertainty). If the cost of cognitive effort in improving sophistication is drawn from the same cognitive stock K that generates empathy for future selves then there may be interesting trade-offs for behavior.

Finally, in this paper we have interpreted the vector $x = (x_0, \cdots, x_T)$ as a consumption stream, but it can also be interpreted as a vector of attributes in a deterministic choice setting, or as an Anscombe-Aumann act in a subjective uncertainty setting with states $s = 1, \cdots, T$, or a probability vector on fixed outcomes $t = 0, \cdots, T$ in a risk setting. Thus, our focus on time preference not withstanding, this paper may also serve as a starting point for thinking about the role of cognitive constraints in other choice domains that are of central economic interest.

A Appendix: Proof of Proposition 1

Given the first order condition $u(x_t) = \varphi_t'(D_x(t))$ on X_s, since the model requires $\varphi_t'(0) = 0$, it must be that if $u(x_t) = 0$ then we have the solution $D_x(t) = 0$. Since φ is strictly convex, it must be that the solution $D_x(t)$ is strictly increasing in $u(x_t)$. Given that u is linear, the FOC satisfies $\alpha u(x_t) = u(\alpha \circ x_t) = \varphi_t'(D_{\alpha \circ x}(t))$ for every $t > 0$ and so we have $D_{\alpha \circ x}(t) \leq D_x(t)$ for any $\alpha \in (0,1)$, with strict inequality for all $t > 0$ s.t. $u(x_t) > 0$.

To establish the second claim observe that when the constraint is binding, it must be that the optimal D_x satisfies

$$\frac{u(x_t)}{u(x_{t'})} = \frac{\varphi_t'(D_x(t))}{\varphi_t'(D_x(t'))} \; \forall t, t' > 0, \; and \; \sum_{t \geq 1} \varphi_t(D_x(t)) = K.$$

Since $u(\alpha \circ x) = \alpha u(x)$, it follows that for any $\alpha \in (0,1)$ we have $\frac{\varphi_t'(D_{\alpha \circ x}(t))}{\varphi_t'(D_{\alpha \circ x}(t'))} = \frac{u(\alpha \circ x_t)}{u(\alpha \circ x_{t'})} = \frac{u(x_t)}{u(x_{t'})} = \frac{\varphi_t'(D_x(t))}{\varphi_t'(D_x(t'))}$. Note that since φ_t' is strictly increasing, these ratios imply that if $D_{\alpha \circ x}(t) >$ (resp $<$)$D_x(t)$ for some $t > 0$ then $D_{\alpha \circ x}(t') >$ (resp $<$)$D_x(t')$ for all $t > 0$. It can never be that $D_{\alpha \circ x}(t) > D_x(t)$ for all t since we obtain the contradiction that $\varphi_t(D_{\alpha \circ x}(t)) > \sum_{t \geq 1} \varphi_t(D_x(t)) = K$. If it is the case that $D_{\alpha_n \circ x}(t) < D_x(t)$ for all t for some sequence $\alpha_n \in (0,1)$ converging to 1, then we have $\varphi_t(D_{\alpha_n \circ x}(t)) < \sum_{t\geq 1} \varphi_t(D_x(t)) = K$ and consequently the constraint is slack for all $\alpha_n \circ x$. But then $\alpha_n \circ x \in X_s$ for all n, implying $x \in X_s$, a contradiction. Conclude that $D_{\alpha\circ x}(t) = D_x(t)$ for all α sufficiently close to 1.

B Appendix: Proof of Proposition 2

Begin with an observation about any $x \in X_s$. Since the constraint is not binding on X_s and since the optimization problem (3) is separable in t, the discount factor $D_x(t)$ is determined separately for each t. Indeed, each $D_x(t)$ can be written as function $D_{u(x_t)}(t)$ of consumption x_t alone. Therefore U can be written as some sum $U(x) = u(x_0) + \sum_{t\geq 1} U_t(x_t)$ on X_s.

To prove the proposition, take $x \in X_s$. Clearly, a reduction in the magnitude of any outcome will keep the stream in X_s. Therefore $U(x_t, 0_{-t}) = D_{u(x_t)}(t)u(x_t)$ and $U(0_t, x_{-t}) = \sum_{t'\neq t} D_{u(x_{t'})}(t')u(x_{t'})$ and we obtain that $U(x) + U(0) = U(x) = U(x_t, 0_{-t}) + U(0_t, x_{-t})$.

Next take $x \notin X_s$. Consider the discounted utilities:

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

$$U(x_t, 0_{-t}) = D_{(x_t,0_{-t})}(t)u(x_t)$$

$$U(0_t, x_{-t}) = \sum_{t'\neq t} D_{(0_t,x_{-t})}(t')u(x_{t'}).$$

The first order condition for D_x requires that for all t, t' where $u(x_t), u(x_{t'}) > 0$,

$$\frac{u(x_t)}{u(x_{t'})} = \frac{D_x(t)}{D_x(t')}.$$

and since $D_x(t) = 0$ whenever $u(x_t) = 0$, we can write the binding constraint as $\sum_{t:u(x_t)>0} \varphi(D_x(t)) = K$. The discount functions for the streams $(x_t, 0_{-t}), (0_t, x_{-t})$ satisfy the same displayed equality of ratios but there are fewer periods with $u(x_t) > 0$ and consequently the capacity constraint applies to fewer discount factors. It follows that $D_{(x_t,0_{-t})}(t) \geq D_x(t)$ and $D_{(0_t,x_{-t})}(t') \geq D_x(t')$ for all $t' \neq t$. Consequently $U(x) + U(0) = U(x) \leq U(x_t, 0_{-t}) + U(0_t, x_{-t})$.

C Appendix: Proof of Proposition 3

(a) For any $\alpha \in (0,1)$, define the discount function $D^{un}_{\alpha\circ x}$ that satisfies the FOC $\alpha u(x_t) = \varphi'_t(D^{un}_{\alpha\circ x}(t))$ for each t. Since φ'_t is continuous and $\varphi'_t(0) = 0$ it follows that $D^{un}_{\alpha\circ x} \to 0$ as $\alpha \to 0$. Since $K > 0$ and φ_t is continuous and satisfies $\sum_{t\geq 1} \varphi_t(0) = 0$, it follows that there exists α^* s.t. $\sum_{t\geq 1} \varphi_t(D^{un}_{\alpha^*\circ x}(t)) < K$. Thus, $D^{un}_{\alpha^*\circ x}$ satisfies the FOC and the capacity constraint is slack at $D^{un}_{\alpha^*\circ x}$. It follows that $D_{\alpha^*\circ x} = D^{un}_{\alpha^*\circ x}$ is the solution of the optimization problem for $\alpha^* \circ x$ and indeed, $\alpha^* \circ x \in X_s$. This completes the proof.

(b) Take any $x \in X_s$ where $u(x_t) > 0$ for some $t > 0$. The solution $D_x(t)$ satisfies the FOC. Take any $y \lneq x$ and define the discount function D^{un}_y that satisfies the FOC. Since φ' is strictly increasing, we must have $D^{un}_y \lneq D_x$. Since φ is strictly increasing, we consequently obtain $\sum_{t\geq 1} \varphi_t(D^{un}_y(t)) < \sum_{t\geq 1} \varphi_t(D_x(t)) \leq K$. Thus D^{un}_y satisfies the FOC and the capacity constraint is slack at D^{un}_y. It follows that $D_y = D^{un}_y$ is the solution of the optimization problem for y and indeed, $y \in X_s$.

D Appendix: Proof of Theorem 1

The necessity of the axioms is relegated to the supplementary appendix (Noor and Takeoka [21]). Below we show sufficiency. We proceed in steps. Denoted the set of all magnitude sensitive and separable streams by $X_{ms} \subset X$.

D.1 Additively Separable Utility Representation on X_{ms}

Lemma 1 *The preference $\succsim|_{\Delta_0}$ is represented by a utility function $u : \Delta \to \mathbb{R}_+$ with $u(0) = 0$ which is continuous, mixture linear, homogeneous (that is, $u(\alpha \circ p) = \alpha u(p)$ for all $\alpha \geq 0$), and the restriction of u on C is strictly increasing. Moreover, the preference $\succsim$ on X is represented by a continuous utility function $U : X \to \mathbb{R}_+$ such that $U(p) = u(p)$ for all $p \in \Delta_0$.*

Proof. By Regularity, $\succsim|_{\Delta_0}$ satisfies the vNM axioms. There exists a continuous mixture linear function $u : \Delta \to \mathbb{R}_+$ which represents $\succsim|_{\Delta_0}$ and which can be chosen so that $u(0) = 0$. Moreover, C-Monotonicity implies that the restriction of u on C is strictly increasing.

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

$$u(\alpha \circ p) = u(\alpha \circ p + (1 - \alpha) \circ 0) = \alpha u(p) + (1 - \alpha)u(0) = \alpha u(p).$$

If $\alpha > 1$, we identify $\alpha \circ p$ with $p' \in \Delta$ satisfying $p = \frac{1}{\alpha} \circ p' + \frac{\alpha-1}{\alpha} \circ 0$. Then, mixture linearity of u implies that $u(p) = \frac{1}{\alpha} u(p')$, that is, $u(\alpha \circ p) = u(p') = \alpha u(p)$, 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 $\succsim$. Moreover, for all $p \in \Delta$, $U(p) = u(p)$. In particular, we have $U(0) = u(0) = 0$.

To show the continuity of U, take any sequence $x^n \to \widehat{x}$. There exists a corresponding present equivalent $c_{x^n} \sim x^n$. Since $U(x^n) = u(c_{x^n})$ and u is continuous, we want to show that $c_{x^n} \to c_{\widehat{x}}$.

Claim 1 *The present equivalent is continuous, that is, if $x^n \to x$, then $c_{x^n} \to c_{\widehat{x}}$.*

Proof. Take any $\overline{c}$ and $\underline{c}$ such that $\overline{c} > c_{\widehat{x}} > \underline{c}$. Let $W = \{x \in X \,|\, \overline{c} \succ x \succ \underline{c}\}$. Since $x^n \to \widehat{x} \sim c_{\widehat{x}}$, by Continuity, we can assume $x^n \in W$ for all n without loss of generality.

Seeking a contradiction, suppose $c_{x^n} \not\to c_{\widehat{x}}$. Then, there exists a neighborhood of $c_{\widehat{x}}$, denoted by $B(c_{\widehat{x}})$, such that $c_{x^m} \notin B(c_{\widehat{x}})$ for infinitely many m. Let $\{x^m\}$ denote the corresponding subsequence of $\{x^n\}$. Since $x^n \to \widehat{x}$, $\{x^m\}$ also converges to $\widehat{x}$. Without loss of generality, we can assume $x^m \in W$, that is, $\overline{c} \succ x^m \sim c_{x^m} \succ \underline{c}$. By C-Monotonicity, $\overline{c} > c_{x^m} > \underline{c}$. Thus, $\{c_{x^m}\}$ belongs to a compact interval $[\underline{c}, \overline{c}]$, and hence, there exists a convergent subsequence $\{c_{x^\ell}\}$ with a limit $\widetilde{c} \neq c_{\widehat{x}}$. On the other hand, since $x^\ell \to \widehat{x}$ and $x^\ell \sim c_{x^\ell}$, Continuity implies $\widehat{x} \sim \widetilde{c}$. Since $c_{\widehat{x}}$ is unique, $c_{\widehat{x}} = \widetilde{c}$, which is a contradiction. ■
■

For each $t \geq 1$, let

$$\Delta_t = \{p \in \Delta \,|\, p^t \in X_{ms}\}.$$

Lemma 2 *On the subdomain $X_{ms} \cup \Delta_0 \subset X$, U can be written as an additively separable utility form, i.e. $U : X_{ms} \cup \Delta_0 \to \mathbb{R}_+$ s.t. for all $x \in X_{ms} \cup \Delta_0$,*

$$U(x) = u(x_0) + \sum_{t \geq 1} U_t(x_t),$$

where u is given as in Lemma 1 and $U_t : \Delta_t \to \mathbb{R}$ are continuous with $U_t(0) = 0$ for each t. Moreover, u is unbounded from above.

Proof. Take any $x \in X_{ms}$, which is denoted by $x = (x_0, x_1, \cdots, x_T)$. There exists some $t > 0$ with $x_t \succ 0$. We start with the case where there are two $x_t, x_s \succ 0$. By notational convenience, denote such a stream by $(x_t, x_s, 0_{-t,-s})$. Since this stream is separable,

$$\frac{1}{2} \circ c_{(x_s, 0_{-s})} + \frac{1}{2} \circ c_{(x_t, 0_{-t})} \sim \frac{1}{2} \circ c_{(x_t, x_s, 0_{-t,-s})} + \frac{1}{2} \circ 0.$$

Since u is mixture linear,

$$\begin{aligned} & u(c_{(x_s,0_{-s})}) + u(c_{(x_t,0_{-t})}) = u(c_{(x_t,x_s,0_{-t,-s})}) + u(0) \\ \Longleftrightarrow\ & U(x_s, 0_{-s}) + U(x_t, 0_{-t}) = U(x_t, x_s, 0_{-t,-s}). \end{aligned}$$

Define $U_t(x_t) = U(x_t, 0_{-t})$ and $U_s(x_s) = U(x_s, 0_{-s})$. Then, we have

$$U(x_t, x_s, 0_{-t,-s}) = U_t(x_t) + U_s(x_s). \tag{10}$$

If a separable stream x has three outcomes $x_t, x_s, x_r \succ 0$, denote it by $x = (x_t, x_s, x_r, 0_{-t,-s,-r})$. By PBT, $(x_t, x_s, 0_{-t,-s}) \in X_{ms}$. From the above argument, we have (10). Since x is separabile,

$$\frac{1}{2} \circ c_{(x_r, 0_{-r})} + \frac{1}{2} \circ c_{(x_t, x_s, 0_{-t,-s})} \sim \frac{1}{2} \circ c_{(x_t, x_s, x_r, 0_{-t,-s,-r})} + \frac{1}{2} \circ 0.$$

Since u is mixture linear,

$$\begin{aligned} & u(c_{(x_r,0_{-r})}) + u(c_{(x_t,x_s,0_{-t,-s})}) = u(c_{(x_t,x_s,x_r,0_{-t,-s,-r})}) + u(0) \\ \Longleftrightarrow\ & U(x_r, 0_{-r}) + U(x_t, x_s, 0_{-t,-s}) = U(x_t, x_s, x_r, 0_{-t,-s,-r}). \end{aligned}$$

Define $U_r(x_r) = U(x_r, 0_{-r})$. Then, we have

$$\begin{aligned} U(x_t, x_s, x_r, 0_{-t,-s,-r}) &= U_r(x_r) + U(x_t, x_s, 0_{-t,-s}) \\ &= U_t(x_t) + U_s(x_s) + U_r(x_r). \end{aligned}$$

By repeating the same argument finitely many times, we have

$$U(x) = \sum_{t \geq 0} U_t(x_t),$$

where $U_t(x_t)$ is defined as $U_t(x_t) = U(x_t, 0_{-t})$. By definition, $U_t(0) = 0$. By PBT, for any $x \in X_{ms}$, if $x_t \succ 0$ then $(x_t)^t \in X_{ms}$, that is, $(x_t)^t \in \Delta_t$. Hence, U_t is defined on Δ_t.

Since U is continuous, U_t is also continuous. Take any $p \in \Delta$ and any sequence $x^n = (0, x_1^n, \cdots, x_T^n) \in X_{ms}$, where $x_t^n \to 0$ for all $t \geq 1$. By PBT, $(p, x_{-0}^n) = (p, x_1^n, \cdots, x_T^n) \in X_{ms}$. Since $(p, x_{-0}^n) \to p \in \Delta_0$, by continuity, $U(p, x_{-0}^n) \to u(p)$ and $U(p, x_{-0}^n) = U_0(p) + \sum_{t \geq 1} U_t(x_t^n) \to U_0(p)$. Thus, $U_0(p) = u(p)$.

Finally, we show that u must be unbounded above. First, we show that u is unbounded above. By seeking a contradiction, suppose otherwise. Then, the range of u is nonempty and has an upper bound. There exists a supremum $\overline{v}$ of the range of u. By Monotonicity, U_t is non-constant, and hence, there exists some $\tilde{p} \in \Delta_t$ with $U_t(\tilde{p}) > 0$. Note that $\tilde{p}^t \in X_{ms}$. Take a lottery $\overline{p} \in \Delta$ such that $\overline{v} - u(\overline{p}) < U_t(\tilde{p})$. Consider the stream $\overline{x}$ which pays $\overline{p}$ in period 0, $\tilde{p}$ in period t, and zero otherwise. By PBT, $\overline{x} \in X_{ms}$. By the representation,

$$U(\overline{x}) = u(\overline{p}) + U_t(\tilde{p}) > \overline{v}.$$

Since $\overline{v}$ is the supremum of $u(\Delta)$, the above inequality contradicts the Present Equivalents axiom. ■

Lemma 3 *The function $U : X_{ms} \cup \Delta_0 \to \mathbb{R}_+$ defined as in Lemma 2 can be written as follows:*

$$U(x) = u(x_0) + \sum_{t \geq 1} D_{u(x_t)}(t) u(x_t),$$

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

Proof. Taking the additive representation from Lemma 2, by Monotonicity, 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 $D_{u(x_t)}(t) = \frac{U_t(u(x_t))}{u(x_t)} > 0$ for any $x_t \in \Delta$ with $x_t \succ 0$. Define $\underline{D}(t) = \inf \{D_{u(p)}(t) \,|\, 0 \prec p \in \Delta_t\}$. Then

$$U(x) = u(x_0) + \sum_{t \geq 1} D_{u(x_t)}(t) u(x_t), \text{ for all } x \in X_{ms} \cup \Delta_0.$$

To see that $D_{u(p)}(t)$ is strictly increasing in $u(p)$, note that for any stream $x \in X_{ms}$ and its present equivalent c_x, by definition of X_{ms}, $\alpha U(c_x) > U(\alpha \circ x)$ for all $\alpha \in (0,1)$ and thus $\alpha U(x) > U(\alpha \circ x)$. Applying this more specifically to a dated reward p^t with $u(p) > 0$ and exploiting mixture linearity of u, we obtain $\alpha D_{u(p)}(t)u(p) > D_{u(\alpha \circ p)}(t)u(\alpha \circ p) = \alpha D_{\alpha u(p)}(t)u(p)$ and thus

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

as desired.

Since u and U_t are continuous, so is $D_{u(p)}(t)$ in $u(p)$ on the domain of $u(p) > 0$. Since $\underline{D}(t)$ is defined as $\inf\{D_{u(p)}(t) \,|\, 0 \prec p \in \Delta_t\}$ and $D_{u(p)}(t)$ is strictly increasing in $u(p)$, $D_{u(p)}(t)$ is indeed continuous for all $u(p) \geq 0$.

Take any $D_{u(p)}(t)$. There exists a corresponding dated reward $p^t \in X_{ms}$. By Impatience, $u(p) = U(p^0) \geq U(p^t) = D_{u(p)}(t)u(p)$, which implies $D_{u(p)}(t) \leq 1$. ■

Define

$$S_t = \{d \in [0,1] \,|\, d = D_{u(p)}(t) \text{ for some } p^t \in X_{ms}\}.$$

By Monotinicity and PBT (ii), if $p^t \in X_{ms}$, then $\alpha \circ p^t \in X_{ms}$ for all $\alpha \in (0,1)$. Thus, S_t is an interval. Note $\underline{D}(t) = \inf S_t$. Denote $\overline{D}(t) = \sup S_t$.

Lemma 4 *The function $U : X_{ms} \cup \Delta_0 \to \mathbb{R}_+$ appeared in Lemma 3 can be written as follows:*

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

$$s.t. \qquad D_x = \arg\max_D \{\sum_{t\geq 1} \big(D(t)u(x_t) - \varphi_t(D(t))\big)\}$$

where for each $t \geq 1$, $\varphi_t : [0, \overline{D}(t)] \to \mathbb{R}_+$ is an increasing convex function that is strictly increasing, strictly convex, and continuously differentiable on $[\underline{D}(t), \overline{D}(t)]$, and satisfies $\varphi_t(\underline{D}(t)) = 0$ and $\varphi_t'(\underline{D}(t)) = 0$.

Proof. By Monotonicity and PBT (ii), if $x \in X_{ms}$, then $(x_t)^t \in X_{ms}$ for $x_t \succ 0$. Thus, φ_t can be derived from the dated rewards at t as follows. The marginal cost function φ_t' on S_t is implicitly defined by the first order condition

$$u(p) = \varphi_t'(D_{u(p)}(t)). \tag{11}$$

Since $D_{u(p)}(t)$ is strictly increasing and continuous in $u(p)$, (11) implies that φ_t' is strictly increasing and continuous. Then, we can monotonically and continuously extend φ_t' to $S_t \cup \{\underline{D}(t), \overline{D}(t)\}$ by defining $\varphi_t'(\overline{D}(t)) = \lim_{d \to \overline{D}(t)} \varphi_t'(d)$ and $\varphi_t'(\underline{D}(t)) = \lim_{d \to \underline{D}(t)} \varphi_t'(d)$. Moreover, the continuity of $D_{u(p)}(t)$ wrt $u(p)$ requires that $0 = \varphi_t'(\underline{D}(t))$. We further extend φ_t' to $[0, \overline{D}(t)]$ by setting $\varphi_t'(d) = 0$ for all $d \in [0, \underline{D}(t)]$.

We claim that $\varphi_t'(\overline{D}(t)) < \infty$. Suppose otherwise. Then, for any $r > 0$, there exists some $p^t \in X_{mx}$ such that $u(p) = r = \varphi_t'(D_{u(p)}(t))$. Since $D_r(t)$ is increasing on S_t, $D_r(t)r$

diverges to infinity. This means that the set $V(t) = \{U(p^t) \in \mathbb{R}_{++} \,|\, p^t \in X_{ms}\}$ is unbounded above. On the other hand, by assumption, there exists some $y \notin X_{ms} \cup \Delta_0$. Since $V(t)$ is unbounded above, there exists some $p^t \in X_{ms}$ such that $U(p^t) > U(0, y_{-0})$, then by PBT (ii), we must have $y \in X_{ms}$, which is a contradiction.

By the fundamental theorem of calculus, φ_t is obtained on $[0, \overline{D}(t)]$ as the integral of φ_t' along with the assumption that $\varphi_t(\underline{D}(t)) = 0$. Then, φ_t is strictly convex on $[\underline{D}(t), \overline{D}(t)]$. The function is by construction continuously differentiable and has a positive slope. By construction, the set $\arg\max_D\{\sum\big(D(t)u(x_t) - \varphi_t(D(t))\big)\}$ is nonempty and moreover, it is a singleton since $\sum\big(D(t)u(x_t) - \varphi_t(D(t))\big)$ is a strictly concave function of D. Thus D_x is a unique solution. ■

D.2 Extension to X

Recall that X_{ms} is the set of all magnitive sensitive and separable streams. Note that any $x \notin X_{ms}$ is magnitude insensitive. Indeed, if otherwise, x is magnitude sensitive, but then by MS Separability, x is separable. Since x is magnitude sensitive and separable, we have $x \in X_{ms}$, which is a contradiction.

Lemma 5 *For any stream $x \in X \setminus \Delta_0$, there exists a unique $\alpha_x \in (0, 1]$ such that*

$$\begin{cases} \alpha \leq \alpha_x \implies \alpha \circ x \in X_{ms}, \\ \alpha > \alpha_x \implies \alpha \circ x \notin X_{ms}. \end{cases}$$

Proof. Let $A = \{\alpha \in (0, 1] \,|\, \alpha \circ x \in X_{ms}\}$. By part (i) of PBT, $A \neq \emptyset$. Let $\alpha_x = \sup A$. We claim that A is an interval with $\inf A = 0$. Take any $\alpha \in A$ and $\beta \in (0, \alpha)$. Since $\alpha \circ x \in X_{ms}$, by part (ii) of PBT, $\beta \circ x = \frac{\beta}{\alpha} \circ (\alpha \circ x) \in X_{ms}$, that is, $\beta \in A$ as desired. Now, by definition of α_x, if $\alpha < \alpha_x$, then $\alpha \in A$, and hence $\alpha \circ x \in X_{ms}$. If $\alpha > \alpha_x$, then $\alpha \notin A$, and hence $\alpha \circ x \notin X_{ms}$. Uniqueness of α_x is obvious. Moreover, if $x \in X_{ms}$, by part (ii) of PBT, $A = (0, 1)$, and hence, $\alpha_x = 1$. ■

Lemma 6 *For any $x \in X \setminus \Delta_0$, take $\alpha_x \in (0, 1]$ which is defined as in Lemma 5. Then,*

$$\begin{cases} \alpha < \alpha_x \implies \alpha \circ c_x \succ \alpha \circ x, \\ \alpha \geq \alpha_x \implies \alpha \circ c_x \sim \alpha \circ x. \end{cases}$$

Proof. Step 1: For all $x \in X \setminus \Delta_0$, $\alpha \circ c_x \succ \alpha \circ x$ implies $\beta \circ c_x \succ \beta \circ x$ for all $\beta \in (0, \alpha]$. By definition, a present equivalent of $\alpha \circ x$, denoted by $c_{\alpha \circ x}$, satisfies $\alpha \circ c_x \succ \alpha \circ x \sim c_{\alpha \circ x}$. For any $\gamma \in (0, 1)$, let $\beta = \gamma\alpha \in (0, \alpha)$. By Weak Homotheticity and Risk Preference,

$$\beta \circ c_x = \gamma\alpha \circ c_x \succ \gamma \circ c_{\alpha \circ x} \succsim \gamma\alpha \circ x = \beta \circ x,$$

as desired.

Step 2: If there exist $\alpha, \beta \in (0, 1)$ such that $\alpha \circ c_x \sim \alpha x$ and $\beta \circ c_{\alpha \circ x} \sim \beta \circ (\alpha \circ x)$, then $\alpha\beta \circ c_x \sim \alpha\beta \circ x$. By definition and the assumption, $\alpha \circ c_x \sim \alpha \circ x \sim c_{\alpha \circ x}$. By Risk Preference, $\alpha\beta \circ c_x \sim \beta \circ c_{\alpha \circ x}$. Hence, by assumption, $\alpha\beta \circ c_x \sim \alpha\beta \circ x$.

Step 3: There exists a unique $\widetilde{\alpha}_x \in (0,1]$ such that

$$\begin{cases} \alpha < \widetilde{\alpha}_x \implies \alpha \circ c_x \succ \alpha \circ x, \\ \alpha \geq \widetilde{\alpha}_x \implies \alpha \circ c_x \sim \alpha \circ x. \end{cases}$$

If $x \in X_{ms}$, $\widetilde{\alpha}_x = 1$ satisfies this condition. Thus, assume $x \notin X_{ms}$. Let $\widetilde{A} = \{\alpha \in (0,1] \mid \alpha \circ c_x \succ \alpha \circ x\}$. By part (i) of PBT, $\widetilde{A}$ is non-empty. Moreover, by Step 1, $\widetilde{A}$ is an interval with $\inf \widetilde{A} = 0$. Let $\widetilde{\alpha}_x$ be a supremum of $\widetilde{A}$. If $\widetilde{A} = (0,1)$, $\widetilde{\alpha}_x = 1$ and this $\widetilde{\alpha}_x$ satisfies the desired property. If $\widetilde{A}$ is a proper subset of $(0,1)$, $\widetilde{\alpha}_x < 1$. Then, there exists a sequence $\alpha^n \to \widetilde{\alpha}_x$ with $\alpha^n > \widetilde{\alpha}_x$. Since $\alpha^n \circ c_x \sim \alpha^n \circ x$, by Continuity, $\widetilde{\alpha}_x \circ c_x \sim \widetilde{\alpha}_x \circ x$, as desired.

Step 4: $\widetilde{\alpha}_x \leq \alpha_x$. Seeking a contradiction, suppose $\widetilde{\alpha}_x > \alpha_x$. Lemma 5 implies $\widetilde{\alpha}_x \circ x \notin X_{ms}$. Since this stream is magnitude insensitive, there exists $\beta \in (0,1)$ such that $\beta \circ c_{\widetilde{\alpha}_x \circ x} \sim \beta \circ (\widetilde{\alpha}_x \circ x)$. Since $\widetilde{\alpha}_x \circ c_x \sim \widetilde{\alpha}_x \circ x$, by Step 2, $\widetilde{\alpha}_x \beta \circ c_x \sim \widetilde{\alpha}_x \beta \circ x$. Since $\widetilde{\alpha}_x \beta < \widetilde{\alpha}_x$, this contradicts to Step 3.

Step 5: $\widetilde{\alpha}_x = \alpha_x$. By Step 4, seeking a contradiction, suppose $\widetilde{\alpha}_x < \alpha_x$. Take any $\alpha \in (\widetilde{\alpha}_x, \alpha_x)$. By Step 3, $\alpha \circ c_x \sim \alpha \circ x$. Moreover, for all γ sufficiently close to one, since $\gamma\alpha \in (\widetilde{\alpha}_x, \alpha_x)$, $\gamma\alpha \circ c_x \sim \gamma\alpha \circ x$. Now, by definition, $c_{\alpha \circ x} \sim \alpha \circ x$, which implies $c_{\alpha \circ x} \sim \alpha \circ c_x$. Since $\alpha \circ x \in X_{ms}$ by Lemma 5, for all $\gamma \in (0,1)$, $\gamma \circ c_{\alpha \circ x} \succ \gamma\alpha \circ x$. Thus, we have

$$\gamma \circ c_{\alpha \circ x} \succ \gamma\alpha \circ x \sim \gamma\alpha \circ c_x$$

for all γ sufficiently close to one. By Risk Preference, $c_{\alpha \circ x} \succ \alpha \circ c_x$, which is a contradiction. ■

Lemma 7 *For all $x, y \in X \setminus \Delta_0$, take $\alpha_x, \alpha_y \in (0,1]$ which are defined as in Lemma 5. If $x_t \sim y_t$ for all $t \geq 1$, then $\alpha_x = \alpha_y$.*

Proof. Since $u(x_t) = u(y_t)$ for all $t \geq 1$, Monotonicity implies $(0, x_{-0}) \sim (0, y_{-0})$. By PBT (ii), $(0, x_{-0}) \in X_{ms}$ if and only if $(0, y_{-0}) \in X_{ms}$. Again, by PBT (ii), $x \in X_{ms}$ if and only if $(0, x_{-0}) \in X_{ms}$. Hence, $x \in X_{ms}$ if and only if $y \in X_{ms}$. Now, if $x, y \in X_{ms}$, we obtain $\alpha_x = \alpha_y = 1$. Next assume that $x \notin X_{ms}$ and $y \notin X_{ms}$. Seeking a contradiction, suppose $\alpha_x \neq \alpha_y$. Without loss of generality, let $\alpha_x > \alpha_y$. For any $\alpha \in (\alpha_y, \alpha_x)$, by Lemma 5, $\alpha \circ x \in X_{ms}$ and $\alpha \circ y \notin X_{ms}$. But, since $u(\alpha \circ x_t) = u(\alpha \circ y_t)$ for all $t \geq 1$, by the same argument as above, we have $\alpha \circ x \in X_{ms}$ if and only if $\alpha \circ y \in X_{ms}$, which is a contradiction. Thus, we have $\alpha_x = \alpha_y$, as desired. ■

As shown in Lemma 5, for any $x \in X \setminus \Delta_0$,

$$\alpha_x = \sup\{\alpha \in [0,1] \mid \alpha \circ x \in X_{ms}\}.$$

Lemma 8 *The function $U : X \to \mathbb{R}_+$ appeared in Section D.1 can be written as*

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

$$s.t. \qquad D_x = \begin{cases} \arg\max_D \{\sum_{t\geq 1} D(t)u(x_t) - \varphi_t(D(t))\} & \text{if } x \in X_{ms} \cup \Delta_0, \\ D_{\alpha_x \circ x} & \text{if } x \notin X_{ms} \cup \Delta_0. \end{cases}$$

Proof. By the result of Section D.1, U has the desired form on $X_{ms} \cup \Delta_0$. Consider the case of $x \notin X_{ms} \cup \Delta_0$. Since $u(\alpha_x \circ c_x) = U(\alpha_x \circ x)$ by Lemma 6,

$$U(x) = u(c_x) = \frac{1}{\alpha_x} U(\alpha_x \circ x). \tag{12}$$

By the representation on X_{ms},

$$U(\alpha_x \circ x) = u(\alpha_x \circ x_0) + \sum_{t\geq 1} D_{\alpha_x \circ x}(t) u(\alpha_x \circ x_t). \tag{13}$$

By combining (12) with (13),

$$\begin{aligned} U(x) = \frac{1}{\alpha_x} U(\alpha_x \circ x) &= \frac{1}{\alpha_x} \left(u(\alpha_x \circ x_0) + \sum_{t\geq 1} D_{\alpha_x \circ x}(t) u(\alpha_x \circ x_t) \right) \\ &= u(x_0) + \sum_{t\geq 1} D_{\alpha_x \circ x}(t) u(x_t), \end{aligned}$$

as desired. ■

From now on, we derive a function $K : X \setminus \Delta_0 \to \mathbb{R}_{++}$ which serves as a capacity constraint for the CCE* representation. Let

$$\varphi(D) := \sum_{t\geq 1} \varphi_t(D(t)).$$

Lemma 9 *There is a function $K : X \setminus \Delta_0 \to \mathbb{R}_{++}$ such that $\succsim$ is represented by*

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

$$s.t. \qquad D_x = \arg\max_{D \in \Lambda_x} \{\sum D(t)u(x_t) - \varphi_t(D(t))\}$$

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

Moreover, (1) the function K_x satisfies $K_x = K_{\lambda\circ x}$ for any x and λ, and (2) for all streams x, y, if $u(x_t) = u(y_t)$ for all $t \geq 1$, then $K_x = K_y$.

Proof. Since $U(p) = u(p)$ for all $p \in \Delta_0$, K does not play any role for consumption stream on Δ_0.

By assumption, there exists $\bar{x} \notin X_{ms} \cup \Delta_0$. By Lemma 5, there exists $\alpha_{\bar{x}} \in (0, 1)$ such that $\alpha_{\bar{x}} \circ x \in X_{ms}$ and $\alpha \circ x \notin X_{ms}$ for all $\alpha > \alpha_{\bar{x}}$. Let $X^{\alpha_{\bar{x}}} = \{y \in X \,|\, y_t \succ \alpha_{\bar{x}} \circ \bar{x}_t, \ \forall t\}$. If $X^{\alpha_{\bar{x}}} \cap X_{ms} \neq \emptyset$, there exists $y \in X^{\alpha_{\bar{x}}} \cap X_{ms}$. By Monotonicity, $y_t \succsim \alpha \circ \bar{x}_t$ for all t for all

$\alpha > \alpha_{\bar{x}}$ sufficiently close to $\alpha_{\bar{x}}$. But, then, PBT implies $\alpha \circ \bar{x} \in X_{ms}$ for such α, which is a contradiction. Hence, we have $X^{\alpha_{\bar{x}}} \cap X_{ms} = \emptyset$, or $X_{ms} \subset X \setminus X^{\alpha_{\bar{x}}}$.

Now take any $y \in X \setminus \Delta_0$. Let $\bar{x}$ be the stream fixed in the above argument. For sufficiently large $\lambda > 0$, we have $\lambda \circ y_t \succ \alpha_{\bar{x}} \circ \bar{x}_t$ for all t, that is, $\lambda \circ y \in X^{\alpha_{\bar{x}}}$. Together with the above observation, $\lambda \circ y \notin X_{ms}$. Let x denote such $\lambda \circ y$. That is, we find $x \notin X_{ms}$ on the same ray of y. For such x, define

$$K_x := \varphi(D_{\alpha_x \circ x}) < \infty.$$

Extend to X_{ms} by requiring $K_x = K_{\lambda \circ x}$ for any $\lambda > 0$.

For all $x \in X \setminus \Delta_0$, by Lemma 5, there exists $\alpha_x > 0$ such that $\alpha_x \circ x \in X_{ms}$. For any $\beta \in (0, \alpha_x)$, since φ is strictly increasing and $D_{u(c)}(t)$ is strictly increasing in $u(c)$, $K_x = \varphi(D_{\alpha_x \circ x}) > \varphi(D_{\beta \circ x}) \geq 0$. Hence, $K_x > 0$.

For any $x \in X \setminus \Delta_0$, define

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

From Lemma 8, for any $x \in X_{ms}$ we have

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

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

There exists $x' \notin X_{ms} \cup \Delta_0$ such that $x = \alpha x'$ for some $\alpha \in (0,1)$. Since φ is strictly increasing and $D_{\alpha x'}$ is increasing in α up to $\alpha_{x'} \circ x'$, $\varphi(D_x) \leq \varphi(D_{\alpha_{x'} \circ x'}) = K_x$, that is, we have $D_x \in \Lambda_x$. Thus, D_x is also the unique maximizer in the constrained problem:

$$D_x = \arg\max_{D \in \Lambda_x} \{\sum D(t) u(x_t) - \varphi_t(D(t))\},$$

thereby establishing the result for $x \in X_{ms}$.

Next consider $x \notin X_{ms} \cup \Delta_0$, and take $\alpha_x \circ x \in X_{ms}$. By definition, note that $K_x < \infty$. By the preceding,

$$D_{\alpha_x \circ x} = \arg\max_{D \in \Lambda_x} \{\sum D(t) u(\alpha_x x_t) - \varphi_t(D(t))\}.$$

For notational simplicity, for any x, let $u(x)$ denote $(u(x_1), \cdots, u(x_T)) \in \mathbb{R}^T_+$. We first prove that

$$D_{\alpha_x \circ x} \in \arg\max_{D \in \Lambda_x} D \cdot u(x). \tag{14}$$

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

$$D \cdot u(\alpha_x \circ x) - \varphi(D) > D_{\alpha_x \circ x} \cdot u(\alpha_x \circ x) - \varphi(D_{\alpha_x \circ x}),$$

contradicting the optimality of $D_{\alpha_x x}$ for $\alpha_x \circ x$, as desired.

To conclude the proof of the lemma, observe that for any $D \in \Lambda_x$ with $D \neq D_{\alpha_x \circ x}$,

$$\begin{aligned}
& D_{\alpha_x \circ x} \cdot u(\alpha_x \circ x) - \varphi(D_{\alpha_x \circ x}) > D \cdot u(\alpha_x \circ x) - \varphi(D) \\
\Longrightarrow\ & D_{\alpha_x \circ x} \cdot u(\alpha_x \circ x) - D \cdot u(\alpha_x \circ x) > \varphi(D_{\alpha_x \circ x}) - \varphi(D) \\
\Longrightarrow\ & \alpha_x [D_{\alpha_x \circ x} \cdot u(x) - D \cdot u(x)] > \varphi(D_{\alpha_x \circ x}) - \varphi(D) \\
\Longrightarrow\ & D_{\alpha_x \circ x} \cdot u(x) - D \cdot u(x) > \varphi(D_{\alpha_x \circ x}) - \varphi(D) \\
& (\text{since } D_{\alpha_x \circ x} \cdot u(x) \geq D \cdot u(x), \text{ by (14)}) \\
\Longrightarrow\ & D_{\alpha_x \circ x} \cdot u(x) - \varphi(D_{\alpha_x \circ x}) > D \cdot u(x) - \varphi(D).
\end{aligned}$$

Thus,

$$D_{\alpha_x \circ x} = \arg \max_{D \in \Lambda_x} \{ \sum D(t) u(x_t) - \varphi_t(D(t)) \},$$

as desired.

By Lemma 7, if $u(x_t) = u(y_t)$ for all $t \geq 1$, $\alpha_x = \alpha_y$. 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 $K_x = K_y$. ■

Define

$$V_{ms} := \{ U(0, x_{-0}) \in \mathbb{R}_{++} \,|\, x \in X_{ms} \}.$$

We claim that V_{ms} is bounded above. Indeed, by assumption, there exists some $y \notin X_{ms} \cup \Delta_0$. If there exists $x \in X_{ms}$ with $U(0, x_{-0}) > U(0, y_{-0})$, then by PBT (ii), we must have $y \in X_{ms}$, which is a contradiction. Hence, for all $x \in X_{ms}$, $U(0, x_{-0}) \leq U(0, y_{-0})$. That is, V_{ms} is bounded above. Hence, there exists a finite $\overline{v} := \sup V_{ms} > 0$. Let

$$X_{\overline{v}} = \{ x \in X \setminus \Delta_0 \,|\, U(0, x_{-0}) \leq \overline{v} \}.$$

The following lemma states that X_{ms} is characterized as the lower contour set of some indifference curve.

Lemma 10 $X_{ms} = X_{\overline{v}}$.

Proof. $X_{ms} \subset X_{\overline{v}}$: Take any $x \notin X_{\overline{v}}$. By definition, $U(0, x_{-0}) > \overline{v}$. Then, we have $x \notin X_{ms}$ because $x \in X_{ms}$ violates the definition of $\overline{v}$.

$X_{\overline{v}} \subset X_{ms}$: Take any $x \in X_{\overline{v}}$ with $U(0, x_{-0}) < \overline{v}$. By definition of $\overline{v}$, there exists $y \in X_{ms}$ with $U(0, x_{-0}) \leq U(0, y_{-0})$. By part (ii) of PBT, $x \in X_{ms}$. Next, take $x \in X_{\overline{v}}$ with $U(0, x_{-0}) = \overline{v}$. For any $\alpha \in (0, 1)$, by Risk Preference and Monotonicity, $\alpha \circ x \prec x$, and hence, $U(0, \alpha \circ x_{-0}) < U(0, x_{-0}) = \overline{v}$, which implies $\alpha \circ x \in X_{\overline{v}}$. By the above argument, $\alpha \circ x \in X_{ms}$. By Lemma 5, $x \in X_{ms}$ as $\alpha \to 1$. ■

Define

$$bd^{+}(X_{ms}) = \{ x \in X_{ms} \,|\, \lambda \circ x \notin X_{ms} \text{ for all } \lambda > 1 \}.$$

Lemma 11 $bd^{+}(X_{ms}) = \{ x \in X_{\overline{v}} \,|\, U(0, x_{-0}) = \overline{v} \}$.

Proof. Take any $x \in bd^+(X_{ms})$. Since $x \in X_{ms} = X_{\overline{v}}$ (by Lemma 10), we have $U(0, x_{-0}) \leq \overline{v}$. Seeking a contradiction, suppose $U(0, x_{-0}) < \overline{v}$. By Continuity, there exists some $\lambda > 1$ such that $U(0, \lambda \circ x_{-0}) < \overline{v}$, which implies $\lambda \circ x \in X_{ms}$ by Lemma 10. However, by definition of $bd^+(X_{ms})$, $\lambda \circ x \notin X_{ms}$. This is a contradiction.

Conversely, take any $x \in X_{\overline{v}}$ satisfying $U(0, x_{-0}) = \overline{v}$. By Lemma 10, we know $x \in X_{ms}$. Seeking a contradiction, suppose $x \notin bd^+(X_{ms})$. Then, there exists some $\lambda > 1$ with $\lambda \circ x \in X_{ms}$. By Monotonicity, $U(0, \lambda \circ x_{-0}) > U(0, x_{-0}) = \overline{v}$. Lemma 10 implies $\lambda \circ x \notin X_{ms}$, a contradiction. ■

Lemma 12 *For any* x,

$$K_x = \sum_{t=1}^{T} \varphi_t\Big((\varphi_t')^{-1}(\lambda_x u(x_t))\Big),$$

and λ_x *is the unique solution to* $\sum_{t=1}^{T} \lambda u(x_t)(\varphi_t')^{-1}(\lambda u(x_t)) = \overline{v}$.

Proof. Take any $x \in X$. Since $(\varphi_t')^{-1}$ is strictly increasing, there exists a unique solution λ_x to $\sum_{t=1}^{T} \lambda u(x_t)[(\varphi_t')^{-1}(\lambda u(x_t))] = \overline{v}$. We observe that $U(0, \lambda_x \circ x) = \overline{v}$ since by the first order condition,

$$U(0, \lambda_x \circ x) = \sum_{t=1}^{T} D_{u(\lambda_x \circ x_t)}(t) u(\lambda_x \circ x_t) = \sum_{t=1}^{T} \lambda_x u(x_t)[(\varphi_t')^{-1}(\lambda_x u(x_t))] = \overline{v}.$$

Hence, by Lemma 11, $\lambda_x \circ x \in bd^+(X_{ms})$. By Lemma 9, we have $K_x = K_{\lambda_x \circ x}$. Thus,

$$K_x = K_{\lambda_x \circ x} = \sum_{t=1}^{T} \varphi_t\Big((\varphi_t')^{-1}(\lambda_x u(x_t))\Big),$$

as desired. ■

Lemma 13 $\overline{D}(t)\varphi_t'(\overline{D}(t)) = \overline{v}$ *for all* $t \geq 1$.

Proof. By definition of $\overline{D}(t)$, there exists some sequence $(p^n)^t \in X_{ms}$ such that $D_{u(p^n)}(t) \to \overline{D}(t)$. By the FOC, $u(p^n) = \varphi_t'(D_{u(p)}(t))$. Since $(p^n)^t \in X_{ms}$, by Lemma 10, $D_{u(p^n)}(t)u(p^n) \leq \overline{v}$, or equivalently, $D_{u(p^n)}(t)\varphi_t'(D_{u(p^n)}(t)) \leq \overline{v}$. Since $D_r(t)$ and $\varphi_t'(d)$ are continuous, we have $\overline{D}(t)\varphi_t'(\overline{D}(t)) \leq \overline{v}$ as $n \to \infty$. Conversely, since $\overline{v} < \infty$, there exists some $\bar{p}^t$ such that $U(\bar{p}^t) = \overline{v}$. By Lemma 10, $\bar{p}^t \in X_{ms}$, and hence, we have $(\varphi_t')^{-1}(u(\bar{p}))u(\bar{p}) = \overline{v}$. From the FOC, this condition is equivalent to $D_{u(\bar{p})}(t)\varphi_t'(D_{u(\bar{p})}(t)) = \overline{v}$. It follows from the definitions of $\varphi_t'(\overline{D}(t))$ and $\overline{D}(t)$ that $\overline{D}(t)\varphi_t'(\overline{D}(t)) \geq D_{u(\bar{p})}(t)\varphi_t'(D_{u(\bar{p})}(t)) = \overline{v}$. This establishes the statement of the lemma. ■

Recall that in Lemma 4, we constructed φ_t on $[0, \overline{D}(t)]$ for each t. We close the proof by extending φ_t to $[0, 1]$ in a manner required by regularity.

Lemma 14 *There exists an extension of* $\{\varphi_t\}$ *to* $[0, 1]$ *such that* $\varphi_t' \leq \varphi_{t+1}'$ *for all* $t < T$.

Proof. First, as a preliminary, we claim that $\overline{D}(t+1) \leq \overline{D}(t)$. Take $\bar{p}^t$ and $\bar{q}^{t+1}$ such that $u(\bar{p}) = \varphi_t'(\overline{D}(t))$ and $u(\bar{q}) = \varphi_{t+1}'(\overline{D}(t+1))$. By Impatieice, $\bar{p}^t \succsim \bar{p}^{t+1}$, which implies $\bar{p}^{t+1} \in X_{ms}$ from PBT (ii), and hence, $u(\bar{p}) = \varphi_{t+1}'(D_{u(\bar{p})}(t+1))$. Since φ_{t+1}' is strictly increasing,

$$u(\bar{p}) = \varphi_{t+1}'(D_{u(\bar{p})}(t+1)) \leq \varphi_{t+1}'(\overline{D}(t+1)) = u(\bar{q}).$$

Therefore, we must have $\bar{q} \succsim \bar{p}$. Now from Lemma 13, $\overline{D}(t)\varphi_t'(\overline{D}(t)) = \overline{v}$ for all $t > 0$. Together with $u(\bar{p}) = \varphi_t'(\overline{D}(t))$ and $u(\bar{q}) = \varphi_{t+1}'(\overline{D}(t+1))$, we have $\overline{D}(t)u(\bar{p}) = \overline{D}(t+1)u(\bar{q})$. Since $u(\bar{q}) \geq u(\bar{p})$, we have $\overline{D}(t+1) \leq \overline{D}(t)$, as desired.

Next, we claim that $\varphi_t'(d) \leq \varphi_{t+1}'(d)$ for all for all $t < T$ and $d \leq \overline{D}(t+1)$. Take any d in the domain of φ_{t+1}', that is, $d \leq \overline{D}(t+1)$. Since $\overline{D}(t+1) \leq \overline{D}(t)$ from the above claim, d also belongs to the domain of φ_t'. This implies that there exists some $p^t \in X_{ms}$ such that $d = D_{u(p)}(t)$. By Impatience, $U(p^t) \geq U(p^{t+1})$. PBT (ii) implies $p^{t+1} \in X_{ms}$. By the representation on X_{ms}, we have $D_{u(p)}(t)u(p) = U(p^t) \geq U(p^{t+1}) = D_{u(p)}(t+1)u(p)$. Thus, we must have $d = D_{u(p)}(t) \geq D_{u(p)}(t+1)$. Since φ_{t+1}' is strictly increasing, we have $\varphi_{t+1}'(d) \geq \varphi_{t+1}'(D_{u(p)}(t+1))$. Moreover, it follows from the FOC that

$$\varphi_t'(d) = \varphi_t'(D_{u(p)}(t)) = u(p) = \varphi_{t+1}'(D_{u(p)}(t+1)) \leq \varphi_{t+1}'(d),$$

that is, $\varphi_t'(d) \leq \varphi_{t+1}'(d)$ for all $d \leq \overline{D}(t+1)$.

To prove the statement of the lemma, we argue by induction. Take any real-valued, strictly increasing and continuous extension of φ_1' to $[0,1]$. This is always possible because $\varphi_1'(\overline{D}(1)) < \infty$ as demonstrated in Lemma 4. Now assume that there exists a desired extension up to $1 \leq t < T$. Since $\varphi_t'(d) \leq \varphi_{t+1}'(d)$ for all $d \leq \overline{D}(t+1)$ by the above claim and $\varphi_{t+1}'(\overline{D}(t+1)) < \infty$, it is possible to find a real-valued, strictly increasing and continuous extension of φ_{t+1}' to $[0,1]$ with preserving $\varphi_t'(d) \leq \varphi_{t+1}'(d)$ for all $d \in [0,1]$. Finally, by the fundamental theorem of calculus, the desired φ_t is obtained on $[0,1]$ as the integral of φ_t'. ■

E Proof of Theorem 2

(1) For any dated reward p^t with $u(p) > 0$, the discount function (which requires $D_{p^t}(t) > 0$ and $D_{p^t}(\tau) = 0$ for $\tau \neq t$) is determined by preference: if $\gamma \in [0,1]$ is such that $\gamma \circ p \sim p^t$, then $D_{p^t}(t) = \gamma$. Thus the discount functions for dated rewards are uniquely pinned down by preference. Therefore, the set $\{D_{p^t}(t) \in [0,1] \,|\, p \succsim 0\}$ defines the effective domain of the cost function φ_t in any representation. Let $\overline{D}(t) = \sup\{D_{p^t}(t) \in [0,1] \,|\, p \succsim 0\}$. Since an optimal discount function $D_x(t)$ is determined as a solution to the equation (7), $\overline{D}^i(t)$ must be the maximum discount factor at time t achievable in the cognitive optimization problem.Therefore, we have $\overline{D}^i(t) = \overline{D}(t)$, as desired.

(2) Since u^1 and u^2 are linear and represent the same preference over lotteries, there exists $\alpha > 0$ such that $u^2 = \alpha u^1$ (Note that we impose a normalization $u^i(0) = 0$ in the definition of the regular tuple).

Take a dated reward $x = p^t$. By the observation given in part (1), $D_x(t)$ is invariant between the two representation. By the first order condition,

$$(\varphi_t^2)'(D_x(t)) = u^2(p) = \alpha u^1(p) = \alpha(\varphi_t^1)'(D_x(t)),$$

which implies $\varphi_t^2|_{\overline{D}(t)} = \alpha\varphi_t^1|_{\overline{D}(t)}$.

By Lemma 4 of Noor and Takeoka [21],

$$\{x \in X \setminus \Delta_0 \,|\, \sum_{t>0} u^1(x_t)((\varphi_t^1)')^{-1}(u^1(x_t)) \leq \overline{v}^1\} = X_{ms}$$

$$= \{x \in X \setminus \Delta_0 \,|\, \sum_{t>0} u^2(x_t)((\varphi_t^2)')^{-1}(u^2(x_t)) \leq \overline{v}^2\}.$$

Since $u^2 = \alpha u^1$ and $\varphi_t^2|_{\overline{D}(t)} = \alpha\varphi_t^1|_{\overline{D}(t)}$,

$$\sum_{t>0} u^2(x_t)((\varphi_t^2)')^{-1}(u^2(x_t)) = \sum_{t>0} \alpha u^1(x_t)((\varphi_t^1)')^{-1}(\frac{1}{\alpha}\alpha u^1(x_t)) = \sum_{t>0} \alpha u^1(x_t)((\varphi_t^1)')^{-1}(u^1(x_t)),$$

and hence, we have

$$\sum_{t>0} u^2(x_t)((\varphi_t^2)')^{-1}(u^2(x_t)) \leq \overline{v}^2 \iff \sum_{t>0} u^1(x_t)((\varphi_t^1)')^{-1}(u^1(x_t)) \leq \frac{\overline{v}^2}{\alpha}.$$

Therefore, we must have $\overline{v}^2 = \alpha\overline{v}^1$.

F Appendix: Proof of Theorem 3

For the necessity, as in Lemmas 3 and 4 of Noor and Takeoka [21], we can establish the equivalence between X_{ms} and the set of streams where the corresponding unconstrained optiaml discount functions are feasible in the capacity constraint. Hence, on X_{ms}, the representation coincides with the CE model in NT. Necessity of Magnitude-Sensitive Homogeneity is proved in NT, and hence, the proof is omitted.

We show sufficiency. In Lemma 4, we already know that $\varphi_t : [0, \overline{D}(t)] \to \mathbb{R}_+$ is an increasing convex function that is strictly increasing, strictly convex, and differentiable on $[\underline{D}(t), \overline{D}(t)]$. Moreover, $D_r(t)$ is strictly increasing in r on

$$R_{ms}(t) = \{r \,|\, r = u(p) \text{ for some } p^t \in X_{ms}\}, \tag{15}$$

and is constant otherwise. We will show that this cost function $\varphi_t : [0, \overline{D}(t)] \to \mathbb{R}_+$ takes the power form for some constants $m > 1$ and $a_t > 0$,

$$\varphi_t(d) = a_t d^m. \tag{16}$$

Since $\succsim$ satisfies Magnitude-Sensitive Homogeneity, by the same proof of Theorem 7 (Appendix E) in NT, we can show that $D_r(t)$ on $R_{ms}(t)$ is written as a power form, that is, $D_r(t) = \kappa_t r^\theta$ for some $\kappa_t > 0$ and $\theta > 0$. Then, $\varphi_t : [0, \overline{D}(t)] \to \mathbb{R}_+$ is rewritten as in (16).

For convenience for the reader, we reproduce the proof of this result here. First we show that $D_r(t)$ is homogeneous on $R_{ms}(t)$:

Lemma 15 *For each $\alpha \in (0,1]$ there is $h(\alpha)$ s.t. for any $p^t \in X_{ms}$,*

$$D_{\alpha u(p)}(t) = h(\alpha) D_{u(p)}(t).$$

Moreover, $h(\alpha)$ is a continuous function on $(0,1]$ with $h(1) = 1$.

Proof. We first note that for any $x \in X$ such that $x \succ 0$ and $\alpha \in (0,1]$, there exists a unique $\gamma_\alpha(x) \in (0,1]$ such that $\gamma_\alpha(x) \circ c_x \sim \alpha \circ x$: Since $x \succ 0$, Risk Preference and Weak Homotheticity imply $c_x \succ \alpha \circ c_x \succsim \alpha \circ x$. Consequently by Continuity and Monotonicity, the desired $\gamma_\alpha(x) \in (0,1]$ exists and is unique.

Take any dated reward $p^t \in X_{ms}$. By definition of X_{ms}, it must be that $p \succ 0$, and by Monotonicity, $p^t \succ 0$. As noted, there exists $\gamma_\alpha(p^t) \in (0,1]$ such that

$$\gamma_\alpha(p^t) \circ c_{p^t} \sim \alpha \circ p^t.$$

We make several observations about γ_α :

(i) $\gamma_\alpha(p^t)$ is independent of p^t for $p \succ 0$, and so can be written it as γ_α.

Magnitude-Sensitive Homogeneity implies that $\gamma_\alpha(p^t)$ is independent of p and t.

(ii) γ_α is strictly increasing, $\gamma_\alpha = 1$ when $\alpha = 1$, and $\lim_{\alpha \to 0} \gamma_\alpha = 0$.

Since $c_{p^t} \sim p^t$ by definition of present equivalents, and since γ_α is defined by $\gamma_\alpha \circ c_{p^t} \sim \alpha \circ p^t$, it follows trivially that $\gamma_\alpha = 1$ when $\alpha = 1$. Moreover, by Risk Preference and Monotonicity, γ_α must be strictly increasing in α, since $\alpha < \alpha'$ implies $\gamma_\alpha \circ c_{p^t} \sim \alpha \circ p^t \prec \alpha' \circ p^t \sim \gamma_{\alpha'} \circ c_{p^t}$. Finally, $\alpha \to 0$ implies $\alpha \circ p^t = (\alpha \circ p)^t \to 0^t = 0$, and so it must be that $\lim_{\alpha \to 0} \gamma_\alpha = 0$.

(iii) γ_α is continuous in α.

By the representation,

$$U_t(\alpha \circ p) = U((\alpha \circ p)^t) = u(\gamma_\alpha \circ c_{p^t}) = \gamma_\alpha u(c_{p^t}) = \gamma_\alpha U(p^t),$$

that is, $U_t(\alpha \circ p) = \gamma_\alpha U_t(p)$. Since U_t is continuous, so is γ_α.

(iv) γ_α satisfies

$$D_{u(\alpha \circ p)}(t) = \frac{\gamma_\alpha}{\alpha} D_{u(p)}(t).$$

We saw above that $U_t(\alpha \circ p) = \gamma_\alpha U_t(p)$. It follows that

$$U_t(\alpha \circ p) = \gamma_\alpha U_t(p) \iff D_{u(\alpha \circ p)}(t) u(\alpha \circ p) = \gamma_\alpha D_{u(p)}(t) u(p)$$

$$\iff \alpha D_{u(\alpha \circ p)}(t) u(p) = \gamma_\alpha D_{u(p)}(t) u(p) \iff D_{u(\alpha \circ p)}(t) = \frac{\gamma_\alpha}{\alpha} D_{u(p)}(t).$$

Defining $h(\alpha) = \frac{\gamma_\alpha}{\alpha}$, we obtain the desired expression. ■

Lemma 16 *Define $\overline{R}(t) := [0, \overline{r}_t]$. There is $\theta \in \mathbb{R}$ s.t. for any $t > 0$, $r \in \overline{R}(t)$ and $\alpha \in (0,1]$,*

$$D_{\alpha r}(t) = \alpha^\theta D_r(t).$$

Proof. From Lemma, for any $\alpha \in (0,1]$, there is $h(\alpha)$ s.t. for any $r \in \overline{R}(t)$,

$$D_{\alpha r}(t) = h(\alpha) D_r(t).$$

Note that $h(1) = 1$ and $h(\alpha)$ is continuous. Moreover, we find that $h(\alpha\gamma)D_r(t) = D_{\alpha\gamma r}(t) = h(\alpha)D_{\gamma r}(t) = h(\alpha)h(\gamma)D_r(t)$. Indeed, h satisfies the multiplicative Cauchy equation:

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

To convert this into a standard Cauchy functional equation on $\mathbb{R}_+$, define $g : \mathbb{R}_+ \to \mathbb{R}$ by $g(\lambda) = \ln h(e^{-\lambda})$ for any $\lambda \in \mathbb{R}_+$. Since h is continuous, so is g. Observe that for any $\lambda, \nu \in \mathbb{R}_+$

$$g(\lambda + \nu) = \ln h(e^{-\lambda}e^{-\nu}) = \ln h(e^{-\lambda})h(e^{-\nu}) = \ln h(e^{-\lambda}) + \ln h(e^{-\nu}) = g(\lambda) + g(\nu),$$

that is, $g(\lambda + \nu) = g(\lambda) + g(\nu)$, and so g satisfies the standard Cauchy functional equation on $\mathbb{R}_+$. By Aczel [1, Section 2.1.1. Theorem 1], there exists $\zeta \in \mathbb{R}$ such that $g(\lambda) = \zeta\lambda$. Define $\theta = -\zeta$ and observe that h satisfies, for any $\alpha \in (0,1]$,

$$\ln \alpha^{\theta} = \zeta \ln \frac{1}{\alpha} = g(\ln \frac{1}{\alpha}) = \ln h(e^{-\ln \frac{1}{\alpha}}) = \ln h(\alpha)$$

that is, $h(\alpha) = \alpha^{\theta}$ for all $\alpha \in (0,1]$. We have thus shown that $D_{\alpha r}(t) = h(\alpha)D_r(t) = \alpha^{\theta} D_r(t)$, as desired. ■

Lemma 17 *For any $t > 0$, there exist $\theta > 0$ and $\kappa_t > 0$ such that for all $r \in \overline{R}(t)$,*

$$D_r(t) = \kappa_t r^{\theta}.$$

Proof. Take any $r \in \overline{R}(t)$. Then $r \leq \overline{r}_t$. By Lemma 16, $D_r(t) = D_{\frac{r}{\overline{r}_t}\overline{r}_t}(t) = \left(\frac{r}{\overline{r}_t}\right)^{\theta} D_{\overline{r}_t}(t)$. We obtain the expression $D_r(t) = \kappa_t r^{\theta}$ by letting $\kappa_t := \left(\frac{1}{\overline{r}_t}\right)^{\theta} D_{\overline{r}_t}(t)$. Since $\overline{R}(t)$ is a non-trivial interval and $D_r(t)$ is strictly increasing on it, it must be that $\theta > 0$. ■

Lemma 18 *D is the solution to the cognitive optimization wrt to some φ_t defined by $a_t > 0$, $m > 1$ and $\varphi_t(d) = a_t d^m$ for all $d \leq \overline{D}(t)$.*

Proof. By Lemma 17, $D_r(t) = \kappa_t r^{\theta}$ for all $r \in \overline{R}(t)$ where $\kappa_t > 0$ and $\theta > 0$. Using the FOC, define φ_t on $[0, \overline{D}(t)]$ as follows. For all $r \in \overline{R}(t)$, let $r = \varphi_t'(\kappa_t r^{\theta})$, so that $\varphi_t'(d) = \left(\frac{d}{\kappa_t}\right)^{\frac{1}{\theta}}$. Together with $\varphi_t(0) = 0$, we have

$$\varphi_t(d) = \frac{\theta}{(1+\theta)\kappa_t^{\frac{1}{\theta}}} d^{\frac{1+\theta}{\theta}}.$$

Let $m = \frac{1+\theta}{\theta} > 1$ and $a_t = \frac{\theta}{(1+\theta)\kappa_t^{\frac{1}{\theta}}} > 0$. Then, $\varphi_t(d) = a_t d^m$ for all $d \in [0, \overline{D}(t)]$, as desired. ■

Next, we show that for any stream $x \in X_{ms}$, the total cost for the optimal discount function D_x is proportional to the utility from period 1 onward.

Lemma 19 *For all $x \in X_{ms}$,*

$$\varphi(D_x) = \frac{1}{m} U(0, x_{-0}).$$

Proof. As shown in Lemma 18, φ_t admits a power form, $\varphi_t(d) = a_t d^m$. For all $x \in X_{ms}$, the FOC implies $m a_t (D_{x_t}(t))^{m-1} = u(x_t)$. Thus,

$$\begin{aligned}\varphi(D_x) &= \sum_{t>0} \varphi_t(D_{x_t}(t)) = \sum_{t>0} a_t \left(\frac{u(x_t)}{m a_t}\right)^{\frac{m}{m-1}} = \frac{1}{m} \sum_{t>0} \left(\frac{u(x_t)}{m a_t}\right)^{\frac{1}{m-1}} u(x_t) \\ &= \frac{1}{m} \sum_{t>0} D_{x_t}(t) u(x_t) = \frac{1}{m} U(0, x_{-0}),\end{aligned}$$

as desired. ■

For any x, there exists a unique $\alpha_x \in (0, 1]$ such that $\alpha_x \circ x \in bd^+(X_{ms})$. By Lemmas 11 and 19,

$$K_x = K_{\alpha_x \circ x} = \varphi(D_{\alpha_x \circ x}) = \frac{1}{m} U(0, \alpha_x \circ x_{-0}) = \frac{\overline{v}}{m}, \tag{17}$$

that is, K_x is constant for all x, as desired.

From now on, let $K > 0$ be the constant number given by (17).

Lemma 20 *For all $t \geq 1$, $a_t(\overline{D}(t))^m = K$.*

Proof. Consider a dated reward $p^t \in X_{ms}$. By the FOC, $u(p) = m a_t D_{u(p)}(t)^{m-1}$. Since p^t is magnitude sensitive,

$$D_{u(p)}(t) = \left(\frac{u(p)}{m a_t}\right)^{\frac{1}{m-1}} \leq \overline{D}(t).$$

Let $\bar{p}^t$ be a dated reward which attains a supremum of $\{u(p) \,|\, p^t \in X_{ms}\}$. Since the capacity constraint will be binding at $\bar{p}^t$,

$$K = a_t \left(\frac{u(\bar{p})}{m a_t}\right)^{\frac{m}{m-1}} = a_t(\overline{D}(t))^m,$$

as desired. ■

Lemma 21 *For all $t < T$, $a_{t+1} \geq a_t$.*

Proof. Lemma 20 implies $1/m \times \overline{D}(t) \times m a_t \overline{D}(t)^{m-1} = K$, or $\overline{D}(t) u(\bar{p}_t) = mK$, where $\bar{p}_t$ satisfies the FOC $u(\bar{p}_t) = m a_t \overline{D}(t)^{m-1}$. By the same argument as in Lemma 14, we can show that $\overline{D}(t+1) \leq \overline{D}(t)$. Thus, Lemma 20 implies $a_{t+1} \geq a_t$. ■

Since the shape of the cost function beyond the capacity constraint K does not have any behavioral implications, we can extend $\varphi_t : [0, \overline{D}(t)]$ by $\varphi_t(d) = a_t d^m$ on the whole unit interval $[0, 1]$. Then, $(u, \{\varphi_t\}_{t \geq 1})$ is a regular tuple by Lemma 21, as desired.

References

[1] Aczel, J. (1966): *Lectures on Functional Equations and Their Applications*, NY Academic Press.

[2] Banerjee, A., and E. Duflo (2007): "The Economic Lives of the Poor," *Journal of Economic Perspectives* 21(1), pp. 141–167.

[3] Becker, G. and C. Mulligan (1997): "The Endogenous Determination of Time Preference," *The Quarterly Journal of Economics* 112(3), pp. 729–758

[4] Browning, M. and T. F. Crossley (2001): "The Life-Cycle Model of Consumption and Saving," *Journal of Economic Perspectives* 15, pp. 3–22.

[5] Brunnermeier, M., and J. Parker (2005): "Optimal Expectations," *American Economic Review* 95, pp. 1092–1118.

[6] Dohmen, T., A. Falk, D. Huffman and U. Sunde (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100(3), pp. 1238–1260.

[7] Ellis, A. (2018): "Foundations for Optimal Inattention," *Journal of Economic Theory* 173, pp. 56–94.

[8] Ergin, H., and T. Sarver (2010): "A Unique Costly Contemplation Representation," *Econometrica* 78(4), pp. 1285–1339.

[9] Fredrick, S., G. Loewenstein and T. O'Donoghue (2002): "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature* 40(2), pp. 351–401.

[10] Gabaix, X. (2014): "A Sparsity-Based Model of Bounded Rationality," *The Quarterly Journal of Economics* 129(4): pp. 1661–1710.

[11] Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics* 18, pp. 141–153.

[12] Koopmans, T. C. (1972): "Representation of Preference Orderings over Time," in C.B. McGuire and R. Radner, editors, *Decision and Organization: A Volume in Honor of Jacob Marschak*. Amsterdam: North-Holland.

[13] Kureishi, W., H. Paule-Paludkiewicz, H. Tsujiyama and M. Wakabayashi (2021): "Time preferences over the Life Cycle and Household Saving Puzzles," *Journal of Monetary Economics* 124, pp. 123-139.

[14] Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics* 112, pp. 443–77.

[15] Liang, M., S. Grant, S. Hsieh (2019): "Costly self-control and limited willpower," *Economic Theory*, pp 1-26.

[16] Maccheroni, F., M. Marinacci and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica* 74, pp. 1447–1498.

[17] Mani, A., S. Mullainathan, E. Shafir and J. Zhao (2013): "Poverty Impedes Cognitive Function," *Science* 341 (6149), pp. 976-980.

[18] Masatlioglu, Y., D. Nakajima, and E. Ozdenoren (2020): "Willpower and Compromise Effect," *Theoretical Economics* 15, pp. 279–317.

[19] Noor, J. and N. Takeoka (2022): "Optimal Discounting," *Econometrica* 90, pp. 585–623.

[20] Noor, J. and N. Takeoka (2022): "Supplement to "Optimal Discounting"," *Econometrica Supplementary Material*.

[21] Noor, J. and N. Takeoka (2020): "Supplementary Appendix to "Constrained Optimal Discounting", " Working paper.

[22] Noor, J. and N. Takeoka (2023): "Optimal Stochastic Discounting," mimeo.

[23] O'Donoghue, T., and M. Rabin (1999), "Doing it Now or Later," *American Economic Review* 89, pp. 103–124.

[24] Ozdenoren, E., S. W. Salant, and D. Silverman (2012), "Willpower and the Optimal Control of Visceral Urges," *Journal of the European Economic Association* 10, pp. 342–368.

[25] Read, D. and M. Scholten (2012): "Tradeoffs Between Sequences: Weighing Accumulated Outcomes Against Outcome-Adjusted Delays," *Journal of Experimental Psychology: Learning, Memory and Cognition* 38(6), pp.1675-88.