Constrained Optimal Discounting*

Jawwad Noor Norio Takeoka

May 29, 2026

Abstract

Noor and Takeoka [28] model an agent whose impatience is determined by an (unconstrained) cognitive optimization problem. This paper considers an extension in which the agent has a limited stock of cognitive resources. The model generates distinctive patterns in how the relative impatience across periods changes with rewards and in violations of Separability for large rewards. The model can also endogenously produce lower impatience with age, providing a cognitive account for anomalous life-cycle saving behavior. In a task-completion setting, the model can generate cycles of activity and inactivity in task completion, whereas standard models predict the completion of all tasks in consecutive periods.

1 Introduction

Deck and Jahedi [10] show that increased cognitive load on subjects leads to increased impatience. Dohmen et al. [11] show that higher cognitive abilities are correlated with lower impatience. Mani et al. [26] discuss evidence that poverty has a negative impact on one's "mental bandwidth", limiting both cognitive abilities and patience (in the form of impulse control). Such studies suggest that impatience potentially draws on cognitive

*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 Numbers JP19KK0308 and JP25K00619. Part of this research was conducted while Norio Takeoka was visiting the Department of Economics, Boston University, whose hospitality is gratefully acknowledged. We thank audiences at the Decision Theory Workshop 2023 (Hitosubashi University) and AMES 2023 (NTU Singapore), and Jiaqi Yang for expert research assistance. The usual disclaimer applies.

resources. A possible channel can be found in Pigou [34], who relates impatience with an inability to imagine future experiences with clarity: "our telescopic faculty is defective, and we, therefore, see future pleasures, as it were, on a diminished scale". This is supported by Hershfield et al. [19, 20], who demonstrate that by making it easier to visually imagine one's future self (such as through a computer-generated image), subjects increase savings. Indeed, anecdotally, visualizing the future consequences of one's current actions provides a means for avoiding procrastination and achieving self-control.

Building on such views, Noor and Takeoka [28] (henceforth NT) hypothesize that costly cognitive effort in visualizing the future may be exerted to reduce impatience,[1] and show that such a hypothesis is consistent with the magnitude effect, the robust finding in the lab and the field that people are more patient when dealing with larger outcomes than smaller ones (Fredrick et al. [16], Andersen et al. [1]). In NT's model, impatience is determined endogeneously by a cognitive optimization problem involving a cognitive cost. In this paper, we extend the model to include limited cognitive capacity. Formally, taking a (static) preference $\succsim$ over the set $X = C^{T+1}$ of finite horizon consumption streams as their primitive, NT introduce 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

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

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

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). Given the metaphor of multiple selves, the discount function D is interpreted as the distribution of empathy across future selves, which is chosen to balance the resulting utility $\sum_{t \geq 1} D(t) u(x_t)$ against the cost $\sum_{t \geq 1} \varphi_t(D(t))$ of generating empathy D. The cost is interpreted as arising from the current self's cognitive act of imagining herself in the shoes of her future selves.[2] This paper extends the CE model to allow for limited cognitive

[1] Becker and Mulligan [4] model an agent who alters her discount function by physically investing, for instance, in education. NT's model is the counterpart where the investment is cognitive.

[2] An alternative model of cognitive optimization may focus on resolving some kind of subjective uncertainty. Our agent's problem faces no uncertainty, but rather a difficulty in caring about the future in a sense that is familiar from self-control problems.

capacity, in which case the cognitive optimization problem (2) is subject to the capacity constraint:

$$\sum_{t\geq 1} \varphi_t(D(t)) \leq K \tag{3}$$

for some $K \leq \infty$. That is, the agent cannot produce empathy D that costs more than K. The model is referred to as a *Constrained Costly Empathy (CCE)* representation for the preference.

A feature of the model is the additivity of the cost function $D \mapsto \sum_{t\geq 1} \varphi_t(D(t))$. This assumption plays an important role for the behavioral identification of the capacity parameter K. The key question is whether the capacity constraint binds in the agent's choice of discount function and, more generally, how the parameter K can be inferred from observed behavior.

Under additive costs, when the capacity constraint does not bind, the optimal discount factors can be chosen independently across periods. Consequently, the preference admits an additively separable representation over time. When the constraint binds, however, the allocation of cognitive resources across periods becomes interdependent, and the optimal discount factors cannot be chosen separately across periods. As a result, the preference generally fails to satisfy time separability. Thus, under additive costs, violations of time separability provide behavioral evidence that the cognitive capacity constraint is binding and thereby allow the parameter K to be identified.

By contrast, with a general (nonseparable) cost function $D \mapsto \varphi(D)$, the behavioral implications are fundamentally different. In this case, violations of time separability may arise either because the capacity constraint binds or because the cost function itself is nonseparable. Consequently, observed nonseparable behavior cannot be attributed to the capacity constraint, and the parameter K cannot be behaviorally identified without further restrictions on the cost function.

Accordingly, the model has the following implications:

1. 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, the agent will always be less impatient when faced with larger streams. However, if the agent has limited cognitive capacity, then at some point the capacity constraint will bind, and it is infeasible to have decreased impatience towards all periods. Consequently, increasing a stream will necessitate a *trade-off* between impatience in different periods.

2. Violations of Separability.[3] 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 $(100, 200, 0)$. In general, the model tells us that binding cognitive constraints can explain such Separability violations that pertain to impatience.

Seeking further economic implications of cognitive constraints, the paper considers several applications in dynamic settings. 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. We also show that cognitive constraints can endogeneously explain the empirical finding that impatience diminishes with age: young age comes with a longer time horizon and the resulting cognitive demands are only relaxed with age. Another set of applications study the task completion problem, uncovering patterns of behavior that are inconsistent with the lack of cognitive constraints.

The remainder of the paper proceeds as follows. We conclude the Introduction with a discussion of related literature. Section 2 describes our basic framework and derives some key properties of the CCE representation. Section 3 presents an axiomatization of the CCE representation with homogeneous cost functions. Sections 4 and 5 provide applications of the dynamic model. Section 6 concludes with possible future directions for research. All proofs are relegated to the appendices and the supplementary appendix (Noor and Takeoka [29]).

[3] 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 (Koopmans [21]). Separability of preference is necessary for the existence of an additively separable representation for preference.

1.1 Related Literature

There are several models in the literature that incorporate subjective optimization, such as those of optimal expectations (Brunnermeier and Parker [7]), optimal contemplation (Ergin and Sarver [15]) and optimal attention (Ellis [13], Gabaix [17], Bolte and Raymond [5]). 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. Ozdenoren et al. [32] consider the cake-eating problem with a fixed initial stock of willpower, which is depleted over time by exercising self-control. In a discrete setting, Masatlioglu et al. [27] 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 et al. [25] 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. While these papers on willpower do not involve a cognitive choice of discount function, NT suggest that willpower may in fact be operationalized by thinking about consequences for future selves, which induces greater patience. The main take-away of the present paper is that, in any context, binding cognitive constraints express themselves behaviorally in trade-off in impatience across periods for large rewards, and in violations of separability for large rewards.

Non-separability in intertemporal preferences has been studied in other contexts. In Kreps and Porteus [22], preferences are modeled recursively, allowing for a non-additive aggregation of current utility and continuation values. However, this model retains a form of time separability. Since continuation values summerize all future consequences, changes in common earlier payoffs do not affect the ranking of streams that differ only from a later period onward. Models of endogenous time preferences such as Shi and Epstein [36] allow the discount factor to depend on past consumption levels, and hence may violate time separability. In our model, non-separability arises in a different way. Since the discount function is chosen optimally for the entire consumption stream under cognitive constraints, changes in earlier payoffs can alter the evalution of later differences, potentially leading to preference reversals.

A related non-separable model is Wakai [37], which introduces multiple discount factors and allows the effective discount factor to depend on continuation utilities. The model is primarily motivated by explaining non-separable choice patterns such as a preference for spreading consumption across periods. In contrast, in our model, non-separability arises endogenously when a cognitive constraint binds. An important behavioral difference concerns the magnitude effect. In Wakai [37], if a consumption stream pays a payoff only in period t, the same discount factor applies regardless of the payoff size, and hence the model cannot generate the magnitude effect.

Wakai [37] and our model also differ in their implications for timing decisions for task completion and procrastination. In the stationary version of Wakai [37], procrastination does not arise in the task environment we consider. That is, if a task is completed immediately under a short deadline, extending the deadline does not induce delay. A similar implication holds in willpower models when a single task must be completed. In our model, by contrast, procrastination may occur even when the cognitive cost is time-invariant. We discuss these behavioral differences in Section 5.

A different approach to explaining magnitude effects is taken by Chakraborty [9]. In the model, magnitude effects are derived through a minimum representation of present equivalents. That is, each dated payoff is converted into a conservative present equivalent and then additively aggregated across periods. While this formulation can generate payoff-dependent magnitude effects, the resulting utility over consumption streams remains additively separable, and hence, the evaluation of a given payoff depends only on its own magnitude and delay. By contrast, in our model, the discount function is chosen optimally for the entire continuation utility stream. Accordingly, the effective discounting applied to a future payoff may vary with the configuration of other future payoffs in the stream, thereby generating stream-dependent magnitude effects and non-separability.

2 Model and General Implications

2.1 Primitives

There are $T+1 < \infty$ periods, starting with period 0. The consumption space C is assumed to be $C = \mathbb{R}_+$. Consider the space of consumption streams $X = C^{T+1}$, endowed with the product topology. A typical element in X is denoted by $x = (x_0, x_1, \cdots, x_T)$. The primitive of our model is a preference $\succsim$ over X.

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) $u : C \to \mathbb{R}_+$ is continuous and strictly increasing with $u(0) = 0$.

(ii) $\varphi_t : [0, 1] \to \mathbb{R}_+$ is an increasing convex function that is strictly increasing, strictly convex and continuously differentiable on $D_t^+ = \{d : \varphi_t(d) > 0\}$, and satisfies $\varphi_t(\underline{d}_t) = 0$, $\varphi_t'(\underline{d}_t) = 0$, where $\underline{d}_t = \inf D_t^+$, 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. The condition $u(0) = 0$ is more than a mere normalization of u since the sign of u matters for the cognitive optimization problem defined below. 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 the interval $[0, \underline{d}_t]$ but must have standard properties of a cost function on $[\underline{d}_t, 1]$, 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 $t < T$.

Under these conditions the maximization problem (2)-(3) admits a unique solution: Since φ_t is continuous and convex, and the constraint set of discount functions defined by (3) is a closed convex subset of $[0, 1]^T$, the problem (2) maximizes a continuous concave function on a compact convex set. Thus it admits a solution. Since $\varphi_t'(\underline{d}_t) = 0$, it must be that $D(t) \geq \underline{d}_t$ for each t and any solution D. So we can restrict attention to the constraint set corresponding to a set of discount functions on which each φ_t is strictly convex. It follows that the solution must be unique.

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 $U : X \to \mathbb{R}_+$ defined by (1)-(3).*

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*. It is worth noting that this model is essentially a constrained optimization problem since $D(t)$ is bounded above by 1 for each $t > 0$. Defining the cost of perfect patience by $K_t := \varphi_t(1)$, this is equivalent to requiring that $\varphi_t(d) \leq K_t$ for each $t > 0$. The CCE representation replaces these T intratemporal constraints by a more interesting single intertemporal constraint (3).

The CCE representation was interpreted in the Introduction, except that it remains to intepret the condition $0 < K \leq \varphi_1(1)$. Given $K_t := \varphi_t(1)$ and the fact that cost functions are increasing in t the condition implies that

$$0 < K \leq K_t, \text{ for all } t > 0.$$

Intuitively, K is exhausted (weakly) before perfect patience can be achieved in any period. The condition yields the substantive computational simplication that solving the maximization problem subject to the intertemporal constraint (3) alone is *sufficient*, in that all the intratemporal constraints are met 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.

A final remark concerns the interpretation of the cognitive cost of discounting. One might wonder whether imagining or evaluating future outcomes is cognitively demanding, rather than a source of pleasure or motivation. Peters and Büchel [33] show that prompting individuals to engage in episodic future thinking, that is, imagining concrete future events, reduces deley discounting. Moreover, Deck and Jahedi [10] report that imposing cognitive load increases impatience in intertemporal choice. These findings suggest that evaluating future consequences requires cognitive resources, and hence, sustaining an appreciation of distant consumption may be cognitively costly. At the same time, we do not claim that this is the primary channel through which cognitive resources are depleted. Another important role emphasized in the literature is to resist immediate temptations as in models of limited willpower. Our framework is complementary to this literature. Rather than focusing on self-control over current consumption, we consider the possibility that cognitive resources are required to sustain attention to future payoffs through the choice of discount factors. Integrating these channels would be an interesting direction for future research. See Section 6 for more details.

2.3 General Implications of Binding Constraints

We make some observations, collected in two Propositions, about the CCE model and in particular to the solution D_x to the cognitive optimization problem (2)-(3) in the CCE model. These observations will suggest observable implications associated with binding cognitive constraints.

Consider the set of streams $\{x \in X : \sum_{t\geq 1} \varphi_t(D_x(t)) < K\}$ where the capacity constraint (3) 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 $t > 0$.

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:*

$$y \quad x \implies D_y \leq D_x.$$

If $x \notin X_s$ then for any $y \geq x$,

$$D_y(t) > D_x(t) \text{ for some } t > 0 \implies D_y(t') < D_x(t') \text{ for some } t' > 0,$$

and, moreover:

$$\frac{u'(x_t)}{u'(x_{t'})} = \frac{u'(y_t)}{u'(y_{t'})} \text{ for all } t, t' > 0 \implies D_x = D_y$$

The first condition says that if the capacity constraint is not binding at stream x, then discounting will be sensitive to incentives provided by the consumption in the stream. Thus lower consumption $y \leq x$ will weakly decrease all $D(t)$.[4] On the other hand, if the constraint binds at x, then increasing incentives will never lead to $D_y \geq D_x$ with strict inequalty for some t. That is, if some $D(t)$ strictly increases then some other $D(t')$ must strictly decrease. Intuitively, the binding constraint forces the agent to trade-off empathy across the different selves. The last part of the Proposition indicates that it is possible that increasing incentives lead to no change in D. This happens if y and x have the same ratios of marginal utilities between different future periods.

The Proposition suggests behavioral patterns consistent with binding cognitive constraints in the CCE model. Given the first two conditions in the Proposition, standard methods of eliciting discount functions can be used to determine whether increases in the stream lead the discount function to monotonically increase or not. The third condition can also be exploited. Consider the 45 degree line in X, corresponding to constant consumption streams $x_\lambda = (\lambda, ..., \lambda)$. The Proposition implies that the marginal rate of substitution along the 45 degree line must be constant in λ beyond some point on the diagonal, that is, beyond the point where the agent becomes constrained.[5]

There is an obvious analogy with production theory, where a utility vector corresponds to a price vector and a discount function corresponds to a production vector. When the agent is constrained, she is choosing a discount function on a "production frontier" defined by $\{D \,|\, \varphi(D) = K\}$. In fact, when the agent is constrained, even if φ is not additive nor differentiable, the cognitive maximization implies a "law of supply" given by $\sum_{t>0}(D_x(t) - D_y(t))(u(x_t) - u(y_t)) \geq 0$.

We now present a second behavioral implication of the model. For any streams $x =$

---

[4] In the model, it will be the case that $D(t)$ will strictly decrease for some $t > 0$. This is a consequence of assuming that φ_t is strictly increasing and differentiable. NT consider a more general model where φ_t may not be differentiable. In this case a small reduction in incentives may not necessarily change D due to a kink in the cost function. However a large enough reduction in incentives will.

[5] When the utility index is of the power form $u(x) = x^\beta$ then this property holds along any ray.

$(x_0, \cdots, x_T)$ and $y = (y_0, \cdots, y_T)$, denote by (x_t, y_{-t}) the stream that pays x_t at time t and pays $y_{t'}$ at all other $t' \neq t$. 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 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.

The Proposition suggests a pattern that may arise when the agent is cognitively constrained in the CCE model. Denote by $c_t x$ the stream that pays c at time t and otherwise pays according to x in all other periods. Koopmans' Separability condition requires that $c_t x \succsim c_t y \implies c'_t x \succsim c'_t y$ for any c, c', x, y, that is, the ranking of two streams is not sensitive to common consumption (Koopmans [21]). If the agent is cognitively constrained, then this condition must be violated, albeit in a manner indicative of limited cognitive resources. To illustrate, consider an agent who may be willing to give up \$10 in period 1 to obtain \$20 in period 3:

$$(0, 10, 0, 0) \prec (0, 0, 0, 20).$$

However, increasing period 2 consumption may lead the agent to behave less patiently, since she is unable to appreciate the period 3 reward sufficiently:

$$(0, 10, c, 0) \succ (0, 0, c, 20).$$

As noted in the Introduction, in a general subjective optimization problem with potentially non-additive costs, there is no meaningful distinction between constrained and unconstrained optimization. The key theoretical insight of this paper is that if we assume additive costs, then constrained optimization begins to have meaning. This meaning is captured in the sharp and natural implications presented in Propositions 1 and 2. The central role of additivity of cost functions is evident. When, and only when, the agent is unconstrained at x, additivity implies that the agent can optimally tailor $D_x(t)$ to x_t independently of the consumption in other periods. This gives rise to the additive separability of the model shown in Proposition 2, and the fact that increases in the stream lead to an increase in $D_x(t)$ for all periods in Proposition 1. Against this backdrop, it emerges that the existence of constraints leads to particular violations of these properties.

2.4 Homogeneous CCE Model

We define the *Homogeneous CCE model* by a CCE model where costs have the power form $\varphi_t(d) = a_t \cdot d^m$. Note that m is independent of t and the requirements of the CCE model impose that a_t is increasing in t.

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,

$$D_x(t) = \left(\frac{u(x_t)}{ma_t}\right)^{\frac{1}{m-1}} = \gamma(t)u(x_t)^{\frac{1}{m-1}} \tag{4}$$

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

$$D_x(t) = \frac{(mK)^{\frac{1}{m}} \gamma(t) u(x_t)^{\frac{1}{m-1}}}{\left\{\sum_{\tau \geq 1} \gamma(\tau) u(x_\tau)^{\frac{m}{m-1}}\right\}^{\frac{1}{m}}} \tag{5}$$

which is not time-separable, that is, $D_x(t)$ depends on the entire stream x and not just payoff at time t.

From the perspective of applications, a virtue of this model is that it admits an extremely clean way of distinguishing whether the cognitive constraint is binding or not. 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, the constraint is not binding at a stream x iff its future payoff $U(x) - u(x_0)$ is less than some threshold.

The preceding observations directly lead to a clean reduced form for the Homogeneous CCE model, which we present next without proof. Write $\gamma(t) := (ma_t)^{-\frac{1}{m-1}}$. Since a_t is increasing, $\gamma(\cdot)$ is a weakly decreasing function.

Proposition 3 *If $\succsim$ admits a Homogeneous CCE representation $(u, \{a_t d^m\}_{t=1}^T, K)$, then*

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

According to the reduced form, when the future utility of a stream is "small" (less than mK), the utility function is additively separable and the future utility index is effectively $u(x_t)^{\frac{m}{m-1}}$, a power transformation of immediate utility u. The parameter m affects 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.

3 Behavioral Foundations for the Homogeneous CCE Model

We have seen that nonseparability and magnitude-dependent impatience are two key implications of the CCE model. To paint a more complete behavioral picture of the agent, we now identify exhaustive implications of its special case, the Homogeneous CCE model.

As before, there are $T + 1 < \infty$ periods, starting with period 0. The space C of outcomes is assumed to be $C = \mathbb{R}_+$. For reasons that will become apparent in Sections

3.1.1 and 3.1.2, we will employ lotteries over consumption. Let Δ denote the set of simple lotteries over C, with generic elements $p, q, \cdots$. We refer to p as consumption. Consider the space of sequences of lotteries $X = \Delta^{T+1}$as consumption streams, endowed with the product topology. This extends deterministic consumption streams by allowing stochastic consumption in each period.[6] A typical element in X is denoted by $x = (x_0, x_1, \cdots, x_T)$. The subset $\Delta_0 \subset X$ is defined as the set of streams of the form $(p, 0, \cdots, 0)$. The primitive of our model is a preference $\succsim$ over X.[7]

A *Homogeneous CCE* representation $U : X \to \mathbb{R}_+$ for a preference $\succsim$ over X is defined by (1)-(3) where the tuple $(u, \{\varphi_t\}, K)$ is such that (i) $u : \Delta \to \mathbb{R}_+$ is mixture linear and its restriction to C is unbounded, continuous and strictly increasing with $u(0) = 0$, (ii) for each $t \geq 1$, the cost function $\varphi_t : [0, 1] \to \mathbb{R}_+$ takes the power form $\varphi_t(d) = a_t \cdot d^m$, (iii) a_t is increasing in t, and (iv) $K \leq a_1$. These properties are familiar from the CCE model, except for the requirement that u is required to be unbounded here. Unboundedness ensures the existence of present equivalents for streams (defined below), which play an important role in the axiomatization.

In the deterministic model, the optimal discount function depended on the stream of utility of consumption $(u(x_1), \cdots, u(x_T))$. In our extension to the lottery setting, the same is true, except that the stream $(u(x_1), \cdots, u(x_T))$ consists of expected utility of consumption in each period. Formally, the agent first evaluates each period's lottery by its expected utility and obtain a stream of expected utilities. She then chooses an optimal discount function for this stream subject to the cognitive constraint, exactly as in the deterministic model. The evaluation of a lottery stream is given by the discounted sum of the expected utilities under this optimally chosen discount function.

One could imagine that the agent potentially "frames" a stream $x = (x_0, x_1, \cdots, x_T)$ as a lottery over deterministic consumption streams and solves the cognitive optimization problem for each deterministic stream separately. For instance, if she thinks of stream as a lottery that yields a deterministic stream $(1, 1, 1)$ with probability α and $(100, 100, 100)$ with probability $1 - \alpha$, then she may solve the cognitive optimization problem for each deterministic stream separately and then take an expectation. We find this less plausible because if the agent has garnered empathy for future selves when thinking about one deterministic stream, then that empathy should in principle extend to all streams in the

[6]Altough one can generally consider distributions over entire consumption streams to allow for intertemporal correlations, such correlations are not relevant for our analysis.

[7]Ideally, foundations for a theory of time preference would not employ lotteries. However, the domain we use has the benefit of yielding a very transparent set of axioms.

support of the lottery.

3.1 Basic Axioms

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: $\alpha \circ x + (1-\alpha) \circ y = (\alpha_0 \circ x + (1-\alpha) \circ y_0, \cdots, \alpha \circ x_T + (1-\alpha) \circ y_T)$.

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.*

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 $t < t'$, the preference $p^t \succsim p^{t'}$ holds.
(d) (C-Monotonicity) For all $c, c' \in C$, $c \geq c' \Longleftrightarrow c \succsim c'$.
(e) (Monotonicity) For any $x, y \in X$, if $(x_t, 0, \cdots, 0) \succsim (y_t, 0, \cdots, 0)$ for all t then $x \succsim y$. 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 $\alpha \in (0, 1]$, if $p \succ p'$ then $\alpha \circ p + (1-\alpha) \circ p'' \succ \alpha \circ p' + (1-\alpha) \circ p''$.
(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.

A key question is how to identify the discount function using behavior, especially considering that discount functions are stream dependent. Given that a temporal property like impatience should be behaviorally defined without reference to risk preferences, NT's

"MRS" approach uses marginal rates of substitution. We follow their "lottery approach" here, creating a gap between our discussion below and that in the deterministic setting of Section 2.3: instead of defining a magnitude effect by a change in impatience with respect to the scaling of consumption, it will be defined with respect to the scaling of the probability of consumption. This can be rectified by restricting the utility from consumption to have the power form, in which case scaling consumption corresponds to scaling the utility from consumption. However, the main issue is of establishing the existence of an additively separable representation for a subdomain $X_{ms} \subset X$ defined below. The classic approach to obtaining additive separability in the deterministic setup requires X_{ms} to have a product structure (see Koopmans [21]), which generically does not hold in the model. We therefore formulate a condition in Section 3.1.2 that exploits the vNM Independence property for risk preferences imposed by Regularity.

3.1.1 Magnitude-Sensitivity

For any $p \in \Delta$ and $\alpha \in [0,1]$ define the mixture $\alpha \circ p := \alpha \circ p + (1-\alpha) \circ 0$. 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 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), leading to the preference $\alpha \circ c_x \succsim \alpha \circ x$. 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:[8]

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

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

[8]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).

NT define magnitude-independent impatience as *Homotheticity,* which requires that $\alpha \circ c_x \sim \alpha \circ x$ holds for all x and α. They define magnitude-decreasing impatience by *Weak Homotheticity*, which requires $\alpha \circ c_x \succsim \alpha \circ x$ for all x and α. We maintain Weak Homotheticity but impose structure on how impatience changes with scaling. A similar condition is used by NT.

Axiom 2 (Magnitude-Sensitive (MS) Homogeneity) *(i) For any $x \in X$ and any $\alpha \in (0,1)$, $c_x \sim x \Longrightarrow \alpha \circ c_x \succsim \alpha \circ x$.*

(ii) For any magnitude-sensitive dated rewards $p^t, q^s \in X$, their present equivalents $c_{p^t} \sim p^t$ and $c_{q^s} \sim q^s$, and any $\alpha, \beta \in (0,1)$, $\beta \circ c_{p^t} \sim \alpha \circ p^t \Longrightarrow \beta \circ c_{q^s} \sim \alpha \circ q^s$.

The second part of 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. This will yield the power form for cost functions.

3.1.2 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 now define Separability and Non-Separability behaviorally.

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 $\frac{1}{2}$- mixtures of present equivalents of $(10, 10, 0)$ and $(0, 0, 10)$ must be indifferent to that of $(10, 10, 10)$ and $(0, 0, 0)$:[9] For all $x \in X$ and all $t > 0$,

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

[9] As noted earlier, in a deterministic setting, the Separability condition in Koopmans [21] 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. A vNM Independence condition could be used in the richer domain of *lotteries over streams.* In order to avoid adopting a richer domain, we formulate a Separability condition that uses mixtures of certainty equivalents.

As suggested in Section 3.2, 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 given by the following *Weak Separability* condition: 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 $t > 0$,*

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

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 $x \in X$, if x is not separable then it is not magnitude-sensitive. In order to avoid stating negatives, we write this key condition in its contrapositive form:

Axiom 3 (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:

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

Immediate rewards cannot be magnitude-sensitive, and so any $x \in X_{ms}$ must be non-trivial in the sense that $x_t \succ 0$ for some $t > 0$. The nontrivial streams outside of X_{ms} 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 nontrivial streams outside of X_{ms} as those at which the constraint is binding.[10]

3.1.3 Threshold Condition

Our final axiom places structure on X_{ms}. It requires that for any nontrivial x outside of X_{ms}, its future consumption can always be "worsened" sufficiently (wrt preference) that it becomes a nontrival y in X_{ms}. Conversely, if x is in X_{ms}, then any such worsening leaves it in X_{ms}. Formally: For any $x \in X$, let $(0, x_{-0})$ denote the stream that pays 0 in period 0 and pays according to x from period 1 onward, that is, $(0, x_{-0}) = (0, x_1, \cdots, x_T)$. We impose:

Axiom 4 (Preference-Based Threshold (PBT)) *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 $(0, x_{-0}) \succsim (0, y_{-0})$.
(ii) if $x \in X_{ms}$ and y satisfy $(0, x_{-0}) \succsim (0, y_{-0})$, 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 stream. 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 [4], 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.

3.2 Representation Result

We show that:

Theorem 1 *A preference $\succsim$ on X satisfies Regularity, MS Homogeneity, MS Separability, and PBT if and only if it admits a Homogenous CCE representation.*

[10] 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.

Although we did not impose Weak Separability in the main theorem, it is in fact implied by the other axioms.[11]

In the literature, Wakai [37] is the closest to this paper in that it features both stream-dependent discounting and non-separability. Though the model is defined recursively, the representation is rewritten as the minimum discounted utility over a set $\mathbb{D}$ of discount functions such as

$$U(x) = \min_{D \in \mathbb{D}} \sum_{t \geq 0} D(t)u(x_t).$$

Thus, the model has the same structure as the maxmin expected utility (Gilboa and Schmeidler [18]), with discount factors playing the role of priors. Under this representation, quasi-concavity of preferences over consumption streams can be interpreted as aversion to intertemporal fluctuations, which corresponds to preference for spreading consumption across periods. Moreover, the model satisfies a "Certainty Independence" property that implies that the ranking of two streams does not change if those streams are mixed with a constant consumption stream period by period.

In contrast, our model generates a different structure. Consider the case where the cognitive constraint binds, that is, $\sum_{t \geq 1} \varphi_t(D_x(t)) = K$. Since the cost term is constant, the cognitive optimization problem is reduced to

$$D_x = \arg\max_{D \in \mathbb{D}_K} \sum_{t \geq 1} D(t)u(x_t),$$

where $\mathbb{D}_K = \{D | \sum_{t \geq 1} \varphi_t(D(t)) = K\}$. Thus, in this subdomain, our representation is written as $U(x) = \max_{D \in \mathbb{D}_K} \sum_{t \geq 0} D(t)u(x_t)$. This representation therefore has a maxmax form rather than a maxmin form. At the level of utility streams, this structure implies the opposite of a preference for spread, rather a preference for concentration.[12]

The two models also differ in their implications for magnitude effects. In our model, the effective discounting depends on the stakes of streams through the cognitive optimization. In Wakai [37]'s model, by contrast, the discount function applied to a stream may vary across streams but do not systematically depend on the stake size. For example, for a dated reward p^t, the same discount factor is applied regardless of the magnitude of the

[11] This follows from Proposition 2 where it was shown that $U(x) \leq U(0_\tau, x_{-\tau}) + U(x_\tau, 0_{-\tau})$ and the obvious facts that $U(\frac{1}{2} \circ c_x + \frac{1}{2} \circ c_0) = \frac{1}{2}U(x) + \frac{1}{2}U(0) = \frac{1}{2}U(x)$ and $U(\frac{1}{2} \circ c_{(x_\tau, 0_{-\tau})} + \frac{1}{2} \circ c_{(0_\tau, x_{-\tau})}) = \frac{1}{2}U(x_\tau, 0_{-\tau}) + \frac{1}{2}U(0_\tau, x_{-\tau})$.

[12] At the level of consumption, however, the curvature of u also matters, and hence, consumption smoothing may still arise if u is sufficiently concave. See Noor and Takeoka [28] for details.

payoff. At the axiomatic level, this difference is closely related to homotheticity. Our model replaces homotheticity with a weaker condition (MS Homogeneity), which plays a key role for identifying magnitude effects. On the other hand, the Certainty Independence axiom imposed by Wakai [37] implies homotheticity, thereby excluding magnitude effects.

The uniqueness properties of the representation are given by:

Theorem 2 *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)* $K^2 = \alpha K^1$*.*

The sharp uniqueness is due to the separability of the representation on the subdomain X_{ms} and the fact that $u(0) = 0$ is presumed in the representation.

We close with a proof sketch of the sufficiency of Theorem 1, and a more general discussion of the foundations for the CCE model. We first establish an additively separable representation $U(x) = \sum_{t\geq 0} U_t(x_t)$ on X_{ms}. NT used Regularity and the separability axiom to establish an additive separable utility function on a product space. Although X_{ms} is not a product space, similar techniques can be applied using the PBT axiom. This additive separable representation can be rewritten as a discounted utility with stream-dependent discount function simply by defining $u(p) = U_0(p)$ and $D_{u(x_t)}(t) = \frac{U_t(x_t)}{u(x_t)} = \frac{V_t(u(x_t))}{u(x_t)}$. Given u and $D_{u(x_t)}$, we can use the cognitive first order condition $u(x_t) = \varphi_t'(D_{u(x_t)}(t))$ for each $t > 0$ to derive an additive cost function $\varphi = \sum \varphi_t$ for which $x \mapsto D_x$ is optimal. Given the first order condition and magnitude sensitibity of $x \in X_{ms}$, each φ_t must be increasing and convex since $D_{u(x_t)}$ is increasing in $u(x_t)$. This establishes that the preference over X_{ms} has an unconstrained CCE representation.

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 (the first part of MS Homogeneity) and PBT 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 third step is to infer a general capacity constraint K_x, depending on a stream 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 $\{\lambda \circ x \mid \lambda > 0\}$, 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 $\lambda \circ x$ 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 constant K. In explaining this step, we first note what a characterization of the CCE model would entail. For any stream x on the boundary of X_{ms}, denoted by $bd(X_{ms})$, the first order condition, together with integration by parts, implies that:

$$\begin{aligned} K_x = \varphi(D_x) = \sum_{t \geq 1} \varphi_t(D_x(t)) &= \sum_{t \geq 1} \int_0^{u(x_t)} \varphi_t'(D_r(t)) \frac{\mathrm{d}D_r}{\mathrm{d}r}(t)\,\mathrm{d}r = \sum_{t \geq 1} \int_0^{u(x_t)} r \frac{\mathrm{d}D_r}{\mathrm{d}r}(t)\,\mathrm{d}r \\ &= \sum_{t \geq 1} \left[u(x_t) D_{u(x_t)}(t) - \int_0^{u(x_t)} D_r(t)\,\mathrm{d}r \right] = U(0, x_{-0}) - \sum_{t \geq 1} \int_0^{u(x_t)} D_r(t)\,\mathrm{d}r. \end{aligned} \tag{6}$$

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 $x, y \in bd(X_{ms})$,

$$U(0, x_{-0}) - \sum_{t \geq 1} \int_0^{u(x_t)} D_r(t)\,\mathrm{d}r = U(0, y_{-0}) - \sum_{t \geq 1} \int_0^{u(y_t)} D_r(t)\,\mathrm{d}r.$$

Although this technical restriction can in principle be expressed behaviorally[13], it does not translate into economically interesting behavior. This suggests that the "constant K"

[13]The expression (6) 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.

feature of CCE may not be of particular economic interest.

Nevertheless, the Homogeneous CCE model can be characterized by transparent behavioral postulates. MS-Homogeneity (ii) implies that φ_t admits a CRRA form with a power index m. Then, $\varphi = \sum \varphi_t$ is homogeneous of degree m, then Euler's Theorem, $m\varphi(D) = \sum \frac{\partial \varphi}{\partial D_t} D_t$, and the cognitive first order condition yield that the expression (6) can be written as

$$K_x = \varphi(D_x) = \frac{1}{m} \sum_{t \geq 1} \varphi_t'(D_{u(x_t)}(t)) D_{u(x_t)}(t) = \frac{1}{m} \sum_{t \geq 1} u(x_t) D_{u(x_t)}(t) = \frac{1}{m} U(0, x_{-0}). \quad (7)$$

But PBT imposes that $U(0, x_{-0}) = U(0, y_{-0})$ 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 $\lambda_x \circ x$.

4 Dynamic Implications: Consumption-Savings

We study properties of the CCE model in a consumption-savings problem. Below we use the terms "sophisticated" and "naive" in the sense of O'Donoghue and Rabin [31].

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 $t = 0, 1$.

4.1 Comparative Statics: Cognitive Capacity

Assume that both selves are cognitively 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 4 *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.

4.2 Age-Decreasing Impatience

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 [6]). A plausible explanation is that discount factors are age-dependent, and in particular increasing with age (Kureishi et al. [23]). 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 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.

This observation suggests that 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 5 *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 $D^{self\ 0}(2)$ may not be at this maximum since she has to spend resources on $D^{self\ 0}(1)$ 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 periords – 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 $D_{x+\epsilon^T}(t) \leq D_x(t)$ 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 6 *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 $D_{x+\epsilon^\tau}(t) \leq D_x(t)$ 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 $t \neq \tau$.

5 Dynamic Implications: Task Completion

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 where each self shares the same K.

5.1 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 7 *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.

Cyclical effort patterns given in Proposition 7 are consistent with the observation in Burger et al. [8]. In their experiment, subjects allocate effort across multiple periods in order to complete a long task. This allocation problem can be interpreted as choosing when to complete multiple subtasks over time. They report substantial fluctuations in effort over time, with periods of activity followed by inactivity. Such cyclical patterns of effort are qualitatively consistent with the implications of our model, in which limited cognitive capacity can generate alternating periods of task completion and inactivity.

The beta-delta model (for instance, O'Donoghue and Rabin [31]) can account for procrastination in the timing of task completion. However, because such a model is also additively separable across time as the CE model, the presence of additional tasks does not reduce the attractiveness of completing a current task.[14] In contrast, in the CCE model, future rewards compete for cognitive resources, so that the presence of many desirable future tasks may crowd out current effort.

Cyclical effort patterns can also arise in models with a stock of willpower (for instance, Ozdenoren et al. [32]). In such models, exerting effort depletes a stock of willpower, and hence, periods of effort are followed by inactivity until the stock recovers. In contrast, the mechanism in our model is forward-looking. Cycles arise not from past effort but from competition for cognitive resources across future rewards. When many future rewards are present, allocating attention to them may crowd out the evaluation of the current task. This difference leads to further behavioral implications, which we explore next.

5.2 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 8 *Suppose that there is only one task, and that the agent would complete it at $t = 0$ 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

[14] A similar limitation arises in Chakraborty [9]. Although discounting may depend on consumption, the preference remains additively separable across periods, and hence, additional future tasks do not crowd out current effort.

deadline is long then there are more such small benefits to be accrued, and these detract from the benefit of doing the task.

Experimental evidence also suggests that excessive flexibility in deadlines may lead to procrastination. Ariely and Wertenbroch [2] show that students perform worse when only a final deadline is imposed than when intermediate deadlines are required. While their interpretation relies on self-control problems, our model provides a different explanation based on limited cognitive capacity.

In the beta-delta models such as O'Donoghue and Rabin [31], extending the deadline does not necessarily generate additional procrastination. In particular, if an agent completes the task immediately when the deadline is short, extending the deadline does not change the optimal timing of completion. In models with a stock of willpower (Ozdenoren et al. [32]), procrastination arises because exerting self-control today depletes a stock that may be needed to resist future temptations. In the single-task enviromment considered here, however, there is no reason to preserve willpower for the future, and hence, extending the deadline does not generate addtional delay. In contrast, in our model, a longer deadline may induce procrastination because future opportunities to complete the task competes for limited cognitive resources.

Our result also contrasts with the implications of Wakai [37] in the task completion environment considered here. In the stationary version of his model, extending the deadline does not generate procrastination in this setting. If a task is completed immediately under a short deadline, the same remains true when the deadline is extended. A non-stationary version of the Wakai model may generate delay, but through a different mechanism based on time-varying discount bounds. A detailed analysis of the implications of the Wakai model for task timing in this environment is provided in the supplementary appendix [29].

6 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. Consider the following examples.

Empirical studies show that poverty is correlated with lower cognitive abilities (Mani et al. [26]). This suggests that exerting self-control or tolerating low consumption may use up resources that would otherwise be used to think about the future. An extension of the CCE model is therefore to allow the effective cognitive capacity available for evaluating future consumption streams, denoted by K_{x_0}, to depend on the current consumption x_0. In particular, one may assume that K_{x_0} is increasing in x_0, reflecting that resisting current temptations consumes cognitive resources. Such a specification is consistent with willpower-based models in which self-control relies on a limited resource. It may also help explain why individuals facing poverty sometimes engage in seemingly myopic consumption choices (e.g. overspending on alcohol, festivals and underspending on food and education, as discussed in Banerjee and Duflo [3]). These behaviors may arise because such activities help bolster cognitive resources, rather than because low cognitive resources make them myopic.

There are two motivating examples for thinking of more general (nonseparable) cost functions over discount functions. First, experimental studies (such as Read and Scholten [35]) suggest that adding a particular future reward to a stream may reduce impatience towards other future periods. This suggests some complementarity across periods, which is ruled out by an additive cost function. A simple cost function that gives rise to such a complementarity is a Leontief-type function given by $\varphi(D) = \max_t \varphi_t(D(t))$.

Second, Ebert and Prelec [12] show that under high cognitive load, discount functions tend to be insensitive to time, that is, $D(t) \approx D(t')$ for all $t, t' > 0$. Under low cognitive load, discount functions are more downward sloping. Picturing a downward sloping D becoming flat after increasing cognitive load, we see the curious implication that $D(t)$ for distant t in fact increases as cognitive load increases. An intuitive explanation is that when resources are low the agent resorts to a (flat) default discount function, and is more discriminating between periods with higher resources (Ebert and Prelec [12], Enke and Graeber [14]). To model this, the additivity of the cost function in the CCE model must be sacrificed: instead of choosing each $D(t)$ at a cost, the agent chooses discount functions D at a cost. The cheaper discount functions would be constant $t \mapsto D(t) = \delta$ with cost increasing in $|\delta - \delta_d|$, where δ_d is a default discount factor. More flexible discount functions would have higher cost.

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 et al. [30] 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 $\varphi_t'(\underline{d}_t) = 0$, it must be that if $u(x_t) = 0$, then we have the solution $D_x(t) = \underline{d}_t$. If $u(x_t) > 0$, then $D_x(t) > \underline{d}_t$. Since φ_t is strictly convex on $d \geq \underline{d}_t$, it must be that the solution $D_x(t)$ is strictly increasing in $u(x_t) > 0$: the FOC satisfies $u(x_t) = \varphi_t'(D_x(t))$ for every $t > 0$ and so we have $D_y(t) \leq D_x(t)$ for any $y \quad x$ with strict inequality for all $t > 0$ s.t. $u(y_t) > 0$.

To establish the second claim observe that when the constraint is binding for stream $x \notin X_s$, it must be that the optimal D_x satisfies

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

Take any $y \quad x$. Denote $\varphi(D) = \sum_{t \geq 1} \varphi_t(D_t)$. It cannot be that $D_y \quad D_x$ since we obtain the contradiction that $\varphi(D_y) > \varphi(D_x) = K$. Suppose by way of contradiction that $D_y \quad D_x$. Then, $\varphi(D_y) < \varphi(D_x) = K$, which means that $y \in X_s$. Hence, $u(y_t) = \varphi_t'(D_y(t))$ for all $t > 0$. Since $u(y_t) \geq u(x_t)$ and φ_t is strictly convex on $d_t \geq \underline{d}_t$, there exists some $\tilde{D}_x(t)$ such that $u(x_t) = \varphi_t'(\tilde{D}_x(t))$ as long as $u(y_t) > 0$. Since φ_t' is increasing, we have $D_y \geq \tilde{D}_x$. Since φ is weakly increasing, $\varphi(\tilde{D}_x) \leq \varphi(D_y) < K$. Therefore, $\tilde{D}_x$ is an unconstrained optimal discount function for the stream x, and hence, $x \in X_s$, which is a contradiction.

Next take any $y \quad x$ such that $\frac{u(y_t)}{u(y_s)} = \frac{u(x_t)}{u(x_s)}$ for all $t, s > 0$. Then $\frac{\varphi_t'(D_y(t))}{\varphi_s'(D_y(s))} = \frac{u(y_t)}{u(y_s)} = \frac{u(x_t)}{u(x_s)} = \frac{\varphi_t'(D_x(t))}{\varphi_s'(D_x(s))}$. Note that since φ_t' is strictly increasing, these ratios imply that if $D_y(t) >$ (resp $<$)$D_x(t)$ for some $t > 0$ then $D_y(s) >$ (resp $<$)$D_x(s)$ for all $t > 0$. Since it can never be that $D_y \quad D_x$ nor $D_y \quad D_x$, we conclude that $D_y = D_x$.

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 (2) 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 $U(x) = U(x_t, 0_{-t}) + U(0_t, x_{-t})$.

Next take $x \notin X_s$. Consider the discounted utilities: $U(x) = u(x_0) + \sum_{t\geq 1} D_x(t)u(x_t)$, $U(x_t, 0_{-t}) = D_{(x_t,0_{-t})}(t)u(x_t)$ and $U(0_t, x_{-t}) = \sum_{t'\neq t} D_{(0_t,x_{-t})}(t')u(x_{t'})$. The first order condition for D_x requires that for all t, t' where $u(x_t), u(x_{t'}) > 0$,

$$\frac{u(x_t)}{u(x_{t'})} = \frac{\varphi_t'(D_x(t))}{\varphi_{t'}'(D_x(t'))},$$

and since $D_x(t) = \underline{d}_t$ and $\varphi_t(\underline{d}_t) = 0$ whenever $u(x_t) = 0$, we can write the binding constraint as $\sum_{t|u(x_t)>0} \varphi_t(D_x(t)) = K$. 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 $U(x) \leq U(x_t, 0_{-t}) + U(0_t, x_{-t})$.

C Appendix: Proof of Theorem 1

Necessity is verified in the supplementary appendix. We show sufficiency here. We proceed in steps. 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. We refer to the first part of MS Homogeneity as Weak Homotheticity.

C.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. C-Monotonicity implies that the restriction of u on C is strictly increasing. The proof of the remaining claims is provided in the online appendix of NT. ■

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. $U(x) = u(x_0) + \sum_{t\geq 1} U_t(x_t)$ for all $x \in X_{ms} \cup \Delta_0$, 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, it follows that $u(c_{(x_s,0_{-s})}) + u(c_{(x_t,0_{-t})}) = u(c_{(x_t,x_s,0_{-t,-s})}) + u(0)$, which is equivalent to $U(x_s, 0_{-s}) + U(x_t, 0_{-t}) = U(x_t, x_s, 0_{-t,-s})$. Define $U_t(x_t) = U(x_t, 0_{-t})$ and $U_s(x_s) = U(x_s, 0_{-s})$. Then, we have

$$U(x_t, x_s, 0_{-t,-s}) = U_t(x_t) + U_s(x_s). \tag{8}$$

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 (8). Since x is separable, $\frac{1}{2} \circ c_{(x_r,0_{-r})} + \frac{1}{2} \circ c_{(x_t,x_s,0_{-t,-s})} \sim \frac{1}{2} \circ c_{(x_t,x_s,x_r,0_{-t,-s,-r})} + \frac{1}{2} \circ 0$. Since u is mixture linear, $u(c_{(x_r,0_{-r})}) + u(c_{(x_t,x_s,0_{-t,-s})}) = u(c_{(x_t,x_s,x_r,0_{-t,-s,-r})}) + u(0)$ and in particular $U(x_r, 0_{-r}) + U(x_t, x_s, 0_{-t,-s}) = U(x_t, x_s, x_r, 0_{-t,-s,-r})$. Define $U_r(x_r) = U(x_r, 0_{-r})$. Then, $U(x_t, x_s, x_r, 0_{-t,-s,-r}) = U_r(x_r) + U(x_t, x_s, 0_{-t,-s}) = U_t(x_t) + U_s(x_s) + U_r(x_r)$. Repeating the argument finitely many times, we obtain $U(x) = \sum_{t\geq 0} U_t(x_t)$, 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 $U(x) = u(x_0) + \sum_{t\geq 1} D_{u(x_t)}(t)u(x_t)$, 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 $U(x) = u(x_0) + \sum_{t\geq 1} D_{u(x_t)}(t)u(x_t)$, 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 $D_{u(p)}(t) > D_{\alpha u(p)}(t)$ 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 $S_t := \{d \in [0,1] \,|\, d = D_{u(p)}(t)$ for some $p^t \in X_{ms}\}$. By Monotonicity 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 $U(x) = u(x_0) + \sum_{t\geq 1} D_x(t)u(x_t)$ subject to $D_x = \arg\max_D\{\sum_{t\geq 1}\big(D(t)u(x_t) - \varphi_t(D(t))\big)\}$, 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

$$u(p) = \varphi_t'(D_{u(p)}(t)). \tag{9}$$

Since $D_{u(p)}(t)$ is strictly increasing and continuous in $u(p)$, (9) 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. ■

C.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*

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

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,*

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

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

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

If $x \in X_{ms}$, $\widetilde{\alpha}_x = 1$ satisfies this condition. Thus, assume $x \notin X_{ms}$. Let $\widetilde{A} = \{\alpha \in (0,1] \,|\, \alpha \circ c_x \succ \alpha \circ x\}$. 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. ■

Recall that Lemma 5 showed that $\alpha_x = \sup\{\alpha \in [0,1] \,|\, \alpha \circ x \in X_{ms}\}$ for any $x \in X \setminus \Delta_0$.

Lemma 8 *The function $U : X \to \mathbb{R}_+$ appeared in Section C.1 can be written as $U(x) = u(x_0) + \sum_{t \geq 1} D_x(t) u(x_t)$ subject to*

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

Proof. By the result of Section C.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,

$$U(x) = u(c_x) = \frac{1}{\alpha_x} U(\alpha_x \circ x). \tag{10}$$

By the representation on X_{ms},

$$U(\alpha_x \circ x) = u(\alpha_x \circ x_0) + \sum_{t \geq 1} D_{\alpha_x \circ x}(t) u(\alpha_x \circ x_t). \tag{11}$$

By combining (10) with (11),

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

as desired. ■

Next we derive a stream-dependent capacity constraint function $K : X \setminus \Delta_0 \to \mathbb{R}_{++}$ which we will later show to be constant. 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 $U(x) = u(x_0) + \sum_{t \geq 1} D_x(t) u(x_t)$ subject to $D_x = \arg\max_{D \in \Lambda_x} \{\sum D(t) u(x_t) - \varphi_t(D(t))\}$, where $\Lambda_x := \{D \in [0,1]^T \mid \varphi(D) \leq K_x\}$. 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 $X^{\alpha_{\bar{x}}} = \{y \in X \mid y_t \succ \alpha_{\bar{x}} \circ \bar{x}_t,\ \forall t\}$. 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 $\Lambda_x := \{D \in [0,1]^T \mid \varphi(D) \leq K_x\}$. From Lemma 8, for any $x \in X_{ms}$ we have $U(x) = u(x_0) + \sum_{t\geq 1} D_x(t)u(x_t)$ subject to $D_x = \arg\max_D\{\sum D(t)u(x_t) - \varphi_t(D(t))\}$. 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

$$D_{\alpha_x \circ x} \in \arg\max_{D\in\Lambda_x} D \cdot u(x). \tag{12}$$

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 $D \neq D_{\alpha_x \circ x}$,

$$\begin{gathered}
D_{\alpha_x \circ x} \cdot u(\alpha_x \circ x) - \varphi(D_{\alpha_x \circ x}) > D \cdot u(\alpha_x \circ x) - \varphi(D) \\
\implies D_{\alpha_x \circ x} \cdot u(\alpha_x \circ x) - D \cdot u(\alpha_x \circ x) > \varphi(D_{\alpha_x \circ x}) - \varphi(D) \\
\implies \alpha_x[D_{\alpha_x \circ x} \cdot u(x) - D \cdot u(x)] > \varphi(D_{\alpha_x \circ x}) - \varphi(D) \\
\implies D_{\alpha_x \circ x} \cdot u(x) - D \cdot u(x) > \varphi(D_{\alpha_x \circ x}) - \varphi(D) \text{ (since } D_{\alpha_x \circ x} \cdot u(x) \geq D \cdot u(x)) \\
\implies D_{\alpha_x \circ x} \cdot u(x) - \varphi(D_{\alpha_x \circ x}) > D \cdot u(x) - \varphi(D).
\end{gathered}$$

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 $K_x = K_y$. ∎

C.3 Properties of the boundary of X_{ms} and of $\overline{D}$

Define

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

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 $\overline{D}(t)\varphi_t'(\overline{D}(t)) = \overline{v}$ *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 13 $\overline{D}(t+1) \leq \overline{D}(t)$ *for all* $0 < t < T$.

Proof. 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 Impatience, $\bar{p}^t \succsim \bar{p}^{t+1}$, 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 12, $\overline{D}(t)\varphi_t'(\overline{D}(t)) = \overline{v}$ 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. ■

C.4 Homogeneity of φ_t

By 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

$$R_{ms}(t) = \{r \,|\, r = u(p) \text{ for some } p^t \in X_{ms}\}, \tag{13}$$

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 $a_t > 0$,

$$\varphi_t(d) = a_t d^m. \tag{14}$$

Since $\succsim$ satisfies Magnitude-Sensitive Homogeneity, by the same proof of Theorem 7 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 (14). For the convenience of the reader, we reproduce the relevant lemmas here, albeit with some proofs omitted.

Lemma 14 *For each $\alpha \in (0,1]$ there is $h(\alpha)$ s.t. $D_{\alpha u(p)}(t) = h(\alpha) D_{u(p)}(t)$ for any $p^t \in X_{ms}$. Moreover, $h(\alpha)$ is a continuous function on $(0,1]$ with $h(1) = 1$.*

Lemma 15 *Define $\overline{R}(t) := [0, \overline{r}_t]$. There is $\theta \in \mathbb{R}$ s.t. $D_{\alpha r}(t) = \alpha^\theta D_r(t)$ for any $t > 0$, $r \in \overline{R}(t)$ and $\alpha \in (0,1]$.*

Lemma 16 *For any $t > 0$, there exist $\theta > 0$ and $\kappa_t > 0$ such that $D_r(t) = \kappa_t r^\theta$ for all $r \in \overline{R}(t)$.*

Proof. Take any $r \in \overline{R}(t)$. Then $r \leq \overline{r}_t$. By Lemma 15, $D_r(t) = D_{\frac{r}{\overline{r}_t}\overline{r}_t}(t) = \left(\frac{r}{\overline{r}_t}\right)^\theta D_{\overline{r}_t}(t)$. 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 17 *There exists $a_t > 0$, $m > 1$ such that $\varphi_t(d) = a_t d^m$ for all $d \leq \overline{D}(t)$ and $t > 0$.*

Proof. By Lemma 16, $D_r(t) = \kappa_t r^\theta$ 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 18 *For all $x \in X_{ms}$, $\varphi(D_x) = \frac{1}{m} U(0, x_{-0})$.*

Proof. As shown in Lemma 17, 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,

$\varphi(D_x) = \sum_{t>0} \varphi_t(D_{x_t}(t)) = \sum_{t>0} a_t \left(\frac{u(x_t)}{m a_t}\right)^{\frac{m}{m-1}} = \frac{1}{m} \sum_{t>0} \left(\frac{u(x_t)}{m a_t}\right)^{\frac{1}{m-1}} u(x_t)$
$= \frac{1}{m} \sum_{t>0} D_{x_t}(t) u(x_t) = \frac{1}{m} U(0, x_{-0})$, 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 18,

$$K_x = K_{\alpha_x \circ x} = \varphi(D_{\alpha_x \circ x}) = \frac{1}{m} U(0, \alpha_x \circ x_{-0}) = \frac{\overline{v}}{m}, \tag{15}$$

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

From now on, let $K > 0$ be the constant number given by (15).

Lemma 19 *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, $u(p) = ma_t D_{u(p)}(t)^{m-1}$. Since p^t is magnitude sensitive, $D_{u(p)}(t) = \left(\frac{u(p)}{ma_t}\right)^{\frac{1}{m-1}} \leq \overline{D}(t)$. 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 $\bar{p}^t$, we have $K = a_t \left(\frac{u(\bar{p})}{ma_t}\right)^{\frac{m}{m-1}} = a_t(\overline{D}(t))^m$, as desired. ■

Lemma 20 *For all $t < T$, $a_{t+1} \geq a_t$.*

Proof. Lemma 19 implies $1/m \times \overline{D}(t) \times ma_t\overline{D}(t)^{m-1} = K$, or $\overline{D}(t)u(\bar{p}_t) = mK$, where $\bar{p}_t$ satisfies the FOC $u(\bar{p}_t) = ma_t\overline{D}(t)^{m-1}$. By Lemma 13, $\overline{D}(t+1) \leq \overline{D}(t)$. Thus, Lemma 19 implies $a_{t+1} \geq a_t$. ■

Given that $\overline{D}(t) \leq 1$, the previous two lemmas imply that in fact $K \leq a_1$. Moreover, 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 $[0, 1]$. This establishes all the desired properties for $(u, \{\varphi_t\}_{t\geq 1}, K)$.

D Proof of Theorem 2

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

Take two Homogeneous CCE representations for the preference. 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 $x = p^t$. As observed above, $D_x(t)$ 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 $\varphi_t^2 = \alpha\varphi_t^1$. In particular, $m_1 = m_2$ and $a_t^2 = \alpha a_t^1$ for all t.

In the supplementary appendix (Noor and Takeoka [29], Section 7), we defined X_m by

$$X_m = \{x \in X \,|\, \sum_{t\geq 1} \gamma(t)u(x_t)^{\frac{m}{m-1}} \leq mK\},$$

and showed that $X_m = X_{ms}$. Therefore, $X_{ms} = \{x \in X \mid \sum_{t\geq 1} \gamma^i(t)u^i(x_t)^{\frac{m}{m-1}} \leq mK^i\}$ for $i = 1, 2$. Since $\gamma^2(t)u^2(x_t)^{\frac{m}{m-1}} = (ma_t^2)^{-\frac{1}{m-1}}u^2(x_t)^{\frac{m}{m-1}} = (m\alpha a_t^1)^{-\frac{1}{m-1}}(\alpha u^1(x_t))^{\frac{m}{m-1}} = \alpha\gamma^1(t)u^1(x_t)^{\frac{m}{m-1}}$, we have $K^2 = \alpha K^1$, as desired.

References

[1] Andersen, S., G. Harrison, M. Lau and E. Rutström (2013): "Discounting Behavior and the Magnitude Effect: Evidence from a Field Experiment in Denmark," *Economica* 80, pp. 670–697.

[2] Ariely, D., and K. Wertenbroch (2002): "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Sciencel* 13, pp. 219–224.

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

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

[5] Bolte, L. and C. Raymond (2024): "Emotional Inattention", mimeo.

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

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

[8] Burger, N., G. Charness, and J. Lynham (2011): "Field and Online Experiments on Self-Control," *Journal of Economic Behavior & Organization* 77, pp. 393–404.

[9] Chakraborty, A. (2021): "Present Bias," *Econometrica* 89, pp. 1921–1961.

[10] Deck, C., and S. Jahedi (2015): "The Effect of Cognitive Load on Economic Decision Making: A Survey and New Experiments," *European Economic Review* 78, pp. 97-119.

[11] 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.

[12] Ebert, J. and D. Prelec (2007): "The Fragility of Time: Time-Insensitivity and Valuation of thre Near and Far Future," *Management Science* 53(9), pp. 1423-1438.

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

[14] Enke, B. and T. Graeber (2023): "Cognitive Uncertainty", *Quarterly Journal of Economics* 138, pp. 2021–2067.

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

[16] 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.

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

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

[19] Hershfield, H., T. Garton, K. Ballard, G. Samanez-Larkin and B. Knutson (2009): "Don't Stop Thinking About Tomorrow: Individual Differences in Future Self-Continuity Account for Saving," *Judgment and Decision Making* 4(4), pp. 280–86.

[20] Hershfield, H., D. Goldstein, W. Sharpe, J. Fox, L. Yeykalis, L. Carstensen and J. Bailensen (2011): "Increasing Saving Behavior Through Age-Progressed Renderings of the Future Self," *Journal of Marketing Research* 48, pp. S23-S37.

[21] 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.

[22] Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice," *Econometrica* 46, pp. 185–200.

[23] 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.

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

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

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

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

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

[29] Noor, J. and N. Takeoka (2023): "Supplementary Appendix for "Constrained Optimal Discounting"," Working paper.

[30] Noor, J., N. Takeoka, and J. Yang (2026): "Stochastic Magnitude Effect in Intertemporal Choice," mimeo.

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

[32] 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.

[33] Peters, J., and C. Büchel (2010): "Episodic Future Thinking Reduces Reward Delay Discounting through an Enhancement of Prefrontal-Mediotemporal Interactions," *Neuron* 66, pp. 138–148.

[34] Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan.

[35] 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.

[36] Shi, S., and L. G. Epstein (1993): "Habits and Time Preference," *International Economic Review* 34, pp. 61–84.

[37] Wakai, K. (2008): “A Model of Utility Smoothing”, *Econometrica* 76, pp. 137–153.