Constrained Optimal Discounting*

Jawwad Noor Norio Takeoka

May 22, 2020

Abstract

Noor and Takeoka [23] model time preference as a current self that incurs a cognitive cost of empathizing with her future selves. Their model unifies disparate well-known experimental findings. They provide behavioral foundations by exploiting the idea that higher stakes provide an incentive for the exertion of higher cognitive effort, which leads to changes in the agent's impatience with respect to the scale of outcomes. The present paper introduces the capacity of limited cognitive resources into the model and investigates its behavioral implications. We show that the behavioral content of limited cognitive resources lies in violations of time-separability.

1 Introduction

Evidence suggests that subjects in experiments are less impatient when dealing with larger rewards.[1] Noor and Takeoka [23] (henceforth NT) hypothesize that impatience may arise from a cognitive process where a larger reward incentivizes higher cognitive effort that gives rise to less impatience. Their model builds on three introspective observations. First, our knowledge of our future selves is not of the same quality as our knowledge of our

*A version of this paper was previously circulated as a part of the paper entitled "Impatience as Selfishness". Noor (the corresponding author) is at the Dept of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Email: jnoor@bu.edu. Takeoka is at the Dept of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. Email: norio.takeoka@r.hit-u.ac.jp. The authors are grateful to Andrew Ellis, Larry Epstein, Bart Lipman, Jay Lu, Samuel McClure, Marco Mariotti, Anna Rubinchik, Bruno Strulovici for very useful discussions, and to seminar participants at Boston University, Yokohama National University, HEC Paris, Queen Mary University of London, Caltech, UC San Diego, Otaru University of Commerce, Brown University, University of Almeria, Hitotsubashi University, Waseda University, Academia Sinica and the audience at RUD (University of Paris, Dauphine), Bounded Rationality in Choice (Queen Mary), SAET (IMPA, Rio de Janeiro), SAET (Faro, Portugal) and Asian Meeting of the Econometric Society (Doshisha University, Kyoto) for helpful comments. Noor gratefully acknowledges financial support from the NSF (grant SES-1627906). Takeoka gratefully acknowledges financial support from Grant-in-Aid for Scientific Research (18K01503). Last but certainly not least, we thank Kevin Cooke and Nick Saponara for expert research assistance. The usual disclaimer applies.

[1] See Fredrick et al [10] for a review of experiments that find a *magnitude effect*.

current self. In fact, our ability to appreciate the well-being of, say, our retired future self requires a process that is analogous to the process by which we appreciate the well-being of other people: we empathize by imagining ourselves in the other's shoes. Second, imagining ourselves as a future self is cognitively costly. Third, there exists a desire to connect with our future selves, whether out of some sense of moral responsibility or a sense of community with future selves coming from a shared identity.

Their model is a multiple selves model where the current self optimally allocates costly empathy across future selves. They take a (static) preference $\succsim$ over the set X consisting of finite horizon streams of lotteries over consumption as primitives. The main result of their paper provides behavioral foundations for the *Costly Empathy (CE)* representation, which is described by an instantaneous consumption utility u and an increasing and convex cognitive cost function φ_t for each t such that $\succsim$ is represented by the function $U : X \rightarrow \mathbb{R}_+$ defined by

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

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

To interpret, consider a consumption stream $x = (x_0, .., x_T)$. The period 0 self evaluates the value of the stream x via the discounted utility formula (1) where the discount function D_x depends on the stream. We interpret the discount function $D_x(t) \in [0, 1]$ as the current self's empathy for self t's instantaneous consumption, with higher values of $D_x(t)$ expressing a higher degree of connection with the future self. The discount function is a cognitive choice in this model, with the following elements:

- Empathy is a cognitively difficult task, involving the cost of imagining oneself in the other's shoes. The cost of any discount function D is assumed to be additive, $\sum_{t\geq 1} \varphi_t(D(t))$, where each φ_t is an increasing and convex cost function. Moreover, φ_t is increasing in t so that empathy costs are increasing with temporal distance.
- In the cognitive stage the agent's choice of discount function maximizes the objective function given in (2), that is, it maximizes the utility derived from the connection with future selves (discounted future utility) less cognitive costs of empathy.

A tractable special case of the CE model studied by NT is the Homogeneous CE model, which is defined by a family $\{\varphi_t\}$ of convex CRRA cost functions.

In the present paper, we extend the Homogeneous CE model to allow for limited cognitive capacity, in which case the cognitive optimization problem (2) is subject to the capacity constraint:

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

That is, the agent cannot produce empathy D that costs more than K. We analyze and provide behavioral foundations for this model, called the *Constrained Costly Empathy (CCE)* representation. The main behavioral difference from the CE model is that this model violates time-separability. Intuitively, the agent may need to trade-off empathy across different selves, rather than separately optimize empathy for each self as in the CE model. Consequently, the weight placed on time t by the optimal discount function may depend on the rewards available in periods $t' \neq t$.

As an intermediate result for the CCE representation, we also investigate a more general representation, called the General CCE representation, where the cost function φ_t, $t \geq 1$, is increasing and convex (and not necessarily CRRA) but the cognitive capacity K_x can vary with the consumption stream x being evaluated. We provide behavioral foundations for this class of representations as well.

This paper is a companion to NT, who add to the theoretical literature on magnitude-dependent discount functions (Noor [22], Baucells and Heukamp [3], Wakai [31], Epstein and Hynes [8]) and the multiple selves model (Strotz [28], Ainslie [1], Laibson [16], O'Donoghue and Rabin [26]).[2] We list below the contributions relative to NT and related literature:

1. NT show that the CE model can unifying various behavioral findings such as the magnitude effect, preference reversal (the common difference effect), concentration bias, etc and can also explain anomalies for the Life-cycle Hypothesis. They also note the related evidence on the role of cognitive abilities for time preference (Dohmen et al [6]). This paper shows that by allowing for cognitive capacity constraints, the model can also explain evidence pertaining to time non-separability, suggesting therefore that there may exist a relationship among the disparate anomalies.

2. The multiple selves model has been considered in psychology and economics and its key application in economics is for the study of self-control problems through the hyperbolic discounting model and its variants (Strotz [28], Ainslie [1], Laibson [16], O'Donoghue and Rabin [26]). It is recognized in the literature this model does not include an expression of the notion of "self-control", which entails effort to reduce the impact of urges on choice. Fudenberg and Levine [11] extend the multiple selves model by introducing a separate executive self that derives utility from the utility of a sequence of myopic short lived selves, and can change the preferences of the short lived selves at a utility cost. This intervention is interpreted as exertion of self-control. NT point out that the CE model also can be interpreted as a multiple selves model that exhibits self-control: by incurring the cost of empathizing with future selves, the current self is able to behave more patiently. The present paper adds to the self-control interpretation by hypothesizing a hard limit in one's capacity for self-control.

3. There are several models that incorporate subjective optimization, such as those of optimal expectations (Brunnermeier and Parker [5]), optimal contemplation (Ergin and Sarver [9]) and optimal attention (Ellis [7], Gabaix [12]). To our knowledge, constrained subjective optimization is considered only in a few papers in the literature of willpower, where the decision maker is assumed to resist to temptation within the constraint of limited

[2] While most models take the discount function as a given feature of preference, Becker and Mulligan [4] provide a model of endogenous impatience, where the discount function can be altered by investment in education, etc. NT also provide a model of endogenous discounting but since the investment decision in Becker and Mulligan [4] draws from the agent's physical budget constraint rather a cognitive budget constraint, their model does not overlap with NT's. See NT for more discussion.

willpower. Ozdenoren, Salant, and Silverman [27] consider the cake-eating problem with a fixed initial stock of willpower, which is depleted over time with exercising self-control. In a discrete setting, Masatlioglu, Nakajima, and Ozdenoren [20] 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 [17] consider a menus of lotteries setting and incorporate the cost of self-control. They show that the content of limited will-power in that setting lies in the violation of the vNM Independence axiom for menus of lotteries. The key finding in the present paper is that binding cognitive constraints express themselves behaviorally as violations of time-separability.

The remainder of the paper proceeds as follows. We overview the related literature in Section 1.1. Section 2 describes our basic framework and the CCE representation. Section 3 provides a behavioral foundation for the representation, and Section 4 investigates how the representation changes according to the changes of parameters of the model. Section 5 relates the model to empirical findings. Section 6 investigates the General CCE representation. All proofs are relegated to the appendices. Noor and Takeoka [25], a supplementary appendix to this paper, provide an extension of the General CCE representation which allows for negative payoffs in order to accommoda the decision maker cares about gains and losses from some reference point.

2 Constrained CE Model

2.1 Primitives

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

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

As a benchmark, we define:

Definition 1 (Discounted Utility Representation) *A Discounted Utility (DU) representation for a preference over X is given by*

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

where $D(t)$ is weakly decreasing in t and satisfies $D_r(t) = 1$ and u is continuous and mixture-linear and satisfies $u(0) = 0$.

A notable feature of the DU model is that the discount function evaluates time independently of the stream of rewards being evaluated. The CCE model relaxes such magnitude-independent discounting.

2.2 Functional Form

Say that a tuple $(u, \{\varphi_t\}_{t\geq 1}, K)$ is *regular* if
(i) $u : \Delta \to \mathbb{R}_+$ is continuous and mixture linear with increasing vNM utility index $u : C \to \mathbb{R}_+$ satisfying (a) $u(0) = 0$ and (b) unboundedness: $u(C) = \mathbb{R}_+$.
(ii) for each $t \geq 1$, the cost function $\varphi_t : [0, 1] \to \mathbb{R}_+$ takes the form

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

where $m > 1$, and $a_t > 0$ is increasing in t.
(iii) $0 < K \leq a_1$.

Condition (i) requires that the utility from consumption should have familiar properties. Condition (i)(a) is a normalization of u. Condition (i)(b) is needed to ensure the existence of a present equivalent of any stream $x \in X$ (see Section 3.1). Condition (ii) requires $\{\varphi_t\}$ to be a family of convex CRRA cost functions that represent the cost of cognitive effort of appreciating future consumption. The degree of appreciation of consumption at time t is given by the period t discount factor $d \in [0, 1]$. The idea that farther consumption is harder to appreciate is expressed by the fact that a_t is increasing with t. This condition ensures that

$$\varphi_t(d) \leq \varphi_{t+1}(d) \text{ for all } d \in [0, 1] \text{ and } 0 < t < T.$$

Condition (iii) introduces the stock K of cognitive resources. Note that the cost of fully appreciating period 1 consumption is $\varphi_1(1) = a_1$. The condition requires that the stock K is exhausted if the agent tries to fully appreciate period 1 consumption: $K \leq a_1$. By condition (ii), it follows that it is exhausted also if the agent tries to fully appreciate consumption in any future period, that is, $K \leq a_t$ for all t. Therefore, a full appreciation of two or more periods of consumption simultaneously is not feasible, although it may be possible to fully appreciate period 1 consumption if $K = a_1$ (and other t s.t. $a_t = a_1$).

Definition 2 (Constrained CE Representation) *A Constrained Costly Empathy (CCE) representation is a regular tuple $(u, \{\varphi_t\}, K)$ such that $\succsim$ is represented by the function $U : X \to \mathbb{R}_+$ defined by*

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

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

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

The homogeneous CE model, introduced by Noor and Takeoka [23], corresponds to $K = \infty$, which is ruled out by condition (iii) in the CCE model. The cognitive optimization problem has a unique solution given the strict convexity of the cost function.

The optimal discount function D_x is chosen subject to two constraints. The first is the *capacity constraint*:

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

where $\varphi(D) = \sum_{t\geq 1}\varphi_t(D(t))$. That is, it must cost at most K. The second constraint,

$$D(t) \leq 1,$$

is called the *boundary constraint.*

We now clarify a role of condition (iii), that is, $K \leq a_1$, of the regular tuple. This condition has two implications. First, it determines the maximum achievable period t discount factor, denoted by $\overline{d}_t$. If all the cognitive abilities are spent for period t discount factor, the capacity constraint implies $\varphi_t(D(t)) = a_t D(t)^m = K$. This equation determines the maximum achievable discount factor $\overline{d}_t$. That is,

$$\overline{d}_t := \left(\frac{K}{a_t}\right)^{\frac{1}{m}} \leq 1. \tag{4}$$

Since a_t is increasing in t, $\overline{d}_t$ is decreasing in t.

Second, the condition makes the model more tractable because meeting the capacity constraint becomes *sufficient* to meet the boundary constraint: when the capacity constraint is satisfied, the discount function satisfies, for any given t,

$$\varphi_t(D_x(t)) \leq \varphi(D_x) \leq K \leq a_1 \leq a_t = \varphi_t(1),$$

which implies $D_x(t) \leq 1$. The tractability comes from the fact that, both in constructing the model and in its applications, one can effectively ignore the boundary constraint.

Regarding the interpretation of the capacity constraint K, the CCE model should be viewed as one where there exists a cap K on how much of cognitive resources are used for *each* stream in her menu. This is different from a model where the agent has a limited pool, and on facing a menu of streams, decides how to optimally allocate these resources across future selves, giving rise to a menu-dependent (as opposed to stream-dependent) discount function. See the concluding section for some elaboration.

2.3 Reduced Form Representation

We explore some properties of the model here. Assume that $\succsim$ admits a CCE representation. As noted earlier, because of strict convexity of the cost function, the cognitive optimization problem has a unique solution. It is instructive to analyze how this optimal discount function changes along rays of streams. Given the additive separability of the cost function, it is useful to consider the special stream referred to as the *dated reward*: a stream that pays some consumption $p \in \Delta$ at time t and 0 otherwise, denoted $p^t \in X$. Since p is a lottery, it can be α-mixed with consumption 0 to yield a lottery $\alpha \circ p + (1-\alpha) \circ 0$, in which case we imagine that it is "scaled down" by a factor of α. For $\alpha \leq 1$, denote by

$$\alpha p^t$$

the dated reward that pays $\alpha \circ p + (1-\alpha) \circ 0$ in period t. For $\alpha > 1$, we use the same notation for any q that satisfies $q = \frac{1}{\alpha} \circ p + (1 - \frac{1}{\alpha}) \circ 0$, that is, αp^t is a "scaled up" version of p when $\alpha > 1$.

When evaluating αp^t, for all values of α for which the capacity constraint is slack the first order condition is given by:

$$u(\alpha \circ p) = \varphi_t'(D(t)),$$

and consequently the optimal discount function is given by:

$$D_{\alpha p^t}(t) = \left(\frac{u(\alpha \circ p)}{m a_t}\right)^{\frac{1}{m-1}} := \gamma(t) u(\alpha \circ p)^{\frac{1}{m-1}},$$

where $\gamma(t) = (m a_t)^{-\frac{1}{m-1}}$. In particular, $D_{\alpha p^t}$ is increasing in $u(\alpha \circ p)$. Intuitively, as the prospect of obtaining utility at t improves, the agent has more incentive to exert effort to overcome selfishness, which leads to a higher D. Her discount function therefore exhibits *magnitude-decreasing impatience* for some range of rewards.

We just saw that $D_{\alpha p^t}$ is increasing in α as long as the capacity constraint is lax. There exists some threshold α_{p^t} where the capacity constraint binds and the discount function ceases to increase in α, since the agent cannot allocate any more empathy. Therefore beyond a threshold, the discount function exhibits magnitude-independent impatience.

More generally, D_x satisfies the FOC of Lagrangian:

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

where $\lambda \geq 0$ is a Lagrange multiplier. By the FOC wrt $D(t)$, we have

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

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

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

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

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

By substituting back λ to $D_x(t)$, we have

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

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

Say that a stream x is *magnitude sensitive* if the capacity constraint is slack at the optimal D_x and *not magnitude sensitive* if it is binding. In the model, a stream x is magnitude sensitive if and only if scaling down leads to a magnitude effect: $D_{\alpha x} < D_x$ for any $\alpha < 1$. It is not magnitude sensitive otherwise. We use this observation in the sequel to identify magnitude-sensitive streams behaviorally.

A very useful feature of the model is that it admits a clean way of distinguishing magnitude sensitive and other streams in terms of the representation. For any stream x, $U(x) - u(x_0)$ is the discounted future utility achieved from x, which is interpreted as the continuation payoff or the future payoff from period 1 onward. In the CCE representation, a stream x is magnitude sensitive iff its future payoff $U(x)-u(x_0)$ is less than some threshold.[3] This is expressed in the next proposition.

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

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

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

This result reveals that when the magnitude of future payoffs are "small", the utility function is additively separable, and future utility from lotteries is a power transformation

[3]This property is reminiscent of Becker and Mulligan [4], that derive an observation about complementarity between time preference and future utilities. In our model, up to the threshold, impatience decreases in future payoffs, it achieves the minimum impatience at the threshold, and becomes invariant beyond that.

of immediate utility. Since $u(p)$ is an expected utility, risk preferences are unchanged over t. However, the parameter m will affect intertemporal substitution. For large magnitude of future payoffs, we find that the utility function is no longer additively separable. Future utility is evaluated using a concave aggregator.

One more remark is that the CCE representation *on large streams* can be interpreted as a maxmin-type (more precisely, maxmax-type) representation à la Gilboa and Schmeidler [13]. For large streams, (3) is binding, that is, $\varphi(D) = K$. Then, (2) reduces to

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

subject to $D \in \mathcal{D}_K := \{D \in [0,1]^T : \varphi(D) = K\}$. The CCE representation on the large streams can be written as

$$U(x) = D_x \cdot u(x) = \max_{D \in \mathcal{D}_K} D \cdot u(x), \tag{5}$$

that is, given each stream, the agent optimally chooses a discount function so as to maximize discounted utilities within the capacity constraint.[4]

It follows from (10) that a CCE representation satisfies convexity for large streams, while it is not necessarily the case for small streams. Nevertheless, we can show that a CCE representation must be star-shaped, $\alpha U(x) \geq U(\alpha x)$, which is a property weaker than convexity. Since our model violates convexity, it goes beyond models of convex preferences in the literature (such as Maccheroni et al [19]). See Noor and Takeoka [25] for more details.

3 Foundations

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.*

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

$$c_x \sim x.$$

Present equivalents will be used instrumentally below.

3.1 Basic Axioms

The following axiom is the same as in NT.

Axiom 1 (Regularity) *(a) (Order).* $\succsim$ *is complete and transitive.*
(b) (Continuity). For all $x \in X$, $\{y \in X : y \succsim x\}$ *and* $\{y \in X : x \succsim y\}$ *are closed.*

[4] To explain consumption smoothing in an intertemporal decision making, Wakai [31] studies an agent who minimizes discounted utilities over a set of discount functions.

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

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

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

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

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

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

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

(f) (Risk Preference). For any $p, p', p'' \in \Delta$ and $\alpha \in (0, 1]$,

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

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

$$c \succsim x.$$

Order and Continuity are standard. Impatience states that consumption is better when received sooner than later. 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.

NT also formulate the behavioral meaning of time separability in this setup. For notational convenience, for all streams x, $y \in X$ and all t, let $x\{t\}y$ denote the stream that pays according to x at t and according to y otherwise. They define

Axiom 2 (Separability) *For all $x \in X$ and all t,*

$$\frac{1}{2} \circ c_{x\{t\}0} + \frac{1}{2} \circ c_{0\{t\}x} \sim \frac{1}{2} \circ c_x + \frac{1}{2} \circ c_0.$$

We refer the reader to NT for an explanation of the indifference condition in the definition. As indicated in Section 2.3 and as we will see formally below, a key property of the CCE model is that it violates Separability.

3.2 Identifying Magnitude-Decreasing Impatience

To identify whether an agent exhibits magnitude-decreasing impatience (that is, greater patience towards larger rewards), we follow the lottery approach considered by NT.[5] For any $p \in \Delta$ and $\alpha \in [0,1]$ define the mixture $\alpha \circ p := \alpha \circ p + (1-\alpha) \circ 0$. In particular, for $c \in C$, we write this lottery as $\alpha \circ c$ in order to distinguish $\alpha \circ c \in \Delta$ with a deterministic consumption $\alpha c \in C$. For any stream $x = (x_0, ..., x_T)$ define

$$\alpha x := (\alpha \circ x_0, ..., \alpha \circ x_T).$$

Intuitively, the stream αx uniformly "scales down" the desirability of x in every period by increasing the chance that it yields 0. Abusing notation, write $\alpha \circ p$ for the stream $(\alpha \circ p, 0, ...0)$. 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. Then if impatience does *not* change in response to scaling down x by α, then it must be that:

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

since the scaling down affects the evaluation of consumption equally for the immediate reward and the stream. The behavioral content of magnitude-independent impatience is therefore:

Axiom 3 (Homotheticity) *For any $x \in X$ and any $\alpha \in (0,1)$,*

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

Indeed, NT prove that a preference over X satisfies Regularity, Separability and Homotheticity if and only if it admits a DU representation.

If Homotheticity is the behavioral meaning of magnitude-independent impatience, then magnitude-dependence of impatience can be defined in terms of its violations. Indeed, if the agent is more patient towards larger rewards, then she would exhibit:

$$c_x \sim x \Longrightarrow \alpha \circ c_x \succsim \alpha x \text{ for all } \alpha \in (0,1].$$

Intuitively, if scaling down the probability of receiving x by α makes the stream less desirable, then an increase in impatience would lead the stream to lose value faster than the immediate reward $\alpha \circ c_x$ (for which impatience is irrelevant). Consequently, the following behavioral condition captures the such magnitude-decreasing impatience:

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

Axiom 4 (Weak Homotheticity) *For any $x \in X$ and any $\alpha \in (0,1)$,*

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

Say that a stream $x \in X$ is *ℓ-Magnitude Sensitive* if the agent's impatience strictly reduces whenever the stream is made less desirable.

Definition 3 (ℓ-Magnitude-Sensitivity) *A stream $x \in X$ is ℓ-Magnitude Sensitive if*

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

The set of all ℓ-Magnitude Sensitive streams is denoted by $X_\ell \subset X$.

By vNM Independence, it is clear that immediate rewards are not ℓ-Magnitude Sensitive. That is, $\Delta_0 \cap X_\ell = \emptyset$.

3.3 Structure on X_ℓ

The characterization of the CCE model involves placing structure on X_ℓ.

As noted earlier (see Section 2.3), the cognitive capacity constraint can lead to a violation of Separability when it binds. In order to highlight this key feature, it is natural to impose Separability on the set of streams for which the constraint does not bind. That set is precisely X_ℓ, the set of streams that are ℓ-magnitude sensitive. Therefore we impose:

Axiom 5 (X_ℓ-Separability) *For all $x \in X_\ell$ and all t,*

$$\frac{1}{2} \circ c_{x\{t\}0} + \frac{1}{2} \circ c_{0\{t\}x} \sim \frac{1}{2} \circ c_x + \frac{1}{2} \circ c_0.$$

The next restriction is familiar from NT. While they require the condition to hold globally for all $x \in X$ in order to characterize the homogeneous CE representation, we require it only to hold on X_ℓ:

Axiom 6 (X_ℓ-Homogeneity) *For any $x, y \in X_\ell$ s.t. $x_0 \sim y_0 \sim 0$, their present equivalents $c_x \sim x$ and $c_y \sim y$, and any $\alpha, \beta \in (0,1)$,*

$$\beta \circ c_x \sim \alpha x \Longrightarrow \beta \circ c_y \sim \alpha y.$$

X_ℓ-Homogeneity places structure on homotheticity violations. It imposes the substantive simplification that if scaling down x by α is as good as scaling down its present-equivalent c_x by β, then β depends on α but not the stream. It is easy to see that, given vNM Independence, this axiom imposes homotheticity on X_ℓ, since for any $x, y \in X_\ell$ s.t. $x_0 \sim y_0 \sim 0$, it must be that $x \sim y \Longrightarrow \alpha x \sim \alpha y$.

For any $x \in X$, recall that $0\{0\}x$ denotes the stream that pays 0 in period 0 and pays according to x from period 1 onward. That is, $0\{0\}x = (0, x_1, \cdots, x_T)$. Intuitively, $0\{0\}x$ is interpreted as the future payoffs obtained from x.

Axiom 7 (X_ℓ-Monotonicity) *For all $x \in X \setminus \Delta_0$, the following hold.*
(i) if $x \notin X_\ell$, then $y \in X_\ell$ for some $y \in X \setminus \Delta_0$ with $0\{0\}x \succsim 0\{0\}y$.
(ii) if $x \in X_\ell$, then $y \in X_\ell$ for all $y \in X \setminus \Delta_0$ with $0\{0\}x \succsim 0\{0\}y$.

This axiom requires that the magnitude sensitivity of a stream x should be associated with the consumption it delivers from period 1 onward ($0\{0\}x$). In particular, it should be associated with the "future utility" of the steam. Specifically, X_ℓ-Monotonicity (i) requires that if x is not magnitude sensitive, then there must exist a stream y with lower future utility that is magnitude-sensitive. X_ℓ-Monotonicity (ii) requires in addition that if x exhibits the magnitude sensitivity then so must every stream having smaller future utility.

This axiom shares a similar idea to Becker and Mulligan [4], who derive an observation about complementarity between time preference and future utilities. In their model, for example, when the wealth of the physical budget constraint expands, the agent invests more resources into the future consumption as well as the future oriented capital, which leads to decreasing impatience for large future utilities.

3.4 Representation Results

Our main result is:

Theorem 1 *A preference $\succsim$ on X satisfies Regularity, Weak Homotheticity, X_ℓ-Separability, X_ℓ-Homogeneity, and X_ℓ-Monotonicity if and only if it admits a CCE representation.*

NT prove that a preference on X satisfies Regularity, Weak Homotheticity, Separability, X_ℓ-Homogeneity, and X_ℓ-Regularity if and only if it admits a homogeneous CE representation.[6] The difference from Theorem 1 is that in NT's result, (i) Separability is assumed on the whole domain rather than on X_ℓ, and (ii) X_ℓ-Regularity is weaker than X_ℓ-Monotonicity.

Before providing a proof sketch, we note that the CCE model has strong uniqueness properties, inherited from the separability of the representation on the subdomain X_ℓ and because $u(0) = 0$ is featured in the representation.

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

The theorem ensures that, not only the cost function, but also the cognitive capacity K is uniquely derived from preference.

[6] X_ℓ-Regularity is even weaker than Strong X_ℓ-Regularity, introduced in Section 6 to characterize the General CCE representation. See Section 6.3 for more details.

3.5 Proof Outline

A proof sketch of sufficiency is as follows. As an intermediate result, we first establish a general representation (the General CCE representation), introduced in Section 6: There exists a basic tuple $(u, \{\varphi_t\}, K_x)$ such that $\succsim$ is represented by the function $U : X \to \mathbb{R}_+$ defined by

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

$$\text{s.t. } D_x = \arg \max_{D \in [0,1]^T} \sum_{t \geq 1} (D(t)u(x) - \varphi_t(D(t))) \text{ subject to } \sum_{t \geq 1} \varphi_t(D(t)) \leq K_x.$$

Here, the cost function φ_t, $t \geq 1$, is increasing and convex (not necessarily CRRA) and the cognitive resource is a function $K : X \to \mathbb{R}_{++} \cup \{\infty\}$ with several properties. An axiomatization of this general representation is investigated in Section 6. See the proof outline in Section 6.5 for more details.

The cognitive resource K_x can be computed via integrals of functions derived from behavior, but it is difficult to find an elegant axiom that would imply that K_x is a constant function. An observation is that K_x exactly coincides with the total empathy costs $\sum_{t \geq 1} \varphi_t(D_x(t))$ when the capacity constraint is binding. Note that x is magnitude-sensitive if and only if $D_x(t)$ is strictly increasing in x, which is in turn equivalent to the capacity constraint being slack at x. Therefore, the capacity constraint hits at the boundary of X_ℓ. To show K_x is constant, we will claim that $K_x = K_y$ for all streams x, y on the boundary of X_ℓ.

Here, X_ℓ-Homogeneity plays a key role. Given the General CCE representation, X_ℓ-Homogeneity restricts the cost function to be a homogeneous function, which implies that the cost function has a CRRA form. Moreover, as noted earlier in the context of the reduced form of the model (Proposition 1), a very convenient property of CRRA costs is that the total empathy costs $\sum_{t \geq 1} \varphi_t(D_x(t))$ associated with a magnitude-sensitive stream x is proportional to its future payoff $U(x) - u(x_0)$. In particular, when the capacity constraint is binding at x, we have

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

Since the capacity constraint is binding at the boundary of X_ℓ, it is enough to show that $U(x) - u(x_0) = U(y) - u(y_0)$ for all streams x, y on the boundary of X_ℓ.

Indeed, X_ℓ-Monotonicity plays the key role in identifying the boundary of X_ℓ. This axiom implies that the magnitude sensitivity of streams is associated with their future payoffs: if a stream gives a smaller future payoff, it tends to be magnitude sensitive. Since the current consumption does not matter for defining X_ℓ by Time-0 Irrelevance, X_ℓ is characterized as a lower contour set of $\succsim$ among all streams with $x_0 = 0$. In particular, the boundary of X_ℓ corresponds to an indifference curve of $\succsim$ among those streams. Therefore, $U(x) - u(x_0) = U(y) - u(y_0)$ for all x, y on the boundary of X_ℓ, as desired.

4 Limiting Cases

In this section, we investigate how the CCE representation changes according to changes of parameters $((a_t)_{t\geq 1}, m, K)$. We see that some models in the literature can be obtained as limiting cases of the CCE representation.

As pointed out in (4), the maximum achievable period t discount factor is defined by $\overline{d}_t = (K/a_t)^{\frac{1}{m}} \leq 1$. From this condition, a_t may be removed from the model as $a_t = K/\overline{d}_t^m$ for all $t \geq 1$. By substituting it into the cost function, we have

$$\varphi_t(d) = K\left(\frac{d}{\overline{d}_t}\right)^m. \tag{6}$$

Moreover, the capacity constraint is reduced to

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

Thus, K is independent of the capacity constraint and can be interpreted as a cost parameter. We can take $((\overline{d}_t)_{t\geq 1}, m, K)$ as the set of parameters of the CCE representation.

4.1 DU Model

Although the DU model is not nested to the CCE representation, we show that it obtains as a limiting case:

Proposition 2 *Assume that $U : X \to \mathbb{R}_+$ is a CCE representation $(u, \{\varphi_t\}, K)$. Then, for all streams x, the optimal discount factor $D_x(t)$ satisfies*

$$D_x(t) \to \overline{d}_t$$

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

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

4.2 Myopic Model

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

$$D_{u(x_t)}(t) = \left(\frac{\overline{d}_t^m u(x_t)}{mK} \right)^{\frac{1}{m-1}}. \tag{7}$$

As $\overline{d}_t \to 0$, $D_{u(x_t)}(t) \to 0$. Since $D_x(0) = 1$, this limit case corresponds to a completely myopic agent.

For large streams, it is easy to see from Proposition 1 that the representation can be written as

$$U(x) = u(x_0) + \left\{ \sum_{t \geq 1} (\overline{d}_t u(x_t))^{\frac{m}{m-1}} \right\}^{\frac{m-1}{m}}. \tag{8}$$

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

If $K \to 0$ and $\overline{d}_t \to 0$ while holding a_t fixed, Proposition 1 implies that all streams become large streams. From (8), the agent becomes myopic as $\overline{d}_t \to 0$.

Finally, consider the case where $a_t \to \infty$ and $K \to \infty$ with keeping $a_1 \geq K$. Proposition 1 implies that all streams become small streams. From (7), we see that the agent becomes myopic as $a_t = K/\overline{d}_t^m \to \infty$.

4.3 Max-max-type Model

If $K \to 0$ and $a_t \to 0$ while keeping $K \leq a_1$ and holding $\overline{d}_t$ fixed, the CCE model reduces to the case where cost functions vanish to zero and only the capacity constraint remains the same, that is,

$$U(x) = D_x \cdot u(x), \text{ where } D_x = \arg\max D \cdot u(x), \tag{9}$$
$$\text{subject to } \mathcal{D} = \left\{ D \in [0,1]^T \, : \, \sum_{t \geq 1} \left(\frac{D(t)}{\overline{d}_t} \right)^m \leq 1 \right\},$$

or $U(x) = \max_{D \in \mathcal{D}} D \cdot u(x)$. This is also clear from Proposition 1 because all streams become large if $K \to 0$. Proposition 1 also implies that (9) can be written more explicitly as (8).

5 Accommodating Evidence

Dohmen et al [6] show that people with lower cognitive abilities are more impatient. In our model, higher cognitive costs φ or lower cognitive capacity for empathy K correspond to greater impatience. See Section 4.2 for related comparative statics.

The CCE model has several behavioral implications similar to the CE model of NT. For example, both models can accommodate the magnitude effect, preference reversal (the common difference effect), and the concentration bias.

The main difference between the CE and the CCE representations is that the former is time separable, while the latter is time separable only on the set of magnitude-sensitive streams. To illustrate this feature, suppose there are only three time periods and that consumption in the final period 2 is fixed at c_2. Consider a prospect of consuming c_1 in period 1 and let the present equivalent $p(c_1; c_2)$ denote the amount received today that would make him indifferent to it:

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

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

Proposition 3 *If $(0, c_1, c_2)$ is small (resp large), then $p(c_1; c_2)$ is constant (resp. decreasing) in $u(c_2)$.*

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

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

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

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

which is decreasing and convex in b. In particular, p is decreasing in $u(c_2)$. ■

The intuition is simply that if the agent's empathy constraint is binding, then higher values of c_2 causes a reallocation of the limited empathy away from self 1 and towards self 2, making her care less about (more impatient towards) receiving c_1. We are not aware of any direct evidence of such an effect, but we note next that it is consistent with the experimental finding of "preference for spread" documented in Loewenstein and Prelec [18]. In particular, there can exist $u(c) < u(c')$ (such that $(0, c', c')$ is a large stream) and

$$\begin{aligned} (c, 0, 0) &\prec (0, c', 0). \\ (c, 0, c') &\succ (0, c', c'). \end{aligned}$$

Increasing final period consumption from 0 to c' caused the agent to become more impatient, as in the above proposition. The preference pattern suggests a preference for spread of the type discussed in Loewenstein and Prelec [18] since the agent appears to reverse her initial preference for c' tomorrow over c today in order to spread out good consumption.[7]

[7]It does not exactly match Loewenstein and Prelec [18], who ask how subjects would spread opportunities to have dinner at a fancy french restaurant F rather than at home H and finds that a majority exhibit $(F, H, H) \prec (H, F, H)$ and $(F, H, F) \succ (H, F, F)$. Each of these preferences require $D > 1$, which can be interpreted as anticipation. While we can allow this in our model, we do not do so for reasons specified in Section 2.2.

6 General Model

6.1 Functional Form

Say that a tuple $(u, \{\varphi_t\}, K)$ is *basic* if

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

(ii) $\varphi_t : [0, 1] \to \mathbb{R}_+ \cup \{\infty\}$ is an increasing convex function that is

  (a) strictly increasing, strictly convex and differentiable on $\{d : 0 < \varphi_t(d) < \infty\}$, and

  (b) satisfies $\varphi_t(0) = 0$, $\varphi_t'(0) = 0$ and $\varphi_t \leq \varphi_{t+1}$ for all $t < T$,

(iii) $K : X \setminus \Delta_0 \to \mathbb{R}_{++} \cup \{\infty\}$ is either $K_x = \infty$ for all x or a continuous function with the following properties:

  (a) For all x, y and $\lambda > 0$,

$$\frac{u(x_t)}{u(y_t)} = \lambda \text{ for all } t \Longrightarrow K_x = K_y,$$

  (b) $K_{p^t} = \varphi_t(\overline{d}_t)$ for all $p \in \Delta$ and $t \geq 1$, where $\overline{d}_t$ is a supremum of the effective domain $\mathrm{eff}(\varphi_t) := \{d_t \in [0, 1] : \varphi_t(d_t) < \infty\}$,

  (c) if $K_x < \infty$, $K_x \leq K_{xS0}$ for all $S \subset \{1, \cdots, T\}$ with $x_t \succ 0$ for some $t \in S$.

Compared to the regular tuple defined in Section 2.2, the cognitive cost function φ_t here is a more general convex function, and moreover, it is possible that $\varphi_t(d) = 0$ for all d in some interval $[0, \underline{d}_t]$. Intuitively, there can be a base-line degree of selflessness (corresponding to a discount function $\underline{d}_t$) that the agent can access costlessly, that is, $\varphi_t(\underline{d}_t) = 0$ for each t.

In condition (iii) of the regular tuple, the capacity constraint is some constant number $K > 0$. The capacity constraint K as given above is more general in that it can now change with the stream. Property (iii)(a) states that K_x depends only on utility streams $(u(x_t))_{t \geq 1}$ and is homogeneous of degree 0. This can be viewed as saying that K_x depends only on the normalized distribution consumption across time. Property (iii)(b) states that the empathy constraints for dated rewards attain the cost for the maximum level of discount factor at each period. Property (iii)(c) requires that a stream x is associated with weakly less capacity than any of its component rewards.[8]

[8] For example, if $x = (x_0, x_1, x_2)$ and $x_{-2} = (x_0, x_1, 0)$, property (iii)(c) requires that $K_x \leq K_{x_{-2}}$. This is justified with the following story: There exists some background pool of capacity for empathy across future selves. The current self faces with more time trade-offs in x than in x_{-2}. The more difficult solving the trade-offs is, the more the background pool is depreciated.

Property (iii) implies that for all x with $x_t \succ 0$ for some t,

$$K_x \leq K_{x\{t\}0} = K_{p^t} = \varphi_t(\overline{d}_t), \tag{10}$$

which means that the empathy constraint K_x is bounded by the empathy constraints for dated rewards. This condition is regarded as a generalization of property (iii) of the regular tuple. Another implication of (iii) is that if $K_x = \infty$ for some x, $K_{p^t} = \infty$ for all t.

Property (iii)(b) is imposed for uniqueness of the representation. Suppose that for a dated reward p^t, $\varphi_t(\overline{d}_t) < K_{p^t}$. Since $D(t) > \overline{d}_t$ is prohibitively costly, the capacity constraint is never achieved in this case. Hence, K_{p^t} can be reduced up to $\varphi_t(\overline{d}_t)$ without changing behavioral implications. On the other hand, suppose $K_{p^t} < \varphi_t(\overline{d}_t)$. Together with property (iii)(c), for all streams x with $K_x < \infty$, $K_x \leq K_{p^t} < \varphi_t(\overline{d}_t)$. Therefore, φ_t does not have any empirical meaning beyond K_{p^t}. Consequently, we can assume $\varphi_t(\overline{d}_t) = K_{p^t}$.[9]

As in the CCE model, for each x the optimal discount function D_x is chosen subject to two constraints: the *capacity constraint* given by

$$\varphi(D) \leq K_x,$$

where $\varphi(D) = \sum_{t \geq 1} \varphi_t(D(t))$, and the *boundary constraint* given by

$$D(t) \leq \overline{d}_t, \text{ for all } t \geq 1.$$

We define the representation as follows:

Definition 4 (General CCE Representation) *A General Constrained Costly Empathy (General CCE) representation is a basic tuple $(u, \{\varphi_t\}, K)$ such that $\succsim$ is represented by the function $U : X \to \mathbb{R}_+$ defined by*

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

$$\textit{s.t. } D_x = \arg \max_{D \in [0, \overline{d}_t]^T} \{\sum_{t \geq 1} D(t) u(x_t) - \varphi_t(D(t))\} \textit{ subject to } \varphi(D) \leq K_x.$$

For each stream x, an optimal discount function D_x is determined by maximizing the discounted utilities net of aggregated costs for the discount function subject to the capacity and boundary constraints. By condition (10), for all t,

$$\varphi(D) \leq K_x \leq K_{p^t} = \varphi_t(\overline{d}_t).$$

Therefore, if D satisfies the capacity constraint, it also satisfies the boundary constraint, that is, the boundary constraint is redundant. Consequently, an optimal discount function for the General CCE representation is determined by the problem:

$$D_x = \arg \max_{D \in \mathbb{R}^T_+} \{\sum_{t \geq 1} D(t) u(x_t) - \varphi_t(D(t))\} \text{ subject to } \varphi(D) \leq K_x.$$

[9] In other words, the definition of the basic tuple implicitly assumes the maximum costs and minimum capacity constraints for representing preferences. On the other hand, in the regular tuple of the CCE representation, $\{\varphi_t\}$ is assumed to be a CRRA family, and the parameters $((a_t)_{t \geq 1}, m, K)$ are enough to pin down the representation. Hence, we don't have to consider the maximum cost functions there.

6.2 Properties of Optimal Discount Functions

For any stream $\overline{x} \in X \setminus \Delta_0$, consider the ray passing through $\overline{x}$:

$$X_{\overline{x}} = \{x \in X \,|\, x = \alpha \overline{x},\ \exists \alpha > 0\}.$$

By property (iii) of the General CCE representation, K_x is constant on $X_{\overline{x}}$. Denote $K = K_x$ for some (any) $x \in X_{\overline{x}}$.

Proposition 4 *(1) For any $x \in X_{\overline{x}}$, if $\varphi(D_x) < K$, D_x is strictly increasing, and is obtained explicitly by the FOC condition:*

$$D_x(t) = (\varphi'_t)^{-1}(u(x_t)).$$

(2) For any $x, y \in X_{\overline{x}}$, if $\varphi(D_x) = \varphi(D_y) = K$,

$$D_x = D_y.$$

Moreover, $D_x(t)$ depends on the capacity cap K and the whole stream x, and not just the payoff at t.
(3) D_x is weakly increasing on $X_{\overline{x}}$. In particular, D_x is strictly increasing if $\varphi(D_x) < K$, and is constant if $\varphi(D_x) = K$.

From part (1), an optimal discount function $D_x(t)$ depends only on the payoff in period t if the capacity constraint is not binding. By substituting it into the General CCE representation, $U(x)$ is written as

$$U(x) = u(x_0) + \sum_{t \geq 1} (\varphi'_t)^{-1}(u(x_t))u(x_t). \tag{11}$$

Thus, $U(x)$ is additively separable if the capacity constraint is not binding. According to increasing in payoffs, $D_x(t)$ grows unless the capacity constraint is binding. Once the constraint hits, $D_x(t)$ stops growing. Afterwards, on the same ray, $D_x(t)$ is constant but depends on the whole stream x. Consequently, the representation $U(x) = u(x_0) + \sum_{t \geq 1} D_x(t)u(x_t)$ is not additively separable.

6.3 Axioms

The Regularity axiom excluding the Monotonicity condition will be referred to as:

Axiom 8 (Weak Regularity) *$\succsim$ satisfies Order, Continuity, Impatience, C-Monotonicity, and Risk Preference. Moreover, for any steam x, there exists $c \in C$ such that*

$$c \succsim x \succsim 0.$$

The last condition is stronger than the Present Equivalent axiom, and requires that 0 is the worst alternative. If $\succsim$ satisfies Monotonicity, this additional requirement is redundant.

Axiom 9 (Strong X_ℓ-Regularity) *For all $x \in X \setminus \Delta_0$, the following hold.*
(i) if $x \notin X_\ell$ then $\alpha x \in X_\ell$ for some $\alpha \in (0,1]$.
(ii) if $x \in X_\ell$ then $\alpha x \in X_\ell$ for all $\alpha \in (0,1)$.

Consider the ray $\{\alpha x \,|\, \alpha \in (0,1]\}$ that contains all the mixtures that lie between x and 0. By Weak Homotheticity, the agent's impatience must be weakly increasing as we go down this ray from x to 0. Strong X_ℓ-Regularity requires that impatience is in fact strictly increasing as we go down the ray, except possibly for being constant near x. Specifically, Strong X_ℓ-Regularity (i) requires that X_ℓ should always intersect with this ray. That is, there always exists some $\alpha \in (0,1]$ for which αx exhibits ℓ-Magnitude Sensitivity.[10] Strong X_ℓ-Regularity (ii) requires in addition that if x exhibits an ℓ-Magnitude Sensitivity then so must every stream in the ray $\{\alpha x \,|\, \alpha \in (0,1]\}$.

A weaker axiom, called X_ℓ-Regularity, is considered by NT, which requires the same conditions only for dated rewards $p^t \succ 0$ instead of general streams $x \in X \setminus \Delta_0$. We show in Appendix B.1 that together with the other axioms, X_ℓ-Monotonicity implies Strong X_ℓ-Regularity.

Moreover, we impose four axioms for streams in X_ℓ. For notational convenience, for all streams x, $y \in X$ and $S \subset \{0, 1, \cdots, T\}$, let xSy denote the stream that pays according to x at $t \in S$ and according to y otherwise.

Axiom 10 (X_ℓ-Time-Invariance) *For all $p, \widehat{p} \in \Delta$ and t, if $p^t, \widehat{p}^t \in X_\ell$, then*

$$p \succsim \widehat{p} \iff p^t \succsim \widehat{p}^t.$$

Axiom 11 (Time-0 Irrelevance) *For any $x \in X$ and any $p \in \Delta_0$,*

$$x \in X_\ell \implies p\{0\}x \in X_\ell.$$

Axiom 12 (X_ℓ-Dominance) *For any $x \in X$ and any $S \subset \{1, \cdots, T\}$ such that $x_t \succ 0$ for some $t \in S$,*

$$x \in X_\ell \implies xS0 \in X_\ell.$$

Axiom 13 (X_ℓ-Continuity) *X_ℓ is closed in $X \setminus \Delta_0$.*

The first axiom requires that rankings over dated rewards in period t are independent of t for ℓ-magnitude sensitive streams. The second requires that ℓ-magnitude sensitivity of a stream x does not rely on x_0 in any way. The third states that if there is an ℓ-magnitude sensitive stream x paying positive outcomes at some periods within S, then the stream that is identical on S and paying nothing elsewhere is also ℓ-magnitude sensitive. The fourth states that the limit of a sequence of ℓ-magnitude sensitive streams is ℓ-magnitude sensitive if the limit is not an immediate reward.

[10] That is, $c_{\alpha x} \sim \alpha x \implies \beta \circ c_{\alpha x} \succ \beta \alpha x$ for all $\beta \in (0,1)$.

6.4 Representation Results

Theorem 3 *A preference $\succsim$ on X satisfies Weak Regularity, Weak Homotheticity, X_ℓ-Separability, Strong X_ℓ-Regularity, X_ℓ-Time-Invariance, Time-0 Irrelevance, X_ℓ-Dominance, and X_ℓ-Continuity if and only if it admits a General CCE representation.*

Under Weak Regularity, $\succsim$ does not necessarily satisfy Monotonicity, and X_ℓ-Time Invariance is a weaker requirement than Monotonicity. Monotonicity can in fact fail depending on how K_x varies across consumption streams. Nevertheless, there are subdomains where Monotonicity holds. One is the set of magnitude-sensitive streams, X_ℓ. On this subdomain, K_x is irrelevant, and the representation is reduced to an additively separable functional form given by (11). Thus, the representation satisfies Monotonicity. Another subdomain is any ray passing through some consumption stream. As shown in Proposition 4 (3), D_x is weakly increasing on the ray. Thus, the representation satisfies Monotonicity on this subdomain. See Noor and Takeoka [25] for more details.

As in the CE representation, the General CCE representation has strong uniqueness properties.

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

6.5 Proof Outline

A proof sketch of sufficiency of Theorem 3 is as follows. Weak Regularity, X_ℓ-Separability, X_ℓ-Time Invariance, and X_ℓ-Dominance yield an additively separable representation $U(x) = \sum_{t\geq 0} U(x_t)$ on the space of magnitude-sensitive streams X_ℓ. This representation can be rewritten in the obvious way (that is, define $u(p) = U_0(p)$ and $D_x(t) = \frac{U_t(x_t)}{u(x_t)}$) so that it looks like a discounted utility as in the desired representation, with the discount function D_x dependent on the stream. Since u and D_x are given, we can use the first order condition $u(x_t) = \varphi_t'(D_x(t))$ for each t to obtain an additive cost function $\varphi = \sum \varphi_t$ for which D_x is optimal. Here, the cost function φ_t, $t \geq 1$, is increasing and convex (and not necessarily CRRA). This yields a representation close to the desired one on the space of magnitude-sensitive rewards X_ℓ.

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, $\{\alpha x \,|\, \alpha > 0\}$. Weak Homotheticity and Strong X_ℓ-Regularity imply that there exists a unique α_x such that αx is magnitude sensitive if and only if $\alpha \leq \alpha_x$. Thus, $\alpha_x x$ can be regarded as being on the "boundary" of X_ℓ. By this property, for any $x \notin X_\ell \cup \Delta_0$, $\succsim$ satisfies Homotheticity when $\alpha > \alpha_x$, which implies $\alpha_x \circ c_x \sim \alpha_x x$. From this condition, the representation can be extended by $U(x) = u(c_x) = U(\alpha_x x)/\alpha_x$. Moreover, since $\alpha_x x \in X_\ell$, $U(\alpha_x x)$ admits an additively separable representation by the first step. Then, $U(x)$ is more explicitly written as $U(x) = u(x_0) + \sum_{t\geq 1} D_{\alpha_x x}(t) u(x_t)$.

The remaining problem is to infer a general capacity constraint K_x and to show that $D_{\alpha_x x}$ can be regarded as an optimal discount function for x under the constraint. As shown above, along a ray $\{\lambda x \mid \lambda > 0\}$, as λ increases, $D_{\lambda x}$ should first strictly increase and eventually become constant once λx crosses the boundary of X_ℓ. The main step in proving the theorem is to find a cognitive constraint Λ_x for which the optimal D has this property. An arbitrary closed and convex set $\Lambda_x \subset [0,1]^T$ satisfying $\Lambda_x = \Lambda_{\lambda x}$ for all λ does not define a model that is consistent with Weak Homotheticity and Strong X_ℓ-Regularity.[11] Instead, we define $K_x = \varphi(D_{\alpha_x x})$. That is, K_x is the total empathy cost of the optimal discount function at the boundary of X_ℓ. We find that the cognitive 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 verify that K_x satisfies property (iii) of the basic tuple. X_ℓ-Continuity is required for proving continuity of K_x.

7 Concluding Remarks

In the CCE representation, the regular tuple assumes that φ_t takes a CRRA form and K is a constant. On the other hand, in the General CCE representation, the basic tuple assumes that φ_t is increasing and convex, while K is stream-dependent. An intermediate class of the two representations, which has not been axiomatized in the present paper, is that φ_t is increasing and convex, and K is constant. As mentioned in Section 3.5, it is difficult to find an elegant axiom that would imply that K_x is a constant function without homogeneity of φ_t.

The following is one attempt to characterize the constant K model. Say that a capacity constraint $K : X \setminus \Delta_0 \to \mathbb{R}_{++}$ is monotonic if

$$u(x_t) \geq u(y_t), \ \forall t \implies K_x \leq K_y.$$

Then, we can show the next proposition.

[11] Fix x and let $\Lambda_x = \Lambda_{\lambda x} = \Lambda$. Weak Homotheticity and Strong X_ℓ-Regularity require that as stakes are increased, the discount function eventually ceases to change. But without additional structure on Λ, this property may not be obtained. To see this suppose D_x is optimal for x, that is, it satisfies $D_x \cdot u(x) - \varphi(D_x) > D \cdot u(x) - \varphi(D)$ for all $D \in \Lambda$ (the strict inequality comes from the strict convexity of the cost function, which yields a unique maximizer). This can be rewritten as

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

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

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

Consequently, even if D_x is on the boundary of Λ, scaling up rewards may change the agent's discount function in a way inconsistent with Weak Homotheticity and Strong X_ℓ-Regularity.

Proposition 5 *Consider $\succsim$ that admits a General CCE representation. Moreover, assume $\succsim$ satisfies Monotonicity. Then, if K is monotonic, $K_x = K_y$ for any $x, y \in X \setminus \Delta_0$.*

Note that monotonicity of K is stronger than property (iii)(c) of the General CCE representation. Since X_ℓ-Dominance is closely related to property (iii)(c), Strong X_ℓ-Dominance, considered in Appendix B.1, might be a behavioral counterpart of monotonic K. Further investigation is left for future research.

The Impatience axiom does not play a significant role in the construction of our representation from our axioms. We impose it because of its central place in the literature, and because we feel that its violation (such as due to anticipation) is not best understood in terms of empathy. Nevertheless, if we drop Impatience, the boundary constraint would cease to exist since we would be allowing $D(t) > 1$. In this case the joint restriction between a_t and K, that is, $K \leq a_1$, is not needed. Each parameter $((a_t)_{t\geq 1}, m, K)$ of the CCE representation is independent to each other, and $0 < K \leq \infty$ in particular.

Our modelling choices not withstanding, constrained cognitive optimization can gives rise to the Weak Axiom of Revealed Preference. Consider a model where the agent has a pool of resources K, and on facing a menu $\{x, y, z, ...\}$ of streams, she invests her resources in some manner on all the available streams in the menu. This then yields an optimal D with which she evaluates all streams in this menu. This gives rise to a menu-dependent discount function, and would correspond to violations of the Weak Axiom of Revealed Preference. We eschew such a study in this paper. In the CCE model, we assume that there is a cap K in the evaluation of each individual stream. This gives rise to stream-dependent discount functions and preserves the Weak Axiom of Revealed Preference. This affords us tractability for applications and keeps us close to the time preference literature. We leave it to future research to extend and analyze a menu-dependent version of our model.

A Appendix: Proof of Proposition 1

We solve the cognitive optimization problem for each x. Let $\varphi_t(d) = a_t d^m$ on $d \in [0,1]$. As explained in Section 2.2, the boundary constraint $D(t) \leq 1$ is effectively ignored by condition (iii) of regularity. For each x, an optimal discount function $\{D_x(t)\}_{t\geq 1}$ is determined by

$$\max_{D\geq 0} \sum_{t\geq 1} D(t)u(x_t) - \sum_{t\geq 1} \varphi_t(D(t)),$$
$$\text{subject to} \sum_{t\geq 1} \varphi_t(D(t)) \leq K.$$

The FOC of the above maximization problem is obtained as the FOC of the following Lagrangian:

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

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

$$D_x(t) = \left(\frac{u(x_t)}{(1+\lambda) m a_t} \right)^{\frac{1}{m-1}}, \tag{12}$$

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

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

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

and

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

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

Next, suppose x is large. Then, the capacity constraint is binding. By substituting (12) into the capacity constraint,

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

By rearrangement,

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

By substituting it into (12),

$$D_x(t) = \frac{K^{\frac{1}{m}} \left(\frac{u(x_t)}{m a_t} \right)^{\frac{1}{m-1}}}{\left\{ \sum_{t \geq 1} a_t \left(\frac{u(x_t)}{m a_t} \right)^{\frac{m}{m-1}} \right\}^{\frac{1}{m}}} = \frac{(mK)^{\frac{1}{m}} \gamma(t) u(x_t)^{\frac{1}{m-1}}}{\left\{ \sum_{t \geq 1} \gamma(t) u(x_t)^{\frac{m}{m-1}} \right\}^{\frac{1}{m}}}.$$

Therefore,

$$\begin{aligned} U(x) &= u(x_0) + \sum_{t \geq 1} D_x(t) u(x_t) \\ &= u(x_0) + (mK)^{\frac{1}{m}} \left\{ \sum_{t \geq 1} \gamma(t) u(x_t)^{\frac{m}{m-1}} \right\}^{\frac{m-1}{m}}. \end{aligned}$$

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

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

Equivalently,

$$\sum_{t\geq 1}\gamma(t)u(x_t)^{\frac{m}{m-1}} = mK.$$

Therefore, at the boundary,

$$U(x) = u(x_0) + \sum_{t\geq 1}\gamma(t)u(x_t)^{\frac{m}{m-1}} = u(x_0) + mK.$$

B Appendix: Proof of Theorem 1

B.1 Sufficiency

Consider the following axiom:

Axiom 14 (Strong X_ℓ-Dominance) *For any $x \in X$,*

$$x \in X_\ell \text{ and } x_t \succsim y_t,\ \forall t \implies y \in X_\ell.$$

This says that if there is an ℓ-magnitude sensitive stream x then the stream that pays smaller payoffs in every period is also ℓ-magnitude sensitive. If we set $y_t = x_t$ on S and $y_t = 0$ elsewhere, this axiom implies X_ℓ-Dominance.

As an intermediate result, we show the following:

Lemma 1 *Assume $X_\ell \subsetneq X \setminus \Delta_0$. If a preference $\succsim$ on X satisfies Regularity, Weak Homotheticity, X_ℓ-Separability, Strong X_ℓ-Regularity, Time-0 Irrelevance, and Strong X_ℓ-Dominance, then it admits a General CCE representation with $K : X \setminus \Delta_0 \to \mathbb{R}_{++}$.*

Proof. Notice that the set of the axioms in this lemma implies all the axioms of Theorem 3 except for X_ℓ-Continuity. Thus, by all the arguments up to Lemma 18, there exists $K : X \setminus \Delta_0 \to \mathbb{R}_{++} \cup \{\infty\}$ such that $\succsim$ is represented by the General CCE representation. We want to show that K can be taken to be finite-valued. We modify the argument in Lemma 18 as follows.

By assumption, there exists $\bar{x} \notin X_\ell \cup \Delta_0$. By Lemma 14, there exists $\alpha_{\bar{x}} \in (0,1)$ such that $\alpha_{\bar{x}} x \in X_\ell$ and $\alpha x \notin X_\ell$ for all $\alpha > \alpha_{\bar{x}}$. Let $X^{\alpha_{\bar{x}}} = \{y \in X \,|\, y_t \succ \alpha_{\bar{x}} \circ \bar{x}_t,\ \forall t\}$. If $X^{\alpha_{\bar{x}}} \cap X_\ell \neq \emptyset$, there exists $y \in X^{\alpha_{\bar{x}}} \cap X_\ell$. 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, Strong X_ℓ-Dominance implies $\alpha\bar{x} \in X_\ell$ for such α, which is a contradiction. Hence, we have $X^{\alpha_{\bar{x}}} \cap X_\ell = \emptyset$, or $X_\ell \subset X \setminus X^{\alpha_{\bar{x}}}$.

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, $\lambda y \in X^{\alpha_{\bar{x}}}$. Together with the above observation, $\lambda y \notin X_\ell$. Let x denote such λy. That is, we find $x \notin X_\ell$ on the same ray of y. For such x, define

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

Extend to X_ℓ by requiring $K_x = K_{\lambda x}$ for any $\lambda > 0$. The rest of the proof of Lemma 18 is the same as before. ■

We next verify that the set of axioms for the CCE representation implies that of Lemma 1.

Lemma 2 *$\succsim$ satisfies Time-0 Irrelevance, Strong X_ℓ-Dominance, and Strong X_ℓ-Regularity.*

Proof. Time-0 Irrelevance is directly implied from X_ℓ-Monotonicity (ii).

By Time-0 Irrelevance, $x \in X_\ell$ if and only if $p\{0\}x \in X_\ell$ for all $p \in \Delta$. X_ℓ-Monotonicity (ii) and Monotonicity immediately imply Strong X_ℓ-Dominance.

Take any $x \in X \setminus \Delta_0$. By Monotonicity, $0\{0\}x \succ 0$. Suppose $x \notin X_\ell$. By X_ℓ-Monotonicity (i), there exists $y \in X_\ell$ such that $0\{0\}x \succsim 0\{0\}y$. By Monotonicity, $0\{0\}y \succ 0$. By Continuity, $0\{0\}y \succsim 0\{0\}\alpha x$ for some sufficiently small $\alpha > 0$, and so by X_ℓ-Monotonicity (ii), $\alpha x \in X_\ell$. On the other hand, if $x \in X_\ell$, by Monotonicity, $0\{0\}x \succsim 0\{0\}\alpha x$, and hence, by X_ℓ-Monotonicity (ii), $\alpha x \in X_\ell$. ■

Therefore, there exists a General CCE representation for $\succsim$. We first show that the cost function on its effective domain takes the power form for some constants $m > 1$ and $a_t > 0$,

$$\varphi_t(d) = a_t d^m. \tag{13}$$

Moreover, $a_t \leq a_{t+1}$ for all $t \geq 1$. We already know that $\varphi_t : [0, 1] \to \mathbb{R}_+ \cup \{\infty\}$ is an increasing convex function that is strictly increasing, strictly convex, and differentiable on $\{d \,|\, 0 < \varphi_t(d) < \infty\}$. Moreover, $D_r(t)$ is strictly increasing in r on

$$R_\ell(t) = \{r \,|\, r = u(p) \text{ for some } p^t \in X_\ell\}, \tag{14}$$

and is constant otherwise. Since $\succsim$ satisfies X_ℓ-Homogeneity, by the same proof of Theorem 7 (Appendix F) in NT, we can show that $D_r(t)$ on $R_\ell(t)$ is written as a power form, that is, $D_r(t) = \gamma_t r^\theta$ for some $\gamma_t > 0$ and $\theta > 0$. Then, $\varphi_t : [0, \overline{d}_t] \to \mathbb{R}_+$ is rewritten as in (13).

Lemma 3 *For all $x \in X_\ell$,*

$$\varphi(D_x) = \frac{1}{m}[U(x) - u(x_0)].$$

Proof. As shown in (13), φ_t admits a power form, $\varphi_t(d) = a_t d^m$. For all $x \in X_\ell$, the FOC implies $m a_t (D_{x_t}(t))^{m-1} = u(x_t)$. Thus,

$$\begin{aligned}
\varphi(D_x) &= \sum_{t\geq 1} \varphi_t(D_{x_t}(t)) = \sum_{t\geq 1} a_t \left(\frac{u(x_t)}{m a_t}\right)^{\frac{m}{m-1}} = \frac{1}{m} \sum_{t\geq 1} \left(\frac{u(x_t)}{m a_t}\right)^{\frac{1}{m-1}} u(x_t) \\
&= \frac{1}{m} \sum_{t\geq 1} D_{x_t}(t) u(x_t) = \frac{1}{m}[U(x) - u(x_0)],
\end{aligned}$$

as desired. ■

We exclude the possibility of $K_x = \infty$ in the CCE representation.

Lemma 4 $X_\ell \subsetneq X \setminus \Delta_0$.

Proof. By seeking a contradiction, suppose $X_\ell = X \setminus \Delta_0$. From (14), $D_r(t) = \gamma_t r^\theta$ for all $r \in R_\ell(t) = \mathbb{R}_+$. But, then, $D_r(t) > 1$ for all $r > \gamma_t^{-\frac{1}{\theta}}$, which contradicts to the Impatience axiom. Hence, $X_\ell \neq X \setminus \Delta_0$, as desired. ■

By Lemma 1, $\succsim$ admits a General CCE representation with a bounded capacity constraint $K : X \setminus \Delta_0 \to \mathbb{R}_{++}$. From now on, we show that K is constant.

Define $V_\ell := \{U(0\{0\}x) \in \mathbb{R}_{++} \,|\, x \in X_\ell\}$. By Lemma 4, V_ℓ is bounded from above. Indeed, take some $y \in (X \setminus \Delta_0) \setminus X_\ell$. If there exists $x \in X_\ell$ with $U(0\{0\}x) > U(0\{0\}y)$, then by X_ℓ-Monotonicity (ii), we must have $y \in X_\ell$, which is a contradiction. Hence, for all $x \in X_\ell$, $U(0\{0\}x) \leq U(0\{0\}y)$. That is, V_ℓ is bounded from above. Hence, there exists $\overline{v} := \sup V_\ell > 0$. Let

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

The following lemma states that X_ℓ is characterized as the lower contour set of some indifference curve.

Lemma 5 $X_\ell = X_{\overline{v}}$.

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

$X_{\overline{v}} \subset X_\ell$: Take any $x \in X_{\overline{v}}$ with $U(0\{0\}x) < \overline{v}$. By definition of $\overline{v}$, there exists $y \in X_\ell$ with $U(0\{0\}x) \leq U(0\{0\}y)$. By part (ii) of X_ℓ-Monotonicity, $x \in X_\ell$. Next, take $x \in X_{\overline{v}}$ with $U(0\{0\}x) = \overline{v}$. For any $\alpha \in (0,1)$, by Risk Preference and Monotonicity, $\alpha x \prec x$, and hence, $U(0\{0\}\alpha x) < U(0\{0\}x) = \overline{v}$, which implies $\alpha x \in X_{\overline{v}}$. By the above argument, $\alpha x \in X_\ell$. By Lemma 14, appeared in the proof of Theorem 3, $x \in X_\ell$ as $\alpha \to 1$. ■

Next, define

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

Since $K_x = K_{\lambda x}$ for all $\lambda > 0$, it suffices to show that $K_x = K_y$ for all $x, y \in bd(X_\ell)$. This claim follows from the two lemmas below.

Lemma 6 $bd(X_\ell) = \{x \in X_{\overline{v}} \,|\, U(0\{0\}x) = \overline{v}\}$.

Proof. Take any $x \in bd(X_\ell)$. Since $x \in X_\ell = X_{\overline{v}}$ by Lemma 5, $U(0\{0\}x) \leq \overline{v}$. Seeking a contradiction, suppose $U(0\{0\}x) < \overline{v}$. By Continuity, there exists some $\lambda > 1$ such that $U(0\{0\}\lambda x) < \overline{v}$, which implies $\lambda x \in X_\ell$ by Lemma 5. However, by definition of $bd(X_\ell)$, $\lambda x \notin X_\ell$. This is a contradiction.

Conversely, take any $x \in X_{\overline{v}}$ satisfying $U(0\{0\}x) = \overline{v}$. By Lemma 5, we know $x \in X_\ell$. By seeking a contradiction, suppose $x \notin bd(X_\ell)$. Then, there exists some $\lambda > 1$ with $\lambda x \in X_\ell$. By Monotonicity, $U(0\{0\}\lambda x) > U(0\{0\}x) = \overline{v}$. Lemma 5 implies $\lambda x \notin X_\ell$, a contradiction. ■

Lemma 7 *For all* $x, y \in bd(X_\ell)$, $K_x = K_y$.

Proof. By Lemmas 3 and 6,

$$\begin{aligned}
& x, y \in bd(X_\ell) \\
\Longrightarrow\ & U(0\{0\}x) = U(0\{0\}y) \\
\Longrightarrow\ & U(x) - u(x_0) = U(y) - u(y_0) \\
\Longrightarrow\ & \varphi(D_x) = \varphi(D_y) \\
\Longrightarrow\ & K_x = K_y.
\end{aligned}$$

The last implication comes from the definition of K_x, which says that $K_x = \varphi(D_x)$ for all $x \in bd(X_\ell)$.[12] ■

From now on, let $K > 0$ be the constant number implied by Lemma 7.

Lemma 8 *For all $t \geq 1$, $a_t \overline{d}_t^m = K$.*

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

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

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

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

as desired. ■

Together with $\overline{d}_1 \leq 1$, this lemma implies $K \leq a_1$. Since the shape of the cost function beyond the capacity constraint K does not have any behavioral implications, we can assume that $\varphi_t(d) = a_t d^m$ on the whole unit interval $[0, 1]$. Therefore, $(u, \{\varphi_t\}_{t\geq 1}, K)$ is a regular tuple, as desired.

B.2 Necessity

For each components $(u, \{\varphi_t\}, K)$, a CCE representation is given as in Definition 2. As shown in Appendix A, its reduced form is obtained as

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

where

$$D_x(t) = \gamma(t) u(x_t)^{\frac{1}{m-1}} \tag{15}$$

[12] See the proof of Lemma 18.

if $U(x) - u(x_0) \leq mK$, and

$$D_x(t) = \frac{(mK)^{\frac{1}{m}} \gamma(t) u(x_t)^{\frac{1}{m-1}}}{\left\{\sum_{\tau \geq 1} \gamma(\tau) u(x_\tau)^{\frac{m}{m-1}}\right\}^{\frac{1}{m}}} \tag{16}$$

if $U(x) - u(x_0) > mK$.

If $U(x) - u(x_0) > mK$, $D_{\alpha x}(t)$ is constant for all α sufficiently close to one. If $U(x) - u(x_0) \leq mK$, $D_x(t) = D_{u(x_t)}(t)$ is strictly increasing in $u(x_t)$. Therefore,

$$X_\ell = \{x \in X \mid \sum_{t \geq 1} \gamma(t) u(x_t)^{\frac{m}{m-1}} \leq mK\}. \tag{17}$$

Note that $D_x(t)$ is continuous in x. By (15) and (16), $D_x(t)$ is strictly increasing in $u(x_t)$ in X_ℓ, and it is constant on a ray in $X \setminus X_\ell$. It is obvious to see that $\succsim$ that U represents satisfies Order, Continuity, C-Monotonicity, Impatience, Present Equivalents, and Risk Preference. Weak Homotheticity requires $\alpha U(x) \geq U(\alpha x)$, which follows from (15) and (16). Since $D_x(t)$ depends only on $u(x_t)$ on X_ℓ, U is additively separable on this subdomain, which implies X_ℓ-Separability. X_ℓ-Homogeneity is implied by the same argument as in Theorem 7 of NT (The proof is found in the supplementary appendix (Noor and Takeoka [24, Section 4])).

To show X_ℓ-Monotonicity, notice that the CCE representation is additively separable between period 0 and period 1 onward. Thus, for all x, $U(0\{0\}x) = U(x) - u(x_0)$. Take any $x \in X_\ell$ and y with $U(0\{0\}y) \leq U(0\{0\}x)$. Since $U(y) - u(y_0) = U(0\{0\}y) \leq U(0\{0\}x) = U(x) - u(x_0) \leq mK$, $U(y) - u(y_0) \leq mK$, that is, $y \in X_\ell$. If $x \notin X_\ell$, $U(0\{0\}x) = U(x) - u(x_0) > mK$. Since $U(0\{0\}\alpha x) \to 0$ as $\alpha \to 0$, there exists some $\alpha \in (0, 1)$ such that $U(0\{0\}\alpha x) \leq U(0\{0\}x)$ and $U(0\{0\}\alpha x) = U(\alpha x) - u(\alpha \circ x_0) \leq mK$, that is, $\alpha x \in X_\ell$.

Finally, we show the following property:

Lemma 9 $\succsim$ *satisfies Monotonicity.*

Proof. Take any x, y such that $u(x_t) \geq u(y_t)$ for all $t \geq 0$. Since $U(x)$ is additively separable between x_0 and everything else, it is enough to show Monotonicity for streams x, y with $u(x_0) = u(y_0) = 0$. From now on, we consider such streams only.

Since $u(x_t)^{\frac{m}{m-1}} \geq u(y_t)^{\frac{m}{m-1}}$ for all t, we have

$$\sum \gamma(t) u(x_t)^{\frac{m}{m-1}} \geq \sum \gamma(t) u(y_t)^{\frac{m}{m-1}}.$$

Thus, if x and y are ℓ-magnitude sensitive, we have the desired result. From now on, suppose that either x or y is not ℓ-magnitude sensitive. Moreover, since $\sum \gamma(t) u(x_t)^{\frac{m}{m-1}} \geq \sum \gamma(t) u(y_t)^{\frac{m}{m-1}}$, we have either (a) neither x nor y is ℓ-magnitude sensitive, or (b) y is

ℓ-magnitude sensitive but x is not. If case (a) holds, the representation implies that

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

as desired.

Suppose that case (b) holds. Define $x(\alpha) = \alpha x + (1-\alpha)y$ for all $\alpha \in (0,1)$. Note that $u(x_t) \geq u(x_t(\alpha)) \geq u(y_t) \geq 0$. By continuity of the representation, there exists $\alpha^* \in (0,1)$ such that $x(\alpha)$ is not ℓ-magnitude sensitive if $\alpha > \alpha^*$ and $x(\alpha)$ is ℓ-magnitude sensitive if $\alpha \leq \alpha^*$. Since $\sum_{t\geq 1} \gamma(t)u(x_t(\alpha^*))^{\frac{m}{m-1}} = mK$, the representation implies

$$\sum_{t\geq 1} D_x(t)u(x_t) = (mK)^{\frac{1}{m}} \left\{\sum_{t\geq 1} \gamma(t)u(x_t)^{\frac{m}{m-1}}\right\}^{\frac{m-1}{m}}$$
$$\geq (mK)^{\frac{1}{m}} \left\{\sum_{t\geq 1} \gamma(t)u(x_t(\alpha^*))^{\frac{m}{m-1}}\right\}^{\frac{m-1}{m}} = \sum_{t\geq 1} \gamma(t)u(x_t(\alpha^*))^{\frac{m}{m-1}}$$
$$\geq \sum_{t\geq 1} \gamma(t)u(y_t)^{\frac{m}{m-1}} = \sum_{t\geq 1} D_y(t)u(y_t).$$

■

C Appendix: 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 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 $x = p^t$. By the above observation, $D_x(t)$ 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.

From (17) in Appendix B.2, for all $i = 1, 2$,

$$X_\ell = \{x \in X \mid \sum_{t \geq 1} \gamma^i(t) u^i(x_t)^{\frac{m}{m-1}} \leq mK^i\}.$$

Since

$$\begin{aligned} \gamma^2(t) u^2(x_t)^{\frac{m}{m-1}} &= (ma_t^2)^{-\frac{1}{m-1}} u^2(x_t)^{\frac{m}{m-1}} = (m\alpha a_t^1)^{-\frac{1}{m-1}} (\alpha u^1(x_t))^{\frac{m}{m-1}} \\ &= \alpha \gamma^1(t) u^1(x_t)^{\frac{m}{m-1}}, \end{aligned}$$

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

D Proof of Proposition 2

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

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

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

Note that from Proposition 1, X_ℓ approaches to X and all streams become ℓ-magnitude sensitive streams as $m \to \infty$. Proposition 1 implies that for all streams x,

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

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

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

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

E Appendix: Proof of Proposition 4

By property (iii), $K_{\overline{x}} = K_x$ if $\overline{x}, x$ belong to the same ray. Thus, denote $K = K_x$ for some (any) x on the ray. For a stream x on the ray, its optimal discount factor D_x is characterized by the FOC of the following Lagrangian:

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

where $\lambda \geq 0$ is a Lagrange multiplier.

If the capacity constraint is not binding, $\lambda = 0$. By differentiating $\mathcal{L}$ with respect to $D(t)$, an optimal D_x satisfies

$$u(x_t) = \varphi_t'(D_x(t)), \ \forall t \geq 1 \text{ with } u(x_t) > 0.$$

We have a closed-form solution such as

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

In particular, $D_x(t)$ depends only on $u(x_t)$.

If the capacity constraint is binding, by differentiating $\mathcal{L}$ with respect to $D(t)$, an optimal D_x satisfies

$$u(x_t) = (1+\lambda)\varphi_t'(D_x(t)), \ \forall t \geq 1 \text{ with } u(x_t) > 0, \tag{18}$$

$$\sum_{t \geq 1} \varphi_t(D_x(t)) = K \tag{19}$$

for some $\lambda \geq 0$. From (18), $D_x(t) = (\varphi_t')^{-1}(u(x_t)/(1+\lambda))$. By substituting it into (19),

$$\sum_{t \geq 1} \varphi_t\left((\varphi_t')^{-1}\left(\frac{u(x_t)}{1+\lambda}\right)\right) = K.$$

This equation is solved for λ, which is denoted by $\lambda(x, K)$. Then, we obtain

$$D_x(t) = (\varphi_t')^{-1}\left(\frac{u(x_t)}{1+\lambda(x, K)}\right).$$

Note that $D_x(t)$ depends on the whole stream x, not only on the payoff at t, as well as the capacity cap K.

Take any other y on the ray such that the capacity constraint is binding at its optimal D_y. From the FOC,

$$u(y_t) = (1+\lambda')\varphi_t'(D_y(t)), \ \forall t \geq 1 \text{ with } u(x_t) > 0,$$

$$\sum_{t \geq 1} \varphi_t(D_y(t)) = K$$

for some positive λ'. Since x and y belong to the same ray, $y = \alpha x$ for some $\alpha > 0$. Since u is linear, by setting $1 + \lambda' = \alpha(1+\lambda)$, D_y satisfies the FOC of (18). Thus, $D_y = D_x$.

(3) It is implied from (1) and (2).

F Appendix: Proof of Theorem 3

F.1 Additively Separable Utility Representation on X_ℓ

Lemma 10 *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, and homogeneous (that is, $u(\alpha \circ p) = \alpha u(p)$ for all $\alpha \geq 0$.) Moreover, the preference $\succsim$ on X is represented by a continuous utility function $U : X \to \mathbb{R}_+$ such that $U(p) = u(p)$ for all $p \in \Delta_0$.*

Proof. By Weak Regularity, $\succsim|_{\Delta_0}$ satisfies the vNM axioms. There exists a continuous mixture linear function $u : \Delta \to \mathbb{R}_+$ which represents $\succsim|_{\Delta_0}$ and which can be chosen so that $u(0) = 0$.

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

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

If $\alpha > 1$, we identify $\alpha \circ p$ with $p' \in \Delta$ satisfying $p = \frac{1}{\alpha} \circ p' + \frac{\alpha-1}{\alpha} \circ 0$. Then, mixture linearity of u implies that $u(p) = \frac{1}{\alpha}u(p')$, that is, $u(\alpha \circ p) = u(p') = \alpha u(p)$, as desired.

For any $x \in X$, the Present Equivalents axiom ensures that there exists $c_x \in C$ such that $c_x \sim x$. Define $U(x) = u(c_x)$. By construction, U represents $\succsim$. Moreover, for all $p \in \Delta$, $U(p) = u(p)$. In particular, we have $U(0) = u(0) = 0$.

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

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

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

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

■

For each $t \geq 1$, let $\Delta_t = \{p \in \Delta \,|\, p^t \in X_\ell\}$.

Lemma 11 *On the subdomain $X_\ell \cup \Delta_0 \subset X$, U can be written as an additively separable utility form, i.e. $U : X_\ell \cup \Delta_0 \to \mathbb{R}_+$ s.t. for all $x \in X_\ell \cup \Delta_0$,*

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

where u is given as in Lemma 10 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 $x \in X_\ell$, 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 $(x_t, x_s, 0, \cdots, 0)$. By X_ℓ-Separability,

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

Since u is mixture linear,

$$\begin{aligned} & u(c_{(0,x_s,0,\cdots,0)}) + u(c_{(x_t,0,\cdots,0)}) = u(c_{(x_t,x_s,0,\cdots,0)}) + u(0) \\ \iff & U(0, x_s, 0, \cdots, 0) + U(x_t, 0, \cdots, 0) = U(x_t, x_s, 0, \cdots, 0). \end{aligned}$$

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

$$U(x_t, x_s, 0, \cdots, 0) = U_t(x_t) + U_s(x_s). \tag{20}$$

If a stream has three outcomes $x_t, x_s, x_r \succ 0$, denote it by $(x_t, x_s, x_r, 0, \cdots, 0)$. By X_ℓ-Dominance, $(x_t, x_s, 0, \cdots, 0) \in X_\ell^*$. From the above argument, we have (20). By X_ℓ-Separability,

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

Since u is mixture linear,

$$\begin{aligned} & u(c_{(0,0,x_r,0,\cdots,0)}) + u(c_{(x_t,x_s,0,\cdots,0)}) = u(c_{(x_t,x_s,x_r,0,\cdots,0)}) + u(0) \\ \iff & U(0, 0, x_r, 0 \cdots, 0) + U(x_t, x_s, 0, \cdots, 0) = U(x_t, x_s, x_r, 0, \cdots, 0). \end{aligned}$$

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

$$\begin{aligned} U(x_t, x_s, x_r, 0, \cdots, 0) &= U_r(x_r) + U(x_t, x_s, 0, \cdots, 0) \\ &= U_t(x_t) + U_s(x_s) + U_r(x_r). \end{aligned}$$

By repeating the same argument finitely many times, we have

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

where $U_t(x_t)$ is defined as $U_t(x_t) = U(0, \cdots, 0, x_t, 0, \cdots, 0)$. By definition, $U_t(0) = 0$. By X_ℓ-Dominance, for any $x \in X_\ell$, if $x_t \succ 0$, $(x_t)^t \in X_\ell$, 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 $x^n = (0, x_1^n, \cdots, x_T^n) \in X_\ell$, where $x_t^n \to 0$ for all $t \geq 1$. By Time-0 Irrelevance, $p\{0\}x^n =$

$(p, x_1^n, \cdots, x_T^n) \in X_\ell$. Since $p\{0\}x^n \to p \in \Delta_0$, by continuity, $U(p\{0\}x^n) \to u(p)$ and $U(p\{0\}x^n) = U_0(p) + \sum_{t\geq 1} U_t(x_t^n) \to U_0(p)$. Thus, $U_0(p) = u(p)$.

Finally, we show that u must be unbounded from above. First, we show that u is unbounded from 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. Since U_t is non-constant by Time Invariance, there exists some $\tilde{p} \in \Delta$ with $U_t(\tilde{p}) > 0$. 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 Time-0 Irrelevance, $\overline{x} \in X_\ell$. 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 to the Present Equivalents axiom. ■

Lemma 12 *The function $U : X_\ell \cup \Delta_0 \to \mathbb{R}_+$ defined as in Lemma 11 can be written as follows:*

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

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

Proof. Taking the additive representation from Lemma 11, by X_ℓ-Time-Invariance, we have that $U_t(x_t)$ can be written as an increasing transformation of $u(x_t)$. So we can write $U_t(x_t)$ as $U_t(u(x_t))$. Define D_x by $D_{u(x_t)}(t) = \frac{U_t(u(x_t))}{u(x_t)} > 0$ for any $x_t \in \Delta$ with $x_t \succ 0$. Define $\underline{d}_t = \inf \{D_{u(p)}(t) \,|\, 0 \prec p \in \Delta_t\}$. Then

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

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

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

as desired.

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

By Impatience, for all p and $t \geq 1$, $u(p) = U(p^0) \geq U(p^t) = D_{u(p)}(t) u(p)$, which implies $D_{u(p)}(t) \leq 1$. ■

Lemma 13 *The function $U : X_\ell \cup \Delta_0 \to \mathbb{R}_+$ appeared in Lemma 12 can be written as follows:*

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

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

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

Proof. By X_ℓ-Dominance, if $x \in X_\ell$, then $xt0 \in X_\ell$ for $x_t \succ 0$. Thus, φ_t can be derived from the dated rewards at t as follows. Define

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

By Strong X_ℓ-Regularity, if $p^t \in X_\ell$, then $\alpha p^t \in X_\ell$ for all $\alpha \in (0,1)$. Thus, S_t is an interval. Note $\underline{d}_t = \inf S_t$. Denote $\overline{d}_t = \sup S_t$. Define $I_t = S_t \cup \{\overline{d}_t, \underline{d}_t\}$. The cost function φ_t on I_t is implicitly defined by the first order condition

$$u(p) = \varphi_t'(D_{u(p)}(t)), \tag{21}$$

along with the assumption that $\varphi_t(\underline{d}_t) = 0$. Moreover, the continuity of $D_{u(p)}(t)$ wrt $u(p)$ requires that $0 = \varphi_t'(\underline{d}_t)$. The function is by construction once differentiable and has a positive slope. Since $D_{u(p)}(t)$ is strictly increasing in $u(p)$, (21) implies that φ_t' is strictly increasing, and hence, φ_t is strictly convex.

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.

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

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

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

By Impatience, for all positive p and for all $t < T$, $D_{u(p)}(t)u(p) = U(p^t) \geq U(p^{t+1}) = D_{u(p)}(t+1)u(p)$. Thus, $D_{u(p)}(t)$ is weakly decreasing wrt t. This observation implies that the effective domain *eff*(φ_t) of φ_t includes that of φ_{t+1}. For any $d := D_{u(c)}(t+1)$ in the effective domain of φ_{t+1}, it follows from the FOC that

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

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

F.2 Extension to X

Lemma 14 *For any stream $x \in X \setminus \Delta_0$, there exists a unique $\alpha_x \in (0,1]$ such that*

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

Proof. Let $A = \{\alpha \in (0,1] \,|\, \alpha x \in X_\ell\}$. By part (i) of Strong X_ℓ-Regularity, $A \neq \emptyset$. 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 $\alpha x \in X_\ell$, by part (ii) of Strong X_ℓ-Regularity, $\beta x = \frac{\beta}{\alpha}(\alpha x) \in X_\ell$, that is, $\beta \in A$ as desired. Now, by definition of α_x, if $\alpha < \alpha_x$, then $\alpha \in A$, and hence $\alpha x \in X_\ell$. If $\alpha > \alpha_x$, then $\alpha \notin A$, and hence $\alpha x \notin X_\ell$. Uniqueness of α_x is obvious. Moreover, if $x \in X_\ell$, by part (ii) of Strong X_ℓ-Regularity, $A = (0,1)$, and hence, $\alpha_x = 1$. ■

Lemma 15 *For any $x \in X \setminus \Delta_0$, take $\alpha_x \in (0,1]$ which is defined as in Lemma 14. Then,*

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

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

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

as desired.

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

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

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

If $x \in X_\ell$, $\widetilde{\alpha}_x = 1$ satisfies this condition. Thus, assume $x \notin X_\ell$. Let $\widetilde{A} = \{\alpha \in (0,1] \,|\, \alpha \circ c_x \succ \alpha x\}$. By part (i) of Strong X_ℓ-Regularity, $\widetilde{A}$ 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 $\alpha^n \circ c_x \sim \alpha^n x$, by Continuity, $\widetilde{\alpha}_x \circ c_x \sim \widetilde{\alpha}_x x$, as desired.

Step 4: $\widetilde{\alpha}_x \leq \alpha_x$. Seeking a contradiction, suppose $\widetilde{\alpha}_x > \alpha_x$. Lemma 14 implies $\widetilde{\alpha}_x x \notin X_\ell$. By definition, there exists $\beta \in (0,1)$ such that $\beta \circ c_{\widetilde{\alpha}_x x} \sim \beta(\widetilde{\alpha}_x x)$. Since $\widetilde{\alpha}_x \circ c_x \sim \widetilde{\alpha}_x x$, by Step 2, $\widetilde{\alpha}_x\beta \circ c_x \sim \widetilde{\alpha}_x\beta x$. 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, $\alpha \circ c_x \sim \alpha x$. Moreover, for all γ sufficiently close to one, since

$\gamma\alpha \in (\widetilde{\alpha}_x, \alpha_x)$, $\gamma\alpha \circ c_x \sim \gamma\alpha x$. Now, by definition, $c_{\alpha x} \sim \alpha x$, which implies $c_{\alpha x} \sim \alpha \circ c_x$. Since $\alpha x \in X_\ell$ by Lemma 14, for all $\gamma \in (0,1)$, $\gamma \circ c_{\alpha x} \succ \gamma\alpha x$. Thus, we have

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

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

Lemma 16 *For all $x, y \in X \setminus \Delta_0$, take $\alpha_x, \alpha_y \in (0,1]$ which are defined as in Lemma 14. If $x_t \sim y_t$ for all $t \geq 1$, then $\alpha_x = \alpha_y$.*

Proof. By the result in Section F.1, the representation on X_ℓ depends only on utility streams $(u(x_t))_{t=0}^T$. Moreover, by Time-0 Irrelevance, x_0 is independent of whether x is ℓ-magnitude sensitive. Since $u(x_t) = u(y_t)$ for all $t \geq 1$, x is ℓ-magnitude sensitive if and only if so is y. If $x, y \in X_\ell$, $\alpha_x = \alpha_y = 1$. Assume next that $x, y \notin X_\ell$. 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 14, αx is ℓ-magnitude sensitive and αy is not ℓ-magnitude sensitive. Since $u(\alpha x_t) = u(\alpha y_t)$ for all t, this contradicts to the above argument. Thus, $\alpha_x = \alpha_y$, as desired. ■

As shown in Lemma 14, for any $x \in X \setminus \Delta_0$,

$$\alpha_x = \sup\{\alpha \in [0,1] \,|\, \alpha x \in X_\ell\}.$$

Lemma 17 *The function $U : X \to \mathbb{R}_+$ appeared in Section F.1 can be written as*

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

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

Proof. By the result of Section F.1, U has the desired form on $X_\ell \cup \Delta_0$. Consider the case of $x \notin X_\ell \cup \Delta_0$. Since $u(\alpha_x \circ c_x) = U(\alpha_x x)$ by Lemma 15,

$$U(x) = u(c_x) = \frac{1}{\alpha_x} U(\alpha_x x). \tag{22}$$

By the representation on X_ℓ,

$$U(\alpha_x x) = u(\alpha_x \circ x_0) + \sum_{t \geq 1} D_{\alpha_x x}(t) u(\alpha_x \circ x_t). \tag{23}$$

By combining (22) with (23),

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

as desired. ■

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

First, consider the case of $X_\ell = X \setminus \Delta_0$. Since $x \in X_\ell \cup \Delta_0$ for all x, Lemma 17 directly delivers the desired representation by setting $K_x = \infty$ for all $x \in X \setminus \Delta_0$. The CCE representation in this case is additively separable on the whole domain. That is,

Claim 2 *Assume $X_\ell = X \setminus \Delta_0$. If a preference $\succsim$ on X satisfies Weak Regularity, Weak Homotheticity, X_ℓ-Separability, Strong X_ℓ-Regularity, X_ℓ-Time-Invariance, Time-0 Irrelevance, and X_ℓ-Dominance, then it admits a General CCE representation with $K_x = \infty$ for all $x \in X \setminus \Delta_0$.*

From now on, assume $X_\ell \subsetneq X \setminus \Delta_0$. Let

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

Lemma 18 *There is a function $K : X \setminus \Delta_0 \to \mathbb{R}_{++} \cup \{\infty\}$ such that $\succsim$ is represented by*

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

$$s.t. \qquad D_x = \arg\max_{D \in \Lambda_x} \{\sum D(t) u(x_t) - \varphi_t(D(t))\}$$
$$\Lambda_x := \{D \in [0,1]^T \,|\, \varphi(D) \leq K_x\}.$$

Moreover, (1) the function K_x satisfies $K_x = K_{\lambda x}$ 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. Take any $x \in X \setminus \Delta_0$. If $\lambda x \in X_\ell$ for all $\lambda > 0$, define $K_x = K_{\lambda x} = \infty$ for all $\lambda > 0$. Otherwise, we can find another x on the same ray with $x \notin X_\ell$. For such x, define

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

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

For all $x \in X \setminus \Delta_0$, by Lemma 14, there exists $\alpha_x > 0$ such that $\alpha_x x \in X_\ell$. For any $\beta \in (0, \alpha_x)$, since φ is strictly increasing and $D_{u(c)}(t)$ is strictly increasing in $u(c)$, $K_x = \varphi(D_{\alpha_x x}) > \varphi(D_{\beta x}) \geq 0$. 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 17, for any $x \in X_\ell$ we have

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

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

There exists $x' \notin X_\ell \cup \Delta_0$ 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 $\alpha_{x'} x'$, $\varphi(D_x) \leq \varphi(D_{\alpha_{x'} x'}) = K_x$, 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 $x \in X_\ell$.

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

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

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

$$D_{\alpha_x x} \in \arg\max_{D \in \Lambda_x} D \cdot u(x). \tag{24}$$

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

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

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

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

$$\begin{aligned}
& D_{\alpha_x x} \cdot u(\alpha_x x) - \varphi(D_{\alpha_x x}) > D \cdot u(\alpha_x x) - \varphi(D) \\
\Longrightarrow\ & D_{\alpha_x x} \cdot u(\alpha_x x) - D \cdot u(\alpha_x x) > \varphi(D_{\alpha_x x}) - \varphi(D) \\
\Longrightarrow\ & \alpha_x [D_{\alpha_x x} \cdot u(x) - D \cdot u(x)] > \varphi(D_{\alpha_x x}) - \varphi(D) \\
\Longrightarrow\ & D_{\alpha_x x} \cdot u(x) - D \cdot u(x) > \varphi(D_{\alpha_x x}) - \varphi(D) \\
& \text{(since } D_{\alpha_x x} \cdot u(x) \geq D \cdot u(x), \text{ by (24))} \\
\Longrightarrow\ & D_{\alpha_x x} \cdot u(x) - \varphi(D_{\alpha_x x}) > D \cdot u(x) - \varphi(D).
\end{aligned}$$

Thus,

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

as desired.

By Lemma 16, 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 $K_x = K_y$. ■

All that remains to be established is to show properties of K: For all $S \subset \{1, \cdots, T\}$, let

$$\varphi_S(D) := \sum_{t \in S} \varphi_t(D(t)).$$

Lemma 19 *(1) $K : X \setminus \Delta_0 \to \mathbb{R}_{++} \cup \{\infty\}$ is continuous.*
(2) For all p and $t \geq 1$, $K_{p^t} = \varphi_t(\bar{d}_t)$.
(3) If $K_x < \infty$, $K_x \leq K_{xS0}$ for all $S \subset \{1, \cdots, T\}$.

Proof. (1) Take any $x \in X \setminus \Delta_0$. First assume $K_x < \infty$. Thus, there exists some λ with $\lambda x \notin X_\ell$. By X_ℓ-Continuity, any consumption stream y in a small neighborhood of x also satisfies $\lambda y \notin X_\ell$, which implies $K_y < \infty$. In this case, by definition, $K_y = \varphi(D_{\alpha_y y})$ for all such y. Moreover, $D_{\alpha_x x}$ is a unique maximizer of

$$\max\{\sum D(t) u(\alpha_x \circ x_t) - \varphi_t(D(t))\}.$$

If α_x is continuous in x, then $D \cdot u(\alpha_x x)$ is continuous and hence the maximum theorem implies that $D_{\alpha_x x}$ is continuous. Since φ is differentiable (and hence continuous), we have the desired result.

From now on, we will claim that α_x is continuous in x.

Claim 3 *α_x is lower semi-continuous in x, that is, if $x^n \to x$, then*

$$\liminf_n \alpha_{x^n} \geq \alpha_x.$$

Proof. Seeking a contradiction, suppose

$$\alpha_x > \alpha^* := \liminf_n \alpha_{x^n}.$$

Take any $\alpha \in (\alpha^*, \alpha_x)$. There exists a subsequence α_{x^m} converging to α^*. Since $\alpha_{x^m} \to \alpha^*$, $\alpha_{x^m} < \alpha$ for all sufficiently large m. By Lemma 15, $\alpha x^m \sim \alpha \circ c_{x^m}$. By continuity of preference and continuity of present equivalents (Claim 1), $\alpha x \sim \alpha \circ c_x$. On the other hand, since $\alpha < \alpha_x$, Lemma 15 implies $\alpha \circ c_x \succ \alpha x$, which is a contradiction. ■

Claim 4 *α_x is upper semi-continuous in x, that is, if $x^n \to x$, then*

$$\limsup_n \alpha_{x^n} \leq \alpha_x.$$

Proof. Seeking a contradiction, suppose

$$\alpha_x < \alpha^* := \limsup_n \alpha_{x^n}.$$

Take any $\alpha \in (\alpha_x, \alpha^*)$. There exists a subsequence α_{x^m} converging to α^*. Since $\alpha_{x^m} \to \alpha^*$, $\alpha < \alpha_{x^m}$ for all sufficiently large m. By Lemma 14, $\alpha x^m \in X_\ell$. Since X_ℓ is closed in $X \setminus \Delta_0$ by X_ℓ-Continuity, $\alpha x \in X_\ell$. On the other hand, since $\alpha_x < \alpha$, Lemma 14 implies $\alpha x \notin X_\ell$, which is a contradiction. ■

Next consider the case of $K_x = \infty$. We want to show that K_{x^n} diverges to infinity as $x^n \to x$. Without loss of generality, assume $K_{x^n} < \infty$ for all n. Seeking a contradiction, suppose that there exists some subsequence x^m such that $K_{x^m} \leq \overline{K}$ for some $\overline{K} < \infty$. There exists y^m on the boundary of X_ℓ corresponding to each x^m. By definition, $K_{x^m} = \varphi(D_{y^m})$. Since $K_x = \infty$, all y on the same ray passing through x belong to X_ℓ. By X_ℓ-Dominance, $p^t \in X_\ell$ for all p. Thus, each φ_t is unbounded above because $\varphi_t'(D_{u(p)}(t)) = u(p)$ for all $u(p)$. Therefore, together with $\overline{K} \geq K_{x^m} = \varphi(D_{y^m})$, the sequence $\{y^m\}_{m=1}^\infty$ must be bounded. We can find a consumption stream $z^m := \lambda^m y^m \notin X_\ell$ with λ^m sufficiently larger than one. In particular, $\tilde{\alpha} z^m \notin X_\ell$ for some $\tilde{\alpha} \in (0,1)$ sufficiently close to one. Moreover, since x^m and z^m are on the same ray, z^m can be taken to converge to some point $z := \lambda x$.

By Lemma 15, together with the above observations, $\tilde{\alpha} \circ c_{z^m} \sim \tilde{\alpha} z^m$. By continuity of preference and continuity of present equivalents (Claim 1), $\tilde{\alpha} \circ c_z \sim \tilde{\alpha} z$. On the other hand, $K_x = \infty$ implies that $z \in X_\ell$, and hence, $\tilde{\alpha} \circ c_z \succ \tilde{\alpha} z$, which is a contradiction. This completes the proof.

(2) Since φ_t is defined by using $p^t \in X_\ell$, by construction, we have $K_{p^t} = \varphi_t(\overline{d}_t)$.

(3) Step 1: $K_x \leq K_{p^t}$. By part (2), it must be that

$$\{D(t) \in [0,1] \mid \varphi_t(D(t)) \leq K_{p^t}\} = \text{eff}(\varphi_t),$$

and in turn,

$$\{D \in [0,1]^T \mid \varphi_t(D(t)) \leq K_{p^t} \text{ for all } t\} = \text{eff}(\varphi).$$

For any stream x, trivially we must have $\{D \mid \varphi(D) \leq K_x\} \subset \text{eff}(\varphi)$, and so

$$\{D \in [0,1]^T \mid \varphi(D) \leq K_x\} \subset \{D \in [0,1]^T \mid \varphi_t(D(t)) \leq K_{p^t} \text{ for all } t\}.$$

To show that $K_x \leq K_{p^t}$ for all t, take any t and any D in $\{D \mid \varphi(D) \leq K_x\}$ that satisfies $D(t') = 0$ for $t' \neq t$. Then the above condition implies

$$\varphi_t(D(t)) \leq K_x \Longrightarrow \varphi_t(D(t)) \leq K_{p^t}.$$

In particular, if $D(t)$ satisfies $\varphi_t(D(t)) = K_x$,

$$K_x = \varphi_t(D(t)) \leq K_{p^t},$$

as desired.

Step 2: For all $x \in X \setminus \Delta_0$, $d = (d_t)_{t=1}^T$, and $S \subset \{1, \cdots, T\}$,

$$\varphi(d) \le K_x \implies \varphi_S(d) \le K_{xS0}.$$

Take any x and d with $\varphi(d) \le K_x$. By the properties of K shown by Lemma 18, for any $\lambda > 0$, if $y := \lambda x$, then $K_x = K_{\lambda x} = K_y$. By definition of K_x, there exists $\lambda > 0$ such that $y = \lambda x$ belongs to the boundary of X_ℓ and $K_y = \varphi(D_y)$. Since X_ℓ-Dominance implies $yS0 \in X_\ell$ for all $S \subset \{1, \cdots, T\}$ such that $y_t \succ 0$ for some $t \in S$, we have $\varphi_S(D_{yS0}) \le K_{yS0}$. Moreover, the value $D_y(t)$ is also optimal for $yS0$, that is, $D_y(t) = D_{yS0}(t)$ for all $t \in S$. Now, for any d with $\varphi(d) \le K_x$, since $K_x = K_y = \varphi(D_y)$, we have $\varphi_S(d) \le \varphi_S(D_y)$. Therefore,

$$\varphi_S(d) \le \varphi_S(D_y) = \varphi_S(D_{yS0}) \le K_{yS0} = K_{(\lambda x)S0} = K_{\lambda(xS0)} = K_{xS0}.$$

Step 3: The result. Take any $d_S \in [0,1]^T$ such that $d_S(t) \ge 0$ for all $t \in S$ and $d_S(t) = 0$ otherwise. Assume also $\varphi(d_S) \le K_x$. By part (2) and Step 1, $\varphi(d_S) \le K_x \le K_{p^t} = \varphi_t(\bar{d}_t)$. Hence, there exists some d_S^* such that $\varphi(d_S^*) = K_x$. It follows from Step 2 that $K_x = \varphi(d_S^*) = \varphi_S(d_S^*) \le K_{xS0}$, as desired. ■

F.3 Necessity

Given a General CCE representation, define the set of ℓ-magnitude sensitive streams $X_\ell \subset X$ by

$$X_\ell = \{x \in X \mid \alpha U(x) > U(\alpha x) \text{ for all } \alpha \in (0,1)\}.$$

First of all, we show that X_ℓ is characterized by the FOC of the unconstrained optimization problem:

$$\max_{D \in \mathbb{R}_+^T} \{\sum_{t \ge 1} D(t)u(x_t) - \varphi_t(D(t))\}.$$

Let D_x^{un} denote an optimal discount function for the unconstrained optimization problem, which is characterized by the FOC, $u(x_t) = \varphi_t'(D_x^{un}(t))$ for all $t \ge 1$ with $u(x_t) > 0$, or equivalently,

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

if $u(x_t) > 0$, and $D_x^{un}(t) = 0$ if $u(x_t) = 0$. Since φ_t' is strictly increasing, $D_x^{un}(t)$ is strictly increasing in $u(x_t)$.

Lemma 20

$$X_\ell = \{x \in X \mid \varphi(D_x^{un}) \le K_x\}.$$

Proof. To show X_ℓ belongs to the right-hand side, take any $x \in X_\ell$. By the representation,

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

where $D_x = \arg\max_{D \in \Lambda_x} \{\sum_{t \geq 1} D(t)u(x_t) - \varphi_t(D(t))\}$. By definition of X_ℓ, for all $\alpha \in (0,1)$, $u(\alpha \circ c_x) > U(\alpha x)$. Together with linearity of u, this implies

$$\sum_{t \geq 1} D_x(t)u(x_t) > \sum_{t \geq 1} D_{\alpha x}(t)u(x_t).$$

Since $u(x_t) \geq 0$ and $D_x \geq D_{\alpha x}$ by Proposition 4 (3), we have $D_x(t) > D_{\alpha x}(t)$ for some t. By definition of D_x, together with properties of the representation,

$$\varphi(D_{\alpha x}) < \varphi(D_x) \leq K_x = K_{\alpha x}.$$

Hence, $D_{\alpha x} = D^{un}_{\alpha x}$. As $\alpha \to 1$, we have $\varphi(D^{un}_x) \leq K_x$, as desired.

Conversely, take any x from the right-hand side. For $\alpha \in (0,1)$, By property (c) of the representation,

$$\varphi(D^{un}_{\alpha x}) < \varphi(D^{un}_x) \leq K_x = K_{\alpha x}.$$

Therefore,

$$D^{un}_{\alpha x} = D_{\alpha x} = \arg\max_{\Lambda_{\alpha x}} \{\sum_{t \geq 1} D(t)u(\alpha \circ x_t) - \varphi_t(D(t))\}.$$

Since $D_x = D^{un}_x > D^{un}_{\alpha x} = D_{\alpha x}$ and u is linear,

$$\begin{aligned} u(\alpha \circ c_x) &= u(\alpha \circ x_0) + \sum_{t \geq 1} D_x(t)u(\alpha \circ x_t) \\ &> u(\alpha \circ x_0) + \sum_{t \geq 1} D_{\alpha x}(t)u(\alpha \circ x_t) = U(\alpha x), \end{aligned}$$

that is, $\alpha \circ c_x \succ \alpha x$. Hence, $x \in X_\ell$. ■

Note that $D_x(t)$ is continuous in x. By Lemma 20, $D_x(t)$ is strictly increasing in $u(x_t)$ on X_ℓ. It is obvious to see that $\succsim$ that U represents satisfies Weak Regularity. Lemma 20 implies Time-0 Irrelevance and X_ℓ-Continuity.

Lemma 21 *$\succsim$ satisfies Weak Homotheticity.*

Proof. Take any stream $x \in X$. By Proposition 4 (3), $D_x(t) \geq D_{\alpha x}(t)$, which implies, with linearity of u, $\alpha U(x) \geq U(\alpha x)$, or $\alpha \circ c_x \succsim \alpha x$, as desired. ■

Lemma 22 *$\succsim$ satisfies Strong X_ℓ-Regularity.*

Proof. Take any $x \notin X \setminus \Delta_0$. Assume $x \notin X_\ell$. Then by Lemma 20, the unconstrained optimal discount function D^{un}_x violates the capacity constraint, that is, $\varphi(D^{un}_x) > K_x$. Since D^{un}_x is strictly increasing in $u(x_t)$, as $\alpha \to 0$, $D^{un}_{\alpha x}(t) \to \underline{d}_t$ of the minimum discount factor. Since $\varphi_t(\underline{d}_t) = 0$, there must exist $\alpha < 1$ for which $\varphi(D^{un}_{\alpha x}) < K_x$. By property (c), $\varphi(D^{un}_{\alpha x}) < K_{\alpha x}$, implying that $\alpha x \in X_\ell$ by Lemma 20.

Next, take any $x \in X_\ell$ and $\alpha \in (0,1)$. By Lemma 20 and property (c) of the representation, $\varphi(D^{un}_{\alpha x}) < \varphi(D^{un}_x) \leq K_x = K_{\alpha x}$. Again by Lemma 20, $\alpha x \in X_\ell$, as desired. ■

Lemma 23 $\succsim$ *satisfies X_ℓ-Separability.*

Proof. From Lemma 20, $D_x = D_x^{un}$ on X_ℓ. Thus, $D_x(t)$ depends only on $u(x_t)$. Therefore, the representation on X_ℓ is additively separable and satisfies X_ℓ-Separability. ■

Lemma 24 $\succsim$ *satisfies X_ℓ-Time-Invariance.*

Proof. Take any outcomes $p, \widehat{p} \in \Delta$. Suppose $p^t, \widehat{p}^t \in X_\ell$. By the representation on X_ℓ, $U(p^t) = D_{u(p)}(t)u(p)$ and $U(\widehat{p}^t) = D_{u(\widehat{p})}(t)u(\widehat{p})$. Since $D_r(t)$ is increasing in r, if $u(p) \geq u(\widehat{p})$, we have

$$U(p^t) = D_{u(p)}(t)u(p) \geq D_{u(\widehat{p})}(t)u(\widehat{p}) = U(\widehat{p}^t).$$

■

Lemma 25 $\succsim$ *satisfies X_ℓ-Dominance.*

Proof. Take any $x \in X_\ell$ and consider an optimal D_x. By Lemma 20, $\varphi(D_x^{un}) \leq K_x$. Take any $S \subset \{1, \cdots, T\}$ with $x_t \succ 0$ for some $t \in S$. Note that $D_x^{un}(t)$ is also optimal for $xS0$, that is, $D_x^{un}(t) = D_{xS0}^{un}(t)$ for all $t \in S$. By property (iii)(c) of the General CCE representation,

$$\varphi_S(D_{xS0}^{un}) = \varphi_S(D_x^{un}) \leq \varphi(D_x^{un}) \leq K_x \leq K_{xS0}.$$

Thus, again by By Lemma 20, $xS0 \in X_\ell$, as desired. ■

G Proof of Theorem 4

Take two General CCE representations for the preference. By the same argument as in Theorem 2, there exists $\alpha > 0$ such that $u^2 = \alpha u^1$ and $\varphi_t^2 = \alpha \varphi_t^1$ on the effective domain by property (iii)(b).

Note that the unconstrained optimal discount function is identical between the two representations. Indeed, from the above observation, $u^2(x_t) = (\varphi_t^2)'(D_x^{un,2}(t))$ if and only if $\alpha u^1(x_t) = \alpha(\varphi_t^1)'(D_x^{un,2}(t))$, which is equivalent to $u^1(x_t) = (\varphi_t^1)'(D_x^{un,2}(t))$. Thus, we have $D_x^{un,1}(t) = D_x^{un,2}$.

By Lemma 20,

$$\{x \in X \,|\, \varphi^1(D_x^{un}) \leq K_x^1\} = X_\ell = \{x \in X \,|\, \varphi^2(D_x^{un}) \leq K_x^2\}.$$

Since $\varphi_t^2 = \alpha \varphi_t^1$,

$$\{x \in X \,|\, \varphi^2(D_x^{un}) \leq K_x^2\} = \{x \in X \,|\, \varphi^1(D_x^{un}) \leq \frac{K_x^2}{\alpha}\}.$$

Therefore, we must have $K_x^2 = \alpha K_x^1$.

H Proof of Proposition 5

First, take any x, y such that $u(x_t) > 0$, $u(y_t) > 0$ for all $t \geq 0$. Consider the ray passing through x, denoted by $X_x = \{\alpha x \,|\, \alpha > 0\}$. By the property of K, $K_z = K_x$ for all $z \in X_x$. By Monotonicity of preference, there exists $\bar{z} \in X_x$ such that $u(y_t) > u(\bar{z}_t)$. Let $X^{\bar{z}} = \{z \in X \,|\, u(\bar{z}_t) \geq u(z_t),\ \forall t\}$. For any sufficiently small $\alpha > 0$, $\alpha y \in X^{\bar{z}}$. Since K is monotonic,

$$K_x = K_{\bar{z}} \leq K_{\alpha y} = K_y.$$

By the symmetric argument, $K_y \leq K_x$. Thus, we have $K_x = K_y$.

Let $\bar{K} = K_x > 0$ for some (any) x such that $u(x_t) > 0$ for all t. Take any $x \in X \setminus \Delta_0$ such that $u(x_t) = 0$ for some t. Take any sequence x^n within the interior which converges to x. Since $K_{x^n} = \bar{K}$ for all n, $K_x = \bar{K}$ by continuity of the function K.

References

[1] Ainslie, G. (1992): *Picoeconomics*, Cambridge University Press.

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

[3] Baucells, M. and F. Heukamp (2012): "Probability and Time Trade-Off", *Management Science* 58(4), pp. 831–842.

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

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

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

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

[8] Epstein, L., and J. Hynes (1983): "The Rate of Time Preference and Dynamic Economic Analysis", *The Journal of Political Economy* 91(4), pp. 611–635.

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

[10] 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.

[11] Fudenberg, D., and D. Levine (2006): 'A Dual Self Model of Impulse Control', *American Economic Review* 96, pp. 1449–1476.

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

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

[14] Gul, F. and W. Pesendorfer (2001): 'Temptation and Self-Control', *Econometrica* 69, pp. 1403–1435.

[15] Hardisty D., K. Appelt and E. Weber (2013): "Good or Bad, We Want it Now: Fixed-cost Present Bias for Gains and Losses Explains Magnitude Asymmetries in Intertemporal Choice," *Journal of Behavioral Decision Making* 26(4), pp. 348–361.

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

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

[18] Loewenstein, G. and D. Prelec (1993): "Preferences for Sequences of Outcomes", *Psychological Review* 100, pp. 91–108.

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

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

[21] Noor, J. (2011): 'Temptation and Revealed Preference', *Econometrica* 79(2), pp. 601–644.

[22] Noor, J. (2011): "Intertemporal Choice and the Magnitude Effect," *Games and Economic Behavior* 72(1), pp. 255–270.

[23] Noor, J. and N. Takeoka (2020): "Optimal Discounting," Working paper.

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

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

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

[27] 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.

[28] Strotz R. (1955): 'Myopia and Inconsistency in Dynamic Utility Maximization', *Review of Economic Studies* XXIII, pp. 165–180.

[29] Sun, C. and J. Potters (2016): "Magnitude Effect in Intertemporal Allocation Tasks," mimeo.

[30] Thaler, R. (1981): 'Some Empirical Evidence on Dynamic Inconsistency', *Economic Letters* 8, pp. 201–207.

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