Optimal Discounting*

Jawwad Noor   Norio Takeoka

April 25, 2021

Abstract

The agent is modelled as a current self that optimally incurs a cognitive cost of empathizing with future selves. The model unifies well-known experimental and empirical findings in intertemporal choice and enriches the multiple selves model with a notion of self-control. The defining feature of the model is magnitude-decreasing impatience: greater patience towards larger rewards. Two behavioral definitions of magnitude-decreasing impatience are provided and the model is characterized under each of them.

1 Introduction

Models of intertemporal choice in economics routinely assume that the utility of an outcome at time t is discounted by a factor $D(t)$ that is independent of the outcome. However, there is substantial evidence that people are more patient when dealing with larger outcomes than smaller ones. The experimental literature on the *magnitude effect* finds this property in discount functions elicited in the lab (Fredrick et al [19], Sun and Potters [49], Hardisty et al [24], Ericson and Noor [18]) and the field (Andersen et al [4]). The property also explains a host of behaviors discussed in economics, including anomalies of the classic Life-cycle Hypothesis (Yaari [52], Benartzi et al [10], Browning and Collado [13], Scholnick [45]) and historical differences in time preference across societies (Galor and Özak [21])

*A version of this paper was previously circulated under the title "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 very grateful to Itzhak Gilboa (the Editor) and three referees, and to Andrew Ellis, Larry Epstein, Shinsuke Ikeda, Bart Lipman, Jay Lu, Samuel McClure, Marco Mariotti, Anna Rubinchik, Bruno Strulovici, Kan Takeuchi and numerous seminar audiences for very useful comments that helped shape this paper. Noor gratefully acknowledges financial support from the NSF (grant SES-1627906) and thanks the Yale economics department for its hospitality where part of this research was completed. Takeoka gratefully acknowledges financial support from JSPS KAKENHI (Grant Numbers JP18K01503 and JP19KK0308). Last but certainly not least, we thank Kevin Cooke and Nick Saponara for expert research assistance. The usual disclaimer applies.

Motivated by such evidence, this paper suggests a theory for why an agent may be more patient when dealing with larger rewards. Viewing the agent as a sequence of distinct selves,[1] we imagine that the current self has a normative desire to connect with the well-being of future selves, whether out of some sense of moral responsibility or a sense of community with them. It achieves this by creating *empathy* for future selves through the cognitively costly act of imagining itself in their shoes.[2] Higher future rewards incentivize the current self to incur the cost of higher empathy, thereby giving rise to higher patience.

Our primitive is the current self's preference $\succsim$ over the set X consisting of finite horizon consumption streams with generic element $x = (x_0, \cdots, x_T)$. A *Costly Empathy (CE)* representation $U : X \to \mathbb{R}$ for $\succsim$ takes the following form:

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

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

The current self evaluates the consumption stream $x = (x_0, \cdots, x_T)$ via the discounted utility formula (1) where u is the instantaneous utility and the discount function D_x depends on the stream. For each $t > 0$, $D_x(t) \in [0, 1]$ is interpreted as the current self's empathy for self t. The discount function is a cognitive choice, arising from the cognitive optimization problem (2) which seeks to maximize the benefit $\sum_{t \geq 1} D(t) u(x_t)$ of being connected with future selves through some discount function D net of the additive cognitive cost $\varphi(D) := \sum_{t \geq 1} \varphi_t(D(t))$ of the discount function. Each φ_t is an increasing convex function of $D(t) \in [0, 1]$. Moreover, φ_t is increasing in t so that empathy costs are increasing with temporal distance.

The functional form admits alternative interpretations. Consider the classic quote by Pigou [43]: "[O]ur telescopic faculty is defective, and we, therefore, see future pleasures, as it were, on a diminished scale". Consider also the notion of salience or focus (see Bordalo et al [12] and Koszegi and Szeidl [29] who respectively model salience and focus in an atemporal context in terms of menu-dependent weights). One can hypothesize that an investment of attention enhances telescopic faculty, salience or focus, and thus increases the weight given to it. We adopt the empathy interpretation given its natural fit in the language of multiple selves. Indeed, discount functions have been interpreted in terms of altruism in previous literature (see for instance Saez-Marti and Weibull [44], Galperti and Strulovici [22]).

[1] The idea that an agent can or should be viewed as a collection of multiple selves has an illustrious history in philosophy, psychology and in economics (see Parfit [42], Strotz [47], Ainslie [2], Laibson [30], O'Donoghue and Rabin [39], Saez-Marti and Weibull [44], Galperti and Strulovici [22]). The model is sometimes interpreted as a metaphor to describe an individual, and sometimes it is taken as a literal description of an individual (Parfit [42]).

[2] Indeed, in the literature, it has been shown that visualizing future selves increases saving rates (Hershfield et al [25, 26]).

Theoretical Results. It is readily determined that if a preference admits a CE representation, then it is also represented by a *Generalized Discounted Utility* (GDU) representation

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

where $D_{u(x_t)}(t)$ is increasing in $u(x_t)$, a property we refer to as *magnitude-decreasing impatience* (MDI). Theorem 1 shows that the converse is also true: any preference that admits a GDU representation with MDI admits a CE representation. Thus, the behavioral content of the CE model is the same as that of the GDU model with MDI.

While the main behavioral content of the GDU model is readily determined to be a Separability property, the heart of the paper lies in exploring the behavioral meaning of MDI. We provide two behavioral definitions. The first builds on an intuition familiar from the standard Discounted Utility model, namely, that D is identified by the marginal rate of intertemporal substitution (MRS), defined relative to streams that are constant ("on the diagonal"). We show that MDI breaks the connection between D and MRS, as it brings u into play in a subtle way. Nevertheless, a notion of MDI can be formulated in terms of MRS. Our second behavioral definition looks at intertemporal trade-offs "off the diagonal", where u and D are necessarily confounded. Assuming a standard homogeneity structure on u, we provide a definition of MDI that is reminiscent of the magnitude effect. We characterize the CE model using each of these behavioral definitions (Theorems 4 and 6).

Contributions. (a) This paper complements the literature on the magnitude effect (Fredrick et al [19], Loewenstein and Prelec [31], Noor [34], Baucells and Heukamp [7]). If an agent finds receiving $\psi(m,t)$ today as good as receiving m at time t, then the magnitude effect is defined by the property that $\frac{\psi(m,t)}{m}$ is increasing in m. This is in fact MDI expressed in discount functions that have been elicited under particular assumptions (namely, that the utility of receiving any m at any t is $D_m(t) \times m$ and that either the agent does not integrate m with background consumption or that background consumption is fixed across time). We present two alternative ways of behaviorally defining MDI, one of which only assumes smoothness of utility and the other that assumes some homogeneity property of preference (either CRRA utility or Expected Utility). Moreover, while the theoretical literature motivated by the magnitude effect often focuses on particular specifications of the GDU model that exhibit MDI, we provide a general model that generates MDI as its central property and has a novel interpretation.

(b) We add to the multiple selves framework in two ways. (i) While the framework has mainly been used to model present bias and dynamic inconsistency (Laibson [30]), we extend the scope of the model to speak to a range of other behaviors: documented anomalies of the DU model (such as magnitude effects and a preference for increasing sequences) and of the classic Life-cycle Hypothesis (such as the annuitization puzzle and the magnitude hypothesis in consumption smoothing). (ii) In economics and in psychology (Strotz [47], Ainslie [2], Laibson [30], O'Donoghue and Rabin [39]), the multiple selves model is understood to capture self-control problems but without the notion of "self-control": each self is sovereign and maximizes its preference, and there is no notion of resisting an urge. The CE model can be viewed as a version of that model where the current self has the urge to be selfish, but exerts self-control to balance this urge with her normative desire to care for her future selves. This optimization of this self-control leads to MDI. We explore the idea of MDI-as-self-control in an application on procrastination.

The remainder of the paper proceeds as follows. We conclude the Introduction with a discussion of related literature. Section 2 defines the CE model formally and presents special cases. Section 3 defines the class of General Discounted Utility (GDU) and shows that the CE class coincides with the subclass of GDU models that exhibit MDI. Sections 4 and 5 respectively provide an "on-diagonal" and "off-diagonal" behavioral definition of MDI and corresponding foundations for the CE model. Section 6 relates the model to empirical findings whereas Section 7 applies the model to procrastination. All proofs are relegated to appendices. Additional results and omitted proofs are provided in a supplementary appendix (Noor and Takeoka [36]).

Related Literature. In a companion paper, Noor and Takeoka [35] extend the CE model to include a capacity constraint on empathy, leading to a constrained version of the cognitive optimization problem. This is reminiscent of the idea that willpower is a limited resource (see Ozdenoren et al [41] for an early formalization of this in the literature). In this *Constrained CE* model, the agent may need to trade-off limited empathy across different selves, rather than optimize empathy for each self separately as in the CE model. This gives rise to a violation of the Separability property satisfied by the CE model. Noor and Takeoka [37] consider a different extension that permits consumption to provide negative utility. The solution to the cognitive optimization problem is sensitive to the sign of utility and that paper studies how to identify the sign of utility from behavior.

Our paper relates to the following literatures.

(i) *Non-Standard Time Preference*: There is a substantial literature that explains behavioral evidence against the standard Discounted Utility model (see for instance Loewenstein and Prelec [31, 32], Noor [34], Baucells and Heukamp [7], Galperti and Strulovici [22], Wakai [51], to name a few). Like other multiple selves models in this literature, our model explains observed violations of Stationarity and a time-invariant dynamic extension of our static model features dynamic inconsistency (Section 6.2). Our model subsumes the prominent models explaining the magnitude effect, such as Benhabib et al [11] who consider a fixed immediate cost of discounting, and Noor [34] and Baucells and Heukamp [7] who write GDU models that exhibit MDI (see Section 3). The model also relates to papers that note the role of cognitive abilities for time preference (Dohmen et al [16]).

Fudenberg and Levine [20] extend the multiple selves model by positing the existence of 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 self-control cost. The model reduces to a representation closely related to Gul and Pesendorfer [23]'s model of temptation. Our model admits a self-control interpretation where, unlike these papers, the self-control cost is incurred at a cognitive level and not paid from consumption utility. This could be because the cognitive process is non-deliberative: the preferences simply appear to the agent in a way that features MDI. Alternatively, it could be that the agent makes a conscious choice to empathize but, at the time of decision, the cost of doing so is sunk and therefore irrelevant for self 0's choice. At the same time, as in the multiple selves model but unlike Galperti and Strulovici [22], self 0 internalizes only the consumption utility of future selves, and not their empathy for farther future selves nor associated empathy costs.

(ii) *Endogenous Time Preference*: The endogenous time preference literature also deviates from the DU model by assuming that the discount function depends on the stream being faced by the agent (see for instance, Epstein and Hynes [17], Becker and Mulligan [9] and Wakai [51]). While the CE model satisfies Separability, models in this literature typically violate it (due to the dependence of the discount factor on current consumption). Another point of comparison is that this literature typically features dynamically consistent preferences, whereas dynamic extensions of our model will generally be dynamically inconsistent. These differences not withstanding, there is an important overlap with Becker and Mulligan [9], who hypothesize that an agent can alter her discount function by physically investing, for instance, in education. Our agent engages instead in cognitive investment. In Becker and Mulligan [9], the physical investment in education draws from the same physical budget constraint that is used for consumption and saving, and thus the optimal discount function is menu-dependent. In contrast, the cognitive investment in the CE model is independent of the agent's physical budget constraint.

2 Costly Empathy Representation

2.1 Primitives

There are $T+1<\infty$ periods, starting with period 0. The consumption space is $C=\mathbb{R}_+$. Let Δ denote the set of simple lotteries over C, with generic elements $p,q,\cdots$. Some of our models will employ lotteries and some will not, and so it is convenient to let Z denote either C or Δ. The space of consumption streams is given by $X=Z^{T+1}$ (endowed with the product topology), with generic stream $x=(x_0,x_1,\cdots,x_T)$. The space X consists of deterministic streams when $Z=C$, and streams of independent lotteries when $Z=\Delta$. Since C can be embedded in Δ, it is meaningful to talk about deterministic consumption $c\in Z$ even when $Z=\Delta$. The primitive of our model is a preference $\succsim$ over X.

Our benchmark is the standard Discounted Utility model. Say that an instantaneous utility $u:Z\to\mathbb{R}_+$ is *regular* if (i) it is continuous and strictly increasing on C and satisfies $u(0)=0$, and (ii) it is also mixture linear when $Z=\Delta$.

Definition 1 (Discounted Utility Representation) *A Discounted Utility (DU) representation for a preference $\succsim$ over X is a tuple (u,D), where $u:C\to\mathbb{R}_+$ is regular and $D:\{1,\cdots,T\}\to[0,1]$ is weakly decreasing in t, such that $\succsim$ is represented by the function*

$U : X \to \mathbb{R}$ *defined by*

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

The DU model includes standard exponential discounting given by $D(t) = \delta^t$ for all t, with $\delta \in [0, 1]$ (Koopmans [28]), hyperbolic discounting given by $D(t) = \frac{1}{1+t}$ for all t, and its variants such as beta-delta discounting (Laibson [30], O'Donoghue and Rabin [39]).

2.2 CE Model

Consider a regular $u : Z \to \mathbb{R}_+$ and for each $t > 0$, a cost function $\varphi_t : [0, 1] \to \mathbb{R}_+ \cup \{\infty\}$. Say that the tuple $(u, \{\varphi_t\}_{t=1}^T)$ is *basic* if for each $t > 0$, the cost function φ_t is represented by the Riemann integral

$$\varphi_t(D(t)) = \int_0^{D(t)} \varphi_t'(\delta)\, \mathrm{d}\delta, \qquad D(t) \in [0, 1],$$

of a marginal cost function $\varphi_t' : [0, 1] \to \mathbb{R}_+ \cup \{\infty\}$ which is (i) left-continuous, (ii) continuous at 0, (iii) weakly increasing in t, and for which (iv) there exist parameters $0 \leq \underline{d}_t \leq \overline{d}_t \leq 1$ such that $\varphi_t'(d)$ is 0 on $[0, \underline{d}_t]$, strictly increasing on $(\underline{d}_t, \overline{d}_t]$, and takes the value ∞ on $(\overline{d}_t, 1]$.

Regularity of u is defined by familiar properties, but while the restriction "$u(c) \geq 0$" is just a normalization in the DU model, in our model it is a substantive restriction:[3] the solution to the cognitive maximization problem in the CE model is sensitive to the sign of $u(c)$. The problem of endogenizing the sign of $u(c)$ is pursued in Noor and Takeoka [37].

The non-negative, extended-real cognitive cost $\varphi_t(D(t))$ of empathy $D(t) \in [0, 1]$ is computed as a Riemann integral of a marginal cost function φ_t'. Since φ_t' takes the value 0 on $[0, \underline{d}_t]$, it must be that $\varphi_t(D(t)) = 0$ for any $D(t) \leq \underline{d}_t$, and so $\underline{d}_t$ is a baseline level of empathy that can be achieved costlessly. Since φ_t' takes the value ∞ on $(\overline{d}_t, 1]$, it must be that $\varphi_t(D(t)) = \infty$ for any $D(t) > \overline{d}_t$, so that $\overline{d}_t$ is an upper bound on empathy. On $(\underline{d}_t, \overline{d}_t]$, φ_t' is strictly increasing (and thus strictly positive). Given these properties of the marginal cost function, its integral – the cost function – is therefore continuous on $[0, \overline{d}_t]$, increasing and weakly convex on $[0, 1]$, but strictly increasing and strictly convex on $[\underline{d}_t, \overline{d}_t]$. Condition (iii) implies $\varphi_t \leq \varphi_{t+1}$ for all $0 < t < T$ (it is more costly to empathize with farther selves) and $\underline{d}_{t+1} \leq \underline{d}_t$ and $\overline{d}_{t+1} \leq \overline{d}_t$ (the base-line and upper bound on empathy decrease with temporal distance).

By condition (i), φ_t' is at best left continuous. Due to possible discontinuities in φ_t', the cost function φ_t can be non-differentiable. The "kinks" in φ_t will play a role in our analysis (see Section 3.2). Condition (ii), which has bite only when $\underline{d}_t = 0$, plays the role of ensuring that there is no consumption stream x for which the optimal discount function in the model is $D_x = 0$.

We use a basic tuple to define the CE model:

[3] We are grateful to a referee for this observation.

Definition 2 (CE Representation) *A Costly Empathy (CE) representation for a preference* $\succsim$ *over* X *is a basic tuple* $(u, \{\varphi_t\})$ *such that* $\succsim$ *is represented by the function* $U : X \to \mathbb{R}$ *defined by*

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

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

This functional form was interpreted in the Introduction. We make some simple observations here. Since the cognitive problem maximizes the sum of distinct $D(t)u(x_t) - \varphi_t(D(t))$ terms, and since the maximization problem is unconstrained, the cognitive problem is equivalent to solving T independent maximization problems (one for each $t = 1, \cdots, T$), whereby for each $t > 0$,

$$D_x(t) = \arg \max_{D(t) \in [0,1]} \{D(t) u(x_t) - \varphi_t(D(t))\}.$$

Intuitively, the agent can optimize $D_x(t)$ for each t separately due to the absence of a limited stock of empathy. Given the continuity of the cost function φ_t on $[0, \overline{d}_t]$, a maximizer exists,[4] yielding the optimal $D_x(t)$. Given condition (iii) in the definition of a basic tuple, $D_x(t)$ must in fact belong to $[\underline{d}_t, \overline{d}_t]$. Since φ_t is strictly convex on $[\underline{d}_t, \overline{d}_t]$, it must be that the objective function is strictly concave on $[\underline{d}_t, \overline{d}_t]$. Thus, $D_x(t)$ is unique for all t. In particular, the cognitive optimization problem has a unique discount function D_x as its solution.

Since the cognitive maximization problem corresponding to period t depends only on consumption x_t in that period, the solution $D_x(t)$ can be written as $D_{u(x_t)}(t)$. In the special case where φ_t is differentiable, this maximization problem has a first order condition given by

$$u(x_t) = \varphi_t'(D_{u(x_t)}(t)), \tag{3}$$

where the marginal benefit $u(x_t)$ of empathy must equal the marginal cost at the solution $D_{u(x_t)}(t)$. Observe that if the highest possible marginal benefit, $\sup_C u(c)$, is strictly lower than the highest possible marginal cost, $\sup_{\delta \in [\underline{d}_t, \overline{d}_t]} \varphi_t'(\delta)$, then the upper bound $\overline{d}_t$ is never obtained as a solution (that is, $D_{u(x_t)}(t) = \overline{d}_t$ never holds) and we could marginally reduce $\overline{d}_t$ and still represent the preference. To avoid this, some of our uniqueness results will consider only representations $(u, \{\varphi_t\})$ that are *maximal* in the sense that

$$\sup_{\delta \in [\underline{d}_t, \overline{d}_t]} \varphi_t'(\delta) \leq \sup_{c \in C} u(c). \tag{4}$$

[4] If $\varphi_t(\overline{d}_t) = \infty$ (that is, φ_t diverges to infinity as $\delta \to \overline{d}_t$) then for each $u(c_t) > 0$, it is possible to truncate $[0, \overline{d}_t)$ effectively to a compact sub-domain. Hence, the maximum exists. See the proof of necessity of Lemma 4 for more details.

2.3 Special Cases

The DU model is the special case where the agent's empathy allocation is fixed at the base-line level across all periods. Specifically, when $\underline{d}_t = \overline{d}_t = d_t$, the marginal cost φ'_t takes value 0 on $[0, d_t]$ and value ∞ on $(d_t, 1]$, and the cost function is given by $\varphi_t(d) = 0$ if $d \leq d_t$ and $\varphi_t(d) = \infty$ otherwise. Hence, the optimal discount function is magnitude-independent as in the DU model:

$$D_x(t) = d_t.$$

Since the CE model permits φ_t to be non-differentiable on $[\underline{d}_t, \overline{d}_t]$, it is not always amenable to standard optimization techniques. A tractable special case of the model can be obtained by setting $\underline{d}_t = 0$ and taking a power form for the family of cost functions. Most of our illustrations and applications will utilize this formulation.

Definition 3 (Homogeneous CE) *A homogeneous CE representation $(u, m, \overline{d}_t, a_t)$ is a CE representation $(u, \{\varphi_t\})$ such that for all t,*

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

where (i) $m > 1$ and (ii) $a_t > 0$ is increasing in t.

Since the cost function in this model is differentiable on $[0, \overline{d}_t]$, the cognitive optimization problem can be solved in the usual way by including the constraint $d \leq \overline{d}_t$.

3 Reduced Form Structure

In order to develop an intuition for its structure, we first establish that the CE model can be nested within a class of representations that maintains the DU model's additive separability across time but permits magnitude-dependent discounting:

Definition 4 (General Discounted Utility Representation) *A General Discounted Utility (GDU) representation for a preference $\succsim$ over X is a tuple (u, D) where $u : C \to \mathbb{R}_+$ is regular and $D_r : \{1, \cdots, T\} \to [0, 1]$ is weakly decreasing in t and continuous for all $r > 0$, such that $\succsim$ is represented by a strictly increasing function $U : X \to \mathbb{R}_+$ defined by*

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

A GDU representation (u, D) is unbounded if $u(C) = \mathbb{R}_+$.

The GDU model permits $D_r(t)$ to increase or decreasing with r, although in the latter case the monotonicity of U requires that $D_r(t)r$ must be strictly increasing. That said, our study will lead us into the following GDU subclass where impatience is decreasing with respect to magnitude of consumption utility:

Definition 5 (Magnitude-Decreasing Impatience) *A GDU representation (u, D) exhibits magnitude-decreasing impatience (MDI) if $D_r(t)$ is weakly increasing in r for all t.*

3.1 An Equivalence Result

The main result in this section is that the CE model is the subclass of GDU models that exhibit MDI:

Theorem 1 *A preference $\succsim$ over X admits a CE representation $(u, \{\varphi_t\})$ if and only if it admits a GDU representation (u, D) that exhibits magnitude-decreasing impatience. Moreover, the discount function D in the GDU representation corresponds to the optimal discount function in the CE representation.*

We provide a proof outline below. Theorem 1 permits us to immediately deduce that the CE model subsumes several GDU style-models in the literature such as Noor [34], Baucells and Heukamp [7], Benhabib et al [11] and the relevant subclass of Chakraborty [14].[5] We also learn from Theorem 1 that the behavioral foundations of the CE model lie in (a) the behavioral foundations for the GDU model, and (b) a behavioral definition for MDI.

The main task in this paper is to behaviorally define MDI (Sections 4 and 5 provide two definitions and corresponding characterizations of the CE model). The behavioral foundations of GDU correspond to standard conditions, and are provided in the supplementary appendix [36]. We note here only that, beyond basic regularity conditions such as order and continuity, the content of the GDU model lies in Separability, the property that consumption at t is evaluated independently of what is consumed in other periods. This yields an additive utility representation $U(x) = u(x_0) + \sum_{t\geq 1} V_t(x_t)$, and the GDU representation is obtained simply by defining $D_{u(x_t)}(t) = \frac{V_t(x_t)}{u(x_t)}$ for any $x_t > 0$. Separability lacks empirical validity – see Loewenstein and Prelec [32] for empirical evidence and Baucells and Zhao [8] for a more recent critique. However, maintaining Separability here allows us to focus in a simple way on how discounting of period t consumption depends on the magnitude of that consumption. In Noor and Takeoka [35], we relax Separability in order to augment the CE model with a limited stock of empathy. That model accommodates Separability violations of the type observed in Loewenstein and Prelec [32].

Given Theorem 1, special cases of the CE model must correspond to special cases of the GDU model. Consider:

[5] Noor [34] and Baucells and Heukamp [7] are written directly as GDU-style models. Less directly, Benhabib et al [11] write a model where there is a fixed cost of delay. In our context, their utility of a stream can be written as

$$u(x_0) + \sum_t [u(x_t) - cI(x_t > 0)],$$

where $c > 0$ and $I(x_t > 0)$ is an indicator function that takes value 1 when $x_t > 0$ and 0 otherwise. Write $D_{x_t}(t) = \frac{u(x_t)-c}{u(x_t)}$ for any $x_t > 0$ to see that the model reduces to a GDU model with $D_{x_t}(t)$ that is increasing in $u(x_t)$. Similarly, the model of Chakraborty [14] given by

$$x_0 + \sum_t V(\min_{v\in V_\delta} v^{-1}(\delta^t v(x_t))),$$

where the agent with ambiguous tastes uses a conservative present equivalent to evaluate future consumption, is a GDU model with discount function $D_{x_t}(t) = \frac{V(\min_{v\in V_\delta} v^{-1}(\delta^t v(x_t)))}{x_t}$.

Theorem 2 *For any preference $\succsim$ over X, and any tuple $(u, m, \overline{d}_t, a_t)$ as in Definition 3, the following statements are equivalent:*

(a) $\succsim$ admits a homogeneous CE representation $(u, m, \overline{d}_t, a_t)$.

(b) $\succsim$ admits a GDU representation (u, D) such that for all t,

$$D_r(t) = \begin{cases} \kappa_t r^{\frac{1}{m-1}} & \text{if } r \leq \overline{r}_t \\ \overline{d}_t & \text{if } r > \overline{r}_t \end{cases},$$

where $\kappa_t = (ma_t)^{-\frac{1}{m-1}}$ and $\overline{r}_t = ma_t\overline{d}_t^{m-1}$.

The result reveals that the homogeneous CE model is equivalent to a GDU model with a particular form of MDI: the discount function is initially strictly increasing and homogeneous up to some threshold level $\overline{r}_t$ of consumption utility, and constant beyond that threshold. Thus, for any stream $x \in X$ that is "large" ($u(x_t) > \overline{r}_t$ for all t) the homogeneous CE model reduces to a DU model:

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

whereas for any stream that is "small" ($u(x_t) \leq \overline{r}_t$ for all t), the model takes a DU form except that the future utility index is a convex transformation of u:

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

3.2 Proof Sketch for Theorem 1

Sufficiency is readily established as follows: As noted in Section 2.2 (in the discussion following Definition 2), the CE representation has the form $U(x) = u(x_0) + \sum_{t \geq 1} D_x(t)u(x_t)$ where, for each $t > 0$, $D_x(t)$ solves its own cognitive optimization problem $D_x(t) = \arg\max_{D(t) \in [0,1]} \{D(t)u(x_t) - \varphi_t(D(t))\}$, and the solution $D_x(t)$ can be written as a function $D_{u(x_t)}(t)$. It is readily determined that $D_{u(x_t)}(t)$ is weakly increasing in $u(x_t)$. Inserting this optimal discount function $t \longmapsto D_{u(x_t)}(t)$ into the utility representation (1) establishes that the CE model is a GDU model that exhibits MDI.[6] Note that the proof does not rely on whether consumption is deterministic or risky, that is, on whether Z is C or Δ.

The converse is less obvious. Take any GDU representation (u, D) that satisfies MDI. It is instructive to first consider the simple case where $D_r(t)$ is *strictly* increasing in r. We show that, in this case, it is possible to construct a CE representation with a *differentiable* cost function φ_t. Indeed, if such a representation exists, there is a first order condition (3) for the cognitive optimization problem:

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

[6] If $\lim_{\delta \searrow \underline{d}_t} \varphi_t'(\delta) > 0$, the optimal discount factor is constant at $D_{u(x_t)}(t) = \underline{d}_t$ for all $u(x_t) \leq \lim_{\delta \searrow \underline{d}_t} \varphi_t'(\delta)$. Thus, in this case, the agent will not exhibit MDI when payoffs are below a threshold. (We thank a referee for suggesting this feature, which was absent in an earlier version of this paper).

When $D_r(t)$ is strictly increasing in r, we can use this equation to construct the cognitive cost function φ_t: define φ_t' using this equation, and use the first fundamental theorem of calculus to obtain φ_t. Such a cost function is differentiable and strictly convex by construction, and it is compatible only with $D_r(t)$ that is strictly increasing in t.

However, MDI only requires $D_r(t)$ to be *weakly* increasing in r. The idea in the proof is to construct a non-differentiable cost function φ_t and use its *kinks* to explain flat sections in the discount functions. Due to the non-differentiability, the first order condition is not well-defined and so we need to find an appropriate way to define the marginal cost φ_t'. In Appendix A, we show how to construct a discontinuous function φ_t' of r.[7] We then define φ_t as the (Riemann) integral of φ_t'. We show that $D_r(t)$ is indeed optimal with respect to φ_t in the sense of solving the cognitive optimization problem.

4 Foundations: On the Diagonal

Following our remarks in Section 3.1, we proceed to present our first behavioral definition of MDI and a corresponding characterization of the CE model. We operate here in a purely deterministic environment ($X = C^{T+1}$) under the assumption that:

Assumption 1 *Preference $\succsim$ over $X = C^{T+1}$ admits a GDU representation (u, D) that is smooth, in that $u(c)$ and $D_r(t)$ are differentiable in c and r respectively.*

Smoothness is a technical requirement for defining the marginal rate of substitution, and is otherwise not of behavioral interest. In what follows, we say that x is *on the diagonal* if it is a constant stream, so that $x_t = c$ for some $c \in C$ and all t. Otherwise we say that x is *off the diagonal.*

4.1 MDI on the Diagonal

Take any deterministic stream $x = (c, c, \cdots, c) \in C^{T+1}$ on the diagonal as the agent's constant background consumption. For any $m, t > 0$ define the *present equivalent* $\psi_x(m, t)$ by the indifference:

$$(\underbrace{c + \psi_x(m,t)}_{\text{period } 0}, c, \cdots, c) \sim (c, \cdots, c, \underbrace{c + m}_{\text{period } t>0}, c, \cdots, c).$$

The ratio $\frac{\psi_x(m,t)}{m}$ defines a rate of substitution between periods 0 and t, given stream x. The agent's *marginal rate of intertemporal substitution (MRS)* is the limit of this ratio as

[7] The construction of φ_t' is natural if D_r increases in r in a way that partitions the space of utility values r into "i-intervals" (on which D is strictly increasing) and "c-intervals" (on which D is constant). However, in general, such a partition may not exist, since there may exist r that are neither in an i-interval nor a c-interval: every neighborhood of r may contain infinitely many i-intervals and c-intervals. The proof of Theorem 1 navigates this possibility.

we take $m \to 0$:

$$MRS_x(t) := \lim_{m \to 0} \frac{\psi_x(m,t)}{m} \geq 0.$$

Given Assumption 1, the MRS is well-defined and corresponds to the slope of the agent's indifference curve at a stream on the diagonal. Since the GDU model is additively separable, $\psi_x(m,t)$ (and therefore $MRS_x(t)$) is in fact *independent* of consumption $x_{t'}$ in other periods, $t' \neq 0, t$.

In the standard DU model, MRS identifies the discount function: the trade-offs directly reveal D since the marginal utility of consumption is the same across all periods for any x on the diagonal, and the use of small changes purges the trade-offs of the effect of the curvature of utility (due to smoothness). Since the discount function is magnitude-independent in the DU model, MRS cannot vary along the diagonal. We use this property to define magnitude-independent impatience, requiring MRS to remain unchanged as we uniformly scale down any stream $x \in X$ on the diagonal by $\alpha \in (0,1)$ to obtain $\alpha x = (\alpha x_1, \ldots.., \alpha x_T) \in X$.

Axiom 1 (Constant MRS) *For any $0 < t \leq T$ and $x \in C^{T+1}$ on the diagonal,*

$$MRS_{\alpha x}(t) = MRS_x(t) \quad \textit{for all } \alpha \in (0,1).$$

If, on the other hand, the agent's impatience decreases as future consumption increases, then we might expect that the MRS would increase along the diagonal:

Axiom 2 (Increasing MRS) *For any $0 < t \leq T$ and $x \in C^{T+1}$ on the diagonal,*

$$MRS_{\alpha x}(t) \leq MRS_x(t) \quad \textit{for all } \alpha \in (0,1).$$

This serves as a behavioral definition of MDI.[8]

4.2 Representation Results

In the context of the smooth GDU model, we first confirm that:[9]

[8] Increasing MRS is closely related to, but distinct from, the "decreasing marginal impatience (DMI)" property discussed in the endogenous time preference literature (for instance, Epstein and Hynes [17]). To clarify the relationship, define the MRS between time t and $t+1$ for a stream $x \in C^{T+1}$ on the diagonal by $MRS_x(t, t+1) := \frac{MRS_x(t+1)}{MRS_x(t)}$. DMI requires that for any $x \in C^{T+1}$ on the diagonal, where $x_t = c$ for all t, it must be that $MRS_x(t, t+1)$ is increasing in c. From the fact that, on the diagonal,

$$MRS_x(t) = MRS_x(0,1) \times \cdots \times MRS_x(t-1, t),$$

it is easy to see that DMI implies Increasing MRS. The converse is not true, since even if $MRS_x(t)$ and $MRS_x(t+1)$ are increasing in c, the ratio $MRS_x(t, t+1) = \frac{MRS_x(t+1)}{MRS_x(t)}$ may not be. The polar opposite case is called "increasing marginal impatience (IMI)". There is a debate about whether impatience to consume should increase or decrease as actual consumption rises (Obstfeld [38]).

[9] Although u^2 and u^1 are generally related by an affine transformation in additively separable representations, our uniqueness result yields the stronger relationship, $u^2 = \lambda u^1$, because of the normalization $u^i(0) = 0$ in GDU representations.

Theorem 3 *Under Assumption 1, $\succsim$ exhibits Constant MRS if and only if $\succsim$ admits a DU representation.*

Furthermore, if there are two DU representations (u^i, D^i), $i = 1, 2$ of the same preference $\succsim$, then $D^1 = D^2$ and there exists $\lambda > 0$ such that $u^2 = \lambda u^1$.

The next result shows that weakening Constant MRS to Increasing MRS characterizes a subclass of CE models. It also establishes that the CE model has strong uniqueness properties. Recall the notion of maximality defined by (4).

Theorem 4 *Under Assumption 1, $\succsim$ exhibits Increasing MRS if and only if $\succsim$ admits a maximal CE representation where the optimal discount function $D_r(t)$ has the property that $D_r(t)r$ is a convex function of r.*

If there are two maximal CE representations $(u^i, \{\varphi_t^i\})$, $i = 1, 2$ of the same preference $\succsim$, then there exists $\lambda > 0$ such that (i) $u^2 = \lambda u^1$, (ii) $\varphi_t^2 = \lambda \varphi_t^1$ for each t.

The sufficiency part of the characterization result tells us that Increasing MRS implies the existence of a CE representation. So, by Theorem 1, MDI is implied. The necessity part reveals, however, that MDI does not always imply Increasing MRS, unless $D_r(t)r$ is a convex function of r. Therefore, Increasing MRS defines a strong version of MDI, and characterizes a corresponding subclass of CE models $(u, \{\varphi_t\})$. We will see shortly (Section 4.3) that this is because, perhaps unexpectedly, MRS fails to identify D when there exists magnitude-dependence. Instead, it reflects the interaction between u and D: even if MDI holds ($\frac{\partial D_r(t)}{\partial r} \geq 0$), MRS may fail to increase along the diagonal if the curvature $\frac{\partial D_r(t)}{\partial r}$ falls at a fast enough rate.

In the special case of the model where φ_t is differentiable, the convexity restriction in Theorem 4 can be stated equivalently in terms of the cost function. By the cognitive first order condition (3), the optimal discount function must satisfy $D_r(t) = (\varphi_t')^{-1}(r)$. Thus the convexity restriction is equivalent to requiring that $(\varphi_t')^{-1}(r)r$ is a convex function of r.

It is instructive to consider whether the Homogeneous CE model – as characterized in Theorem 2 – satisfies Increasing MRS. It is easy to confirm that $D_r(t)r = \kappa_t r^{\frac{1}{m-1}+1}$ is strictly convex on the subdomain $[0, \overline{r}_t]$, and convex $D_r(t)r = \overline{d}_t r$ on the subdomain $(\overline{r}_t, \infty)$. However, there is a kink in $D_r(t)r$ at $\overline{r}_t$, which violates not only convexity but also Assumption 1. In the supplementary appendix [36] we define a smooth version of the Homogeneous model and provide a characterization result.[10]

Finally, we observe that the uniqueness part of Theorem 4 ensures that the curvature or elasticity of φ_t is uniquely identified by preference.

[10] Unlike the Homogeneous CE model, the cost function in the smooth version must approach ∞ from the left in a smooth manner:

$$\varphi_t(d) = \begin{cases} a_t d^m & \text{if } d \in [0, \widetilde{d}_t], \\ -A_t \ln(\overline{d}_t - d) + C_t & \text{if } d \in (\widetilde{d}_t, \overline{d}_t), \\ \infty & \text{if } d \in [\overline{d}_t, 1], \end{cases}$$

where $A_t, C_t, \widetilde{d}_t$ are parameters defined by $m, a_t, \overline{d}_t > 0$.

4.3 Proof Sketch for Theorems 3 and 4

Theorems 3 and 4 are corollaries of Theorem 1 combined with the following observations.

Given Assumption 1, we can derive the following key relationship: for any $t > 0$ and $x \in X$ on the diagonal,

$$MRS_x(t) = \frac{\partial D_r(t)r}{\partial r}|_{r=u(x_t)} = \left(\frac{\partial D_r(t)}{\partial r}|_{r=u(x_t)}\right) u(x_t) + D_{u(x_t)}(t). \tag{5}$$

The first equality states that MRS is the derivative of discounted utility $D_r(t)r$ with respect to r evaluated at consumption x_t. The product rule yields the second equality. In the standard DU model (where $\frac{\partial D_r(t)}{\partial r} = 0$), expression (5) reduces to the well-known fact that for any x on the diagonal:

$$MRS_x(t) = D(t).$$

But, under magnitude-dependence, this equality fails. Expression (5) reveals that MRS confounds D with the slope $\frac{\partial D_r(t)}{\partial r}$ scaled by $u(x_t)$. Indeed, it is possible that MDI holds ($\frac{\partial D_r(t)}{\partial r} \geq 0$) and Increasing MRS fails if $\frac{\partial D_r(t)}{\partial r}|_{r=u(x_t)}$ falls steeply enough with x_t.

Write the derivative of $D_{u(x_t)}u(x_t)$ wrt $u(x_t)$ as $[D_r(t)r]'$ and note that Increasing MRS is equivalent to the statement that

$$[D_r(t)r]'\Big|_{r=u(x_t)} \geq [D_r(t)r]'\Big|_{r=u(y_t)}$$

for all $u(x_t) \geq u(y_t) > 0$. Thus, Increasing MRS implies that the derivative of $D_r(t)r$ is increasing, and therefore, that $D_r(t)r$ is a convex function. The proof of Theorem 4 exploits the convexity of $D_r(t)r$ to show that $D_r(t)$ is increasing. If Constant MRS holds, then $D_r(t)r$ is more specifically an affine function, and so $D_r(t)r = d_t r + A_t$ for some constants d_t and A_t. The proof of Theorem 3 establishes that $A_t = 0$ and so $D_r(t) = d_t$.

It is worth noting that in general,[11]

$$D_r(t) \text{ is increasing in } r \iff D_r(t)r \text{ is a star-shaped function of } r, \tag{6}$$

since for all $\alpha \in [0, 1]$, $D_{\alpha r}(t) \leq D_r(t)$ is equivalent to $D_{\alpha r}(t)\alpha r \leq \alpha D_r(t)r$. Star-shapeness is weaker than convexity, and so MDI does not necessarily imply Increasing MRS.

5 Foundations: Off the Diagonal

Section 4 defines the MDI property in terms of intertemporal trade-offs centered around streams on the diagonal. We now consider defining it in terms of trade-offs involving streams *off* the diagonal. For such streams, the marginal utility of consumption varies across periods and so the trade-offs confound the interaction of u and D. This *necessitates*

[11] A function $f : \mathbb{R}_+ \to \mathbb{R}$ with $f(0) = 0$ is said to be star-shaped if $f(\alpha r) \leq \alpha f(r)$ for all $r \in \mathbb{R}_+$ and $\alpha \in [0, 1]$.

some structure on u in order to extract information about D from the trade-offs. We adopt a standard homogeneity assumption on u, requiring that when c is "scaled down" by a factor of $\alpha \in (0,1]$ then its utility scales down to $f(\alpha)u(c)$, for some increasing function f. Compared to the classic experimental literature (see Fredrick et al [19]) that elicits discount functions by assuming that u is linear for the range of small rewards offered to subjects and that background consumption lies on the diagonal, we relax linearity, consider all possible rewards and permit background consumption to lie off the diagonal. The homogeneity assumption appears in the experimental literature in the form of CRRA u (for instance, Andreoni and Sprenger [6]) and Expected Utility with respect to u (for instance, Andersen et al [3]).

In what follows, the meaning of "scaled down" will depend on whether consumption is presumed to be deterministic or risky, that is, whether the space of consumption Z is C or Δ. If $Z = C$, then consumption is scaled down in the usual sense as in the previous section, and homogeneity of u is in the usual sense that $u(c) = \gamma c^{\theta}$ for some θ. If $Z = \Delta$, then risky consumption is evaluated using Expected Utility, which implies a homogeneity property (specifically, linearity) with respect to "scaling down" probabilities rather than consumption, without requiring any restriction on the curvature of u. Either route has its merits and demerits. The CRRA form is a very specific restriction on u, albeit one that is standard, weaker than the linearity, and keeps us in the deterministic environment. On the other hand, assuming Expected Utility allows us to leverage a benchmark model of choice under risk in order to leave u unrestricted, but the resulting theory of time preference becomes tied to a theory of risk preference.[12]

We allow for both routes simultaneously, and leave it to the reader to adopt their preferred route. We only assume that:

Assumption 2 *The preference $\succsim$ over $X = Z^{T+1}$ admits a GDU representation (u, D) where*

(i) if $Z = C$ then $u(c)$ is homogeneous in c,[13]

(ii) if $Z = \Delta$ then u is unbounded,

Since $C \subset \Delta(C)$ via a suitable embedding, it is meaningful to talk about deterministic consumption in Z even if Z consists of lotteries. Below, we use p to denote any element of Z, whether it is deterministic or not, and reserve $c \in Z$ to denote a deterministic element.

5.1 MDI Off the Diagonal

Use p^t to denote the stream in X that pays $p \in Z$ at time t and 0 in all other periods. Such a stream is called a *dated reward*. We write immediate rewards p^0 simply as p. So p denotes both consumption $p \in Z$ and $(p, 0, \cdots, 0) \in X$, and in particular 0 also denotes the stream $(0, \cdots, 0)$.

[12] We thank a referee and the Editor for emphasizing the importance of the distinction between the two routes.

[13] A CRRA utility index u can be obtained from conditions on preference – see footnote 26.

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

$$c_x \sim x.$$

These present equivalents capture, for any given (generically off-diagonal) stream x, how much current consumption must be given to the agent to compensate her for losing her future consumption.[14] In contrast, the present-equivalent ψ defined in the on-diagonal approach specified, for any given stream on the diagonal, how much extra consumption today is as good as receiving an extra amount in the future. Note that Assumption 2 posits unboundedness of u in (ii), and unboundedness is also implied by (i) (given monotonicity of GDU, it must be that $u(c) = \gamma c^{\theta}$ for $\theta > 0$). The role of unboundedness is just to ensure that every stream has a present equivalent.

We posit the existence of a "scaling operation" that takes each $p \in Z$ and $\alpha \in [0,1]$ into some $\alpha \circ p \in Z$ that is less desirable than p. Specifically, in the deterministic context $Z = C$, the scaling operation is just the scaling of *consumption*, that is, $\alpha \circ c := \alpha c$, as in the on-diagonal approach. In the context of risky consumption $Z = \Delta$, the *probability of nonzero consumption* is scaled down: lottery p is α-mixed with 0 to obtain the lottery $\alpha \circ p := \alpha \circ p + (1-\alpha) \circ 0$. The scaling of a *stream* $x = (x_0, \cdots, x_T) \in X$ is naturally defined by

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

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 her impatience. If her impatience is independent of the stream then, given the homogeneity property, scaling both c_x and x down by α will not change the relative desirability of either. We therefore obtain a behavioral expression of magnitude-independent impatience:

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

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

Now suppose that the agent is more patient towards larger streams. Scaling down x makes the stream less desirable, and a corresponding increase in impatience would cause the stream $\alpha \circ x$ to lose value faster than the immediate reward $\alpha \circ c_x$. This suggests a behavioral definition of MDI:

[14] If "consumption" is interpreted as the "change in consumption relative to a reference consumption level", then zero consumption just stands for the "reference consumption level", and is therefore not literally zero. Alternatively, if consumption is taken to be absolute, then the zero future consumption requirement can be relaxed: Suppose that, for every $t > 0$, the agent's impatience is magnitude-*independent* for all consumption below some threshold $c_t^* > 0$ (which happens when $\underline{d}_t > 0$ and $\lim_{\delta \searrow \underline{d}_t} \varphi_t'(\delta) > 0$). Then, for any $c_t \leq c_t^*$, the optimal discount factor is constant at the base-line, that is, $D_{u(c_t)}(t) = \underline{d}_t > 0$. In this case, present equivalents can be defined with future consumption fixed at $c_t^* > 0$ in each t. This satisfies the key implicit requirement for our behavioral definition of MDI below, namely, that scaling down the present equivalent (along with the future stream $c_1^*, \cdots, c_T^*$) does not change the corresponding optimal discount function. See the supplementary appendix [36] for formal details.

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

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

In a deterministic context, Weak Homotheticity is reminiscent of the magnitude effect (Fredrick et al [19]). Given subjects' background consumption, these experiments elicit, for instance, the present equivalent $\$c_m$ of receiving $\$m$ at time t, and observe how c_m changes as a function of m. The magnitude effect can be described by the requirement that, for $\alpha < 1$,

$$[c_m \text{ at time } 0\ ] \sim [m \text{ at time } t] \Longrightarrow [\alpha c_m \text{ at time } 0] \ \succ [\alpha m \text{ at time } t],$$

where the consumption here is interpreted as the "change in consumption relative to background consumption". For instance, the average choices of the subjects in Thaler [50] were \$15 now$\sim$ \$60 in a year and \$3000 now$\sim$ \$4000 in a year. These preferences imply that scaling the future \$4000 by $\alpha = 0.015$ to the value \$60 is worse than scaling an immediate \$3000 by $\alpha = 0.015$ to the value of \$45. If we interpret consumption in our model as "change in consumption" and restrict attention to dated rewards instead of streams, then the Weak Homotheticity axiom exactly corresponds to the magnitude effect.

In the context of risky consumption, there is mixed evidence on Weak Homotheticity: in Öncüler [40] and Anderson and Stafford [5], subjects become more impatient under risk, and Sun and Li [48] find that subjects exhibit the magnitude effect even when rewards are risky. Keren and Roelofsma [27] find the opposite. The preceding suggests a perspective on this mixed evidence. While we maintain Expected Utility, experimental findings routinely confirm the Allais paradox, a behavioral pattern contradicting Expected Utility and suggestive of an inordinate preference for certainty. MDI works in favor of Weak Homotheticity but the Allais paradox works against it: when c_x is deterministic (as in the experiments), then the Allais paradox implies that it steeply loses value in the second comparison ($\alpha \circ c_x$ vs. $\alpha \circ x$), and this may well lead to the preference $\alpha \circ c_x \precsim \alpha \circ x$.

5.2 Representation Results

As a first step we show that

Theorem 5 *Under Assumption 2, $\succsim$ satisfies Homotheticity if and only if $\succsim$ admits a DU representation. Moreover, DU representations have the same uniqueness property as in Theorem 3.*

Therefore Homotheticity expresses the magnitude-independence of discounting in the DU representation.[15] The magnitude-dependence in the CE model is captured by:

[15] If Z consists of lotteries, it is easily seen that the DU representation is equivalent to GDU augmented with the vNM Independence axiom: for all streams $x, y, z \in X$ and $\alpha \in (0,1)$, $x \succsim y \Longleftrightarrow \alpha \circ x + (1-\alpha) \circ z \succsim \alpha \circ y + (1-\alpha) \circ z$. In particular, Homotheticity is equivalent to vNM Independence under GDU.

Theorem 6 *Under Assumption 2, $\succsim$ satisfies Weak Homotheticity if and only if it admits a CE representation. Moreover, CE representations have the same uniqueness property as in Theorem 4.*

The proof of this result involves the simple verification that Weak Homotheticity yields the star-shapeness property in (6), which we have already seen characterizes MDI. Then, by Theorem 1, there must exist a CE representation for the preference. Relative to the uniqueness result in Theorem 4, we do not need to impose maximality of φ_t, since φ_t is always maximal when u is unbounded (which is presumed in Assumption 2).

It is worth observing that, when consumption is risky, the Expected Utility assumption that defines the CE model could have been formulated in an alternative way:[16] while we assumed that a lottery p received at t would be evaluated as $D_{u(p)}(t)u(p)$, one could alternatively posit that it is evaluated as $E_p[D_{u(c_t)}(t)u(c_t)]$. Such an alternative model changes the behavioral implications of the model. For example, it will be consistent with Homotheticity despite magnitude-dependent D. Moreover, unlike our model, the magnitude-dependence of D will contribute to the agent's risk attitude, since the vNM utility index is effectively $D_{u(c_t)}(t)u(c_t)$. In particular, if $\succsim$ exhibits the Increasing MRS axiom on C^{T+1}, then Theorem 4 yields that $D_r(t)r$ is convex in r, and hence we have the property that $D_{u(c)}(t)u(c)$ is less concave than $u(c)$ with respect to c. Indeed, the agent will be more risk tolerant when choice is delayed, generating a finding in the "risk and time" literature spawned by Keren and Roelofsma [27].

5.3 Special Case: Homogeneous CE

Next, consider the homogeneous CE model. Say that a stream $x \in X$ is *Magnitude Sensitive* if the agent's impatience strictly reduces whenever the stream is made less desirable.

Definition 6 (Magnitude-Sensitivity) *A stream $x \in X$ is Magnitude Sensitive if*

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

The set of all Magnitude Sensitive streams is denoted by $X^* \subset X$.

It is clear that immediate consumption (whether due to monotonicity of u in the deterministic context or the Expected Utility assumption in the risky context) is not Magnitude Sensitive.

The homogeneous CE model is characterized by the structure it places on X^*. Consider:

Axiom 5 (X^*-Regularity) *For any $p \in Z$ and $t > 0$,*
(i) if $p^t \notin X^*$, then $\alpha \circ p^t \in X^*$ for some $\alpha \in (0,1]$, and
(ii) if $p^t \in X^*$, then $\alpha \circ p^t \in X^*$ for all $\alpha \in (0,1)$.

[16] We thank a referee for pointing out the possibility of this alternative formulation.

Consider the ray $\{\alpha \circ p^t \,|\, \alpha \in (0,1]\}$ defined by the mixtures that lie between p^t and 0. By Weak Homotheticity, D weakly increases along this ray. X^*-Regularity requires that it is in fact strictly increasing, possibly becoming constant as we approach p^t. Specifically, X^*-Regularity (i) requires that if p^t is not already in X^*, there exists some $\alpha \in (0,1]$ for which $\alpha \circ p^t$ exhibits Magnitude Sensitivity. X^*-Regularity (ii) requires in addition that if p^t exhibits an Magnitude Sensitivity then so must every dated reward in the ray $\{\alpha \circ p^t \,|\, \alpha \in (0,1]\}$.

Next, consider:

Axiom 6 (X^*-Homogeneity) *For any dated rewards $p^t, q^s \in X^*$, their present equivalents $c_{p^t} \sim p^t$ and $c_{q^s} \sim q^s$, and any $\alpha, \gamma \in (0,1)$,*

$$\gamma \circ c_{p^t} \sim \alpha \circ p^t \implies \gamma \circ c_{q^s} \sim \alpha \circ q^s.$$

X^*-Homogeneity places structure on homotheticity violations, requiring that if scaling down p^t by α is as good as scaling down its present-equivalent c_{p^t} by γ, then γ depends on α but not the stream. It is easy to see that this axiom imposes homotheticity on dated rewards in X^*, since for any $p^t, q^s \in X^*$, it must be that $p^t \sim q^s \Longrightarrow \alpha \circ p^t \sim \alpha \circ q^s$.

We close this section with:

Theorem 7 *Under Assumption 2, $\succsim$ satisfies Weak Homotheticity, X^*-Regularity and X^*-Homogeneity if and only if $\succsim$ admits a homogeneous CE representation.*

Moreover, if there are two homogeneous CE representations $(u^i, m^i, \overline{d}^i_t, a^i_t)$, $i = 1,2$ of the same preference $\succsim$, then there exists $\lambda > 0$ such that (i) $u^2 = \lambda u^1$, (ii) $\overline{d}^2_t = \overline{d}^1_t$, $a^2_t = \lambda a^1_t$, and $m^2 = m^1$ for each t.

6 Accommodating Evidence

6.1 Magnitude Effect

Recall the notation in Section 4.1. The *magnitude effect* is defined by the property that $\frac{\psi_x(m,t)}{m}$ is increasing in m. Loewenstein and Prelec [31] show that this can arise from the curvature of utility for money: in the DU model, the present equivalent $\psi_x(m,t)$ satisfies $u(x_0 + \psi_x(m,t)) - u(x_0) = D(t)[u(x_t + m) - u(x_t)]$), which implies that

$$\frac{\psi_x(m,t)}{m} = \frac{u^{-1}\left(D(t)[u(x_t+m) - u(x_t)] + u(x_0)\right) - x_0}{m}.$$

Noor [34] provides a calibration theorem to show that the curvature of utility is not an adequate explanation for the magnitude effect.[17] The CE model can readily produce the magnitude effect due to the magnitude-dependence of D, even without any curvature in u.

[17] The calibration theorem implies that for an arbitrary discount function, concave utility, and arbitrary background stream of wealth, if the agent exhibits, say, \$15 now $\sim$ \$60 in a year then the following must hold: give \$60 in a year unconditionally to the agent, and she will subsequently never give up \$$x$ today in return for \$$4x$ (=\$$\frac{60}{15}x$) return in a year, for any value of x.

For completeness, we note the mixed evidence for the magnitude effect for losses: Thaler [50], for instance, finds that subjects are less impatient towards larger losses, whereas Hardisty et al [24] find that subjects are more impatient towards larger losses. Although our model does not speak to losses relative to a reference point, it is worth considering whether these patterns can be understood in terms of optimal empathy. The key issue is that, in the cognitive optimization problem, if a stream yields negative discounted utility, then the optimal discount function must be 0, but the evidence requires it to be strictly positive. To accommodate the pattern in Hardisty et al [24], consider the CE model with negative payoffs, but maintain that the utility from all payoffs is positive. Then the model generates higher impatience towards payoffs with lower utility, and in particular, towards larger losses. The pattern in Thaler [50] is accommodated in Noor and Takeoka [37] where the cognitive optimization problem is extended so that the agent considers the absolute value of utility from outcomes.

6.2 Preference Reversals

For any $c \in C$ and any stream x with $x_T = 0$, let cx denote the stream $(c, x_0, \cdots, x_{T-1})$. It is well-known that the behavioral expression of Exponential Discounted Utility (Definition 1) is Stationarity (Koopmans [28]):

Axiom 7 (Stationarity) *For any streams x, y such that $x_T = y_T = 0$ and any c,*

$$x \succsim y \Longleftrightarrow cx \succsim cy.$$

Stationarity is routinely violated in experiments. A notable finding is that of *preference reversals*, also known as the common difference effect, immediacy effect and present bias (Fredrick et al [19]), defined by $c < \widehat{c}$ and $d > 0$ such that

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

suggesting a bias towards the present. Our model explains this as follows:[18] if empathy is sufficiently costly, the current self will be selfish and prefer the immediate option as in the first comparison. If the cost function for empathizing with self t is not much different from that for self $t+d$, the agent empathizes with self t and self $t+d$ to a similar degree, and therefore prefers the higher outcome.

The attention received by preference reversals notwithstanding, there is *substantial* evidence of the reverse (Fredrick et al [19]):

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

[18] For example, assume a homogeneous CE model. When $(\widehat{c})^d$ is small, the two comparisons are $u(c)$ vs $\kappa_d u(\widehat{c})^{\frac{m}{m-1}}$ and $u(c)$ vs $\left(\frac{\kappa_{t+d}}{\kappa_t}\right)^{\frac{m-1}{m}} u(\widehat{c})$. Depending on parameter values, the agent can exhibit preference reversals or their converse.

This behavior is in fact natural from the perspective of our model. Impatience is sufficient to explain the second choice – the current self's empathy for future selves is decaying at some rate. If the period 0 self has sufficiently high empathy for the first d selves then a large reward $(\widehat{c})^d$ may be chosen over a smaller immediate reward c.

While such behavior is frequently observed in experiments, it is all but ignored in theory and applications, presumably because the lens of present bias is dominant in the literature. We note that even the beta-delta model can give rise to such behavior: that model is silent on how long period 0 is, and if it is long enough, then the agent would exhibit "future bias".

Given that the CE model does not place any particular restriction on Stationarity violations, we close by providing a special case of the model that produces preference reversals:

Axiom 8 (Quasi-Stationarity) *For any streams x, y such that $x_0 = y_0 = x_T = y_T = 0$ and any c,*

$$x \succsim y \Longleftrightarrow cx \succsim cy.$$

The condition requires the conclusion of Stationarity to hold only in comparisons where immediate consumption is not relevant. Observe that the condition requires at least 4 periods ($T \geq 3$) for it to meaningfully restrict the agent's discount function.

Proposition 1 *Suppose that $T \geq 3$. A homogeneous CE representation $(u, m, \overline{d}_t, a_t)$ satisfies Quasi-Stationarity iff there exist $c^* > 0$, $0 < \delta \leq 1$, and $0 < \beta \leq 1/(\delta u(c^*)^{\frac{1}{m-1}})$ such that for each t,*

$$a_t = \frac{1}{m\beta^{m-1}(\delta^{m-1})^t}, \text{ and } \overline{d}_t = \beta\delta^t u(c^*)^{\frac{1}{m-1}},$$

and the optimal discount function takes the form:

$$D_c(t) = \begin{cases} \beta\delta^t u(c)^{\frac{1}{m-1}} & \text{if } c \leq c^*, \\ \beta\delta^t u(c^*)^{\frac{1}{m-1}} & \text{if } c > c^*. \end{cases}$$

See the supplementary appendix (Noor and Takeoka [36]) for the proof. According to the proposition, Quasi-Stationarity requires that $D_c(t)$ takes the familiar beta-delta form augmented with a transformation of $u(c)$. This is an increasing transformation for all c below some threshold $c^* > 0$. Beyond the threshold, the beta-delta form is multiplied with a constant $u(c^*)^{\frac{1}{m-1}}$. We can view δ as the long run discount factor and $\beta u(c)^{\frac{1}{m-1}}$ as reflecting the urge for immediate gratification. This urge reduces as c increases and beyond c^*, it is constant, as in the beta-delta model. In the proof of the Proposition, a key step is to establish that c^* cannot in fact depend on t.

We illustrate how the CE model accommodates preference reversals under Quasi-Stationarity by using the optimal discount function in Proposition 1. Consider a typical pattern such as $(c, 0, 0) \succ (0, c + d, 0)$ and $(0, c, 0) \prec (0, 0, c + d)$. If c is sufficiently large, the optimal discount function is magnitude-independent. Then, the preference reversal choice pattern is equivalent to assuming $\beta u(c^*)^{\frac{1}{m-1}} < 1$ as in the beta-delta model. On the other hand, if c is sufficiently small, a preference reversal may arise even when $\beta u(c^*)^{\frac{1}{m-1}} \not< 1$. From the representation, the choice pattern is equivalent to the inequalities $u(c) > \beta\delta u(c+d)^{\frac{m}{m-1}}$ and $u(c)^{\frac{m}{m-1}} < \delta u(c+d)^{\frac{m}{m-1}}$. To see that these inequalities can hold simultaneously, observe that if $u(c+d)$ is less than one, then its convex transformation $u(c+d)^{\frac{m}{m-1}}$ must be less than $u(c+d)$. If the small payoff $c+d$ is discounted very severely, then the inequalities can hold even if $\beta \geq 1$. Indeed, if $\beta = 1$, preference reversals arise in the model purely because smaller rewards are discounted at a steeper rate than larger rewards.

6.3 Consumption Smoothing vs Preference for Increasing Sequences

Loewenstein and Prelec [32] demonstrate that subjects prefer increasing sequences of consumption to constant or decreasing sequences with the same present value. Assuming three periods, our next proposition shows that if self 2 is better off than self 1, then under certain conditions the agent may be willing to reduce self 1's welfare to improve self 2's further, suggesting a preference for increasing sequences.[19]

Proposition 2 *Assume that $\succsim$ admits a homogeneous CE representation. Suppose there are only three periods and u is linear. If $c_1 < c_2$ and if a_2/a_1 is sufficiently close to one, then for all ε in some positive interval,*

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

The idea is simply that a convex transformation $\frac{m}{m-1}$ can cause the marginal utility at time t to be increasing.[20]

The evidence for increasing sequences not withstanding, it is standard in economics to assume consumption smoothing. This is formally defined by the convexity of upper contour sets:

Definition 7 *A preference $\succsim$ exhibits consumption smoothing if for any $\alpha \in [0,1]$ and for all deterministic streams x, $y \in C^{T+1}$ and $\alpha x + (1-\alpha)y \in C^{T+1}$,*

$$x \sim y \implies \alpha x + (1-\alpha)y \succsim x.$$

[19] Proof: The desired preference obtains if:

$$\kappa_1[(u(c_1)+\varepsilon)^{\frac{m}{m-1}} - (u(c_1)-\varepsilon)^{\frac{m}{m-1}}] < \kappa_2[(u(c_2)+\varepsilon)^{\frac{m}{m-1}} - (u(c_2)-\varepsilon)^{\frac{m}{m-1}}].$$

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

[20] Note that whether marginal utility from consumption increases or not depends on both $\frac{m}{m-1}$ and the curvature of vNM function u over consumption. In the proposition, we assume that u is linear over consumption.

We now study conditions under which the homogeneous CE model exhibits consumption smoothing. As shown in Theorem 2, the utility at time t in this model is given by $D_{u(x_t)}(t)u(x_t) = \kappa_t u(x_t)^{\frac{m}{m-1}}$ if $x_t \leq \overline{x}_t := u^{-1}(m a_t \overline{d}_t^{m-1})$, and $D_{u(x_t)}(t)u(x_t) = \overline{d}_t u(x_t)$ otherwise.

Proposition 3 *Assume that $\succsim$ admits a homogeneous CE representation. If $u(c)^{\frac{m}{m-1}}$ is concave in $c \in \mathbb{R}_+$, then $\succsim$ exhibits consumption smoothing. Conversely, if $\succsim$ exhibits consumption smoothing, then at least T of functions $u(x_0)$, $D_{u(x_1)}u(x_1), \cdots, D_{u(x_T)}u(x_T)$ are concave. Moreover, $u : \mathbb{R}_+ \to \mathbb{R}_+$ is concave if there are $t, s \geq 1$ such that $\overline{x}_t \neq \overline{x}_s$.*

See the supplementary appendix (Noor and Takeoka [36]) for the proof. The first part of Proposition 3 is a sufficient condition for consumption smoothing behavior. Since $z^{\frac{m}{m-1}}$ is a convex transformation, the presumption requires that u is concave and its curvature dominates the convexity of $z^{\frac{m}{m-1}}$. The converse is a direct implication of Yaari [53].[21] If $D_{u(x_t)}(t)$ is independent of $u(x_t)$ (as in the DU model), concavity of $D_{u(x_t)}(t)u(x_t)$ immediately implies concavity of $u(x_t)$. The last statement in the Proposition is obtained from the fact that if the interval domains of two concave functions have a non-trivial overlap, then the function is concave on the whole domain.

In the DU model, it is well-known that consumption smoothing is equivalent to concave u. According to the Proposition 3, this observation is "generically" carried over to the case of homogeneous CE representations. The only exception is the case that $\overline{x}_t = \overline{x}_s$ for all t, s. [22]

6.4 Annuitization Puzzle and Concentration Bias

In a seminal paper, Yaari [52] shows that, under some assumptions, rational agents with no bequest motive should convert all their retirement wealth into annuities at retirement. Subsequent literature (see Benartzi et al [10] for references) shows that, under much weaker assumptions, households should hold a substantial proportion of their wealth in annuities. The literature also empirically documents the absence of a demand for annuities, which is dubbed the *annuitization puzzle*. In an experimental setting, Dertwinkel-Kalt et al [15] demonstrate a related behavior: a preference for concentrating outcomes in one period over spreading them, which they dub *concentration bias*. Our model is consistent with these findings since it requires that smaller payments over a long horizon should be discounted at a higher rate than larger payments over a possibly smaller one.

[21] Yaari [53] shows that if an additively separable function $F(x_1, \cdots, x_S) = \sum_s f_s(x_s)$ is quasi-concave, then at least $S-1$ of functions $f_1, \cdots, f_S$ are concave. Yaari also provides a counterexample for the converse.

[22] When $\overline{x}_t = \overline{x}_s$ for all t, s, u is not necessarily concave even if $D_{u(x_t)}u(x_t)$ is concave for all t. Here is a counterexample: Let $u(x) = \sqrt{x}$ if $x \leq 1$ and $u(x) = x$ if $x > 1$. Assume $m = 2$, $a_t = \frac{1}{2}$, and $\overline{d}_t = 1$. Then, $\kappa_t = (m a_t)^{-\frac{1}{m-1}} = 1$, $\overline{x}_t = u^{-1}(m a_t \overline{d}_t^{m-1}) = 1$ for all t, and $D_{u(x_t)}(t)u(x_t) = \kappa_t u(x_t)^{\frac{m}{m-1}} = x_t$ if $x_t \leq 1$ and $D_{u(x_t)}(t)u(x_t) = x_t$ if $x_t > 1$. Therefore, $D_{u(x_t)}(t)u(x_t)$ is concave for all t, x_t, but u is not concave.

6.5 Magnitude Hypothesis in Consumption Smoothing

There is considerable evidence that consumption tends to respond to anticipated income increases more than what is implied by standard models of consumption smoothing. Moreover, this response is inversely correlated with the size or magnitude of anticipated income increases, that is, for small income changes, consumption tends to overreact to them, while consumption pattern tends to be consistent with consumption smoothing for medium or large income changes. This evidence is known as the *magnitude hypothesis in consumption smoothing* (Browning and Collado [13], Scholnick [45]), which has been attributed to bounded rationality or costs associated with the mental processing of small anticipated income changes.

The CE representation may predict a similar behavioral pattern where $\succsim$ exhibits preference for concentration in streams with small payoffs, whereas it exhibits preference for consumption smoothing otherwise. Consider the reduced form of the homogeneous CE model (Theorem 2). Note that $D_{u(x_t)}(t)u(x_t)$ has a flatter curvature up to $x_t \leq u^{-1}(\overline{r}_t)$ because of the convex transformation over u, and gets more concave beyond the threshold. For example, let $T = 2$ for simplicity. Assume $u(c) = \sqrt{c}$ and $m = 2$. Then, $U(0, x_1, x_2) = \kappa_1 x_1 + \kappa_2 x_2$ if both x_1 and x_2 are below the thresholds, $U(0, x_1, x_2) = \overline{d}_1\sqrt{x_1} + \kappa_2 x_2$ if x_1 is above the threshold and x_2 is below the threshold, and $U(0, x_1, x_2) = \overline{d}_1\sqrt{x_1} + \overline{d}_2\sqrt{x_2}$ if both x_1 and x_2 are beyond the thresholds. Thus, the agent is more likely to choose a skewed consumption stream over a smoothed consumption stream when these streams are small.[23]

6.6 Negative Time Preference

There is evidence of negative discount rates for both positive and negative outcomes. For instance, subjects in Loewenstein and Prelec [32] would rather have a fancy french dinner later than sooner, and subjects in Hardisty et al [24] prefer losing \$9.10 now over losing \$9 in a week. Negative discount rates have been interpreted in the literature in terms of anticipation: delaying positive consumption leads to savoring, whereas expediting negative consumption avoids dread. This interpretation can be recast in terms of empathy. We can readily extend our model to allow for $D_r(t) > 1$, permitting the possibility that the current self weights the future self's utility higher than her own, albeit at the cost of sufficient cognitive effort.[24] Thus, empathy can be used to model negative discount rates, coupled with the hypothesis that the discount rates can be magnitude-dependent.

[23] In this example, it is more appropriate to interpret zero as a reference point or base-line consumption and x as a gain from the reference point. See Noor and Takeoka [37] for such an extension.

[24] We can drop the Impatience axiom in the axiomatization of the GDU representation in the supplementary appendix (Noor and Takeoka [36]) to permit $D_r(t) > 1$. As long as $D_r(t)$ is bounded above for each t, a counterpart of Theorem 1 can be proved for an appropriate generalization of the CE model.

6.7 Cognitive Costs and Impatience

Our paper relates to the literature connecting impatience with cognitive abilities. Dohmen et al [16] show that people with lower cognitive abilities are more impatient. Shiv and Fedorikhin [46] report that cognitive load weakens self-control and makes subjects more present-biased. Becker and Mulligan [9] propose that future-oriented capital such as education or health investment can reduce impatience. Cognitive ability, cognitive load and education can be viewed as determinants of the marginal cognitive cost of empathy in our model, and therefore of impatience.

7 Self-Control and Procrastination

As explained in the Introduction, the CE model can be viewed as a multiple selves model that features self-control. Self-control in this context refers to a cognitive act by the current self that enhances empathy for future selves and enables her to forgo immediate consumption. In the context of procrastination (O'Donoghue and Rabin [39]) we illustrate how such self-control can influence choice. The proof of the results can be found in the supplementary appendix (Noor and Takeoka [36]).

Suppose the time horizon consists of periods 0, 1, 2 and 3. The agent has background consumption $b > 0$ in every period, and evaluates consumption by a weakly concave, strictly increasing utility index u satisfying $u(0) = 0$. Suppose there are two tasks that can be completed within the time horizon. Each task takes effort $e > 0$, where $b - 2e > 0$, and yields a return $R > 0$. Choices are made only in periods 0 and 2, and these determine the consumption in periods 1 and 3. In period 0, the agent may take on zero, one or both tasks. Either task can be completed either in period 0 or in period 2, and two tasks can be completed simultaneously. If the agent completes $n \in \{0, 1, 2\}$ tasks in period $t = 0, 2$, then the agent's utility is $u(b - ne)$ in period t and $u(b + nR)$ in period $t + 1$.

We study the dynamic behavior of sophisticated DU and CE agents. The dynamic behavior can be modelled as a subgame perfect equilibrium of an extensive-form game between self 0 and self 2. To simplify exposition, adopt the tie-break rule that if any self is indifferent between completing n vs m tasks, then she completes the higher number of tasks.

Begin with the DU benchmark. Denote by $V_t(n|m)$ the utility of self $t = 0, 2$ of completing n tasks in period t out of m uncompleted tasks in that period. We first show that if self 2 would be unwilling to complete 1 task, then in the DU model, no task will get completed by either self:

Proposition 4 *Consider a sophisticated DU agent. Suppose self 2 would not complete a task when there is one to be done. On the path of a unique equilibrium, neither self 0 nor self 2 completes any tasks.*

The result is driven by the concavity of u. If receiving R tomorrow does not compensate for exerting effort e today, then neither will $2R$ today compensate for effort $2e$ today.

Moreover, the problem of self 0 and self 2 are identical: each must consider the trade-off between today and tomorrow, from their respective temporal vantage points.

The behavior of the CE agent can differ substantially. Consider the homogeneous CE model where the optimal discount factor is given by $D_{u(c)}(t) = \kappa_t u(c)^{\frac{1}{m-1}}$ for the range of utilities being considered (Theorem 2). Let $U_t(n|m)$ denote the utility of self $t = 0, 2$ of completing n tasks in period t given that there are m uncompleted tasks in that period. We show that, under the hypothesis of Proposition 4, there are parameter values for which the CE agent would behave like the DU agent, but for other parameter values, the CE agent can exhibit novel behaviors.

Proposition 5 *Consider a sophisticated CE agent. If self 2 would not complete a task when there is one to be done, then the model permits three possibilities on the path of a unique equilibrium:*
(i) Neither of self 0 nor self 2 completes any tasks.
(ii) Self 0 completes no task and self 2 completes 2 tasks.
(iii) Self 0 completes 2 tasks.

The Proposition gives rise to two new possibilities beyond those in the DU model. The first is that self 0 procrastinates on two tasks in the knowledge that the high returns of completing the two tasks together will motivate self 2 to complete them. The second is that the returns of completing the two tasks together are high enough to motivate self 0 to complete them immediately herself. She does not wish to delay the returns, even though doing so would delay her effort as well.

In summary, the CE model implies that each self may have a stronger incentive to complete several tasks together. Moreover, a current self may exploit such attitudes of future selves and strategically procrastinate several tasks for future selves to complete.

A Appendix: Proof of Theorem 1

We showed in the text that the CE model reduces to a GDU model with MDI. To prove the converse, suppose $\succsim$ admits a GDU representation (u, D) that exhibits MDI.

Note that Z is either C or Δ. From now on, a generic element of Z is denoted by z regardless of whether consumption is deterministic or risky.

We typically use $r \in \mathbb{R}_+$ to denote the utility level $r = u(z)$ at some consumption level $z \in Z$. Denote the set of all utility levels by

$$R = u(Z) = \{r \in \mathbb{R}_+ \,|\, r = u(z) \text{ for some } z \in Z\}.$$

For any given $t > 0$, R is the domain of $D_r(t)$ when viewed as a function of r. The image of R in this function is:

$$I(t) = \{\delta \in [0,1] \,|\, \delta = D_{u(z)}(t) \text{ for some } z \in Z\}.$$

Say that a non-empty closed interval $[s_i, s_j] \subset R$ (with $s_i < s_j$) in the domain is a *c-interval* if D is constant over it:

$$D_r(t) = D_{r'}(t) \text{ for all } r, r' \in [s_i, s_j] \subset R.$$

Since a c-interval is the inverse image of a singleton $\delta \in I(t)$, it is always closed by continuity of $D_r(t)$ in r. Say that a non-empty open interval $(s_i, s_j) \subset R$ in the domain is an *i-interval* if D is strictly increasing over it:

$$D_r(t) > D_{r'}(t) \text{ for all } r, r' \in (s_i, s_j) \subset R \text{ with } r > r'.$$

For any c-interval in R there corresponds a unique δ. For any i-interval in R there corresponds a unique open interval (δ, γ) in the range of D (since the image of an open interval in a strictly monotone continuous function is open). In general, there will not exist a finite number of c-intervals and i-intervals partitioning R. For any $r \in R$, there may exist a c-interval containing it, or an i-interval containing, but also, neither may be true: every neighborhood of r may contain c-intervals and i-intervals. Our argument below needs to account for this.

With this in mind, for any fixed t, let $I_c = \{\delta \in I(t) \,|\, \delta \text{ is the image of some } c\text{-interval}\}$, the set of discount factors whose inverse image generates a c-interval (as opposed to a point). For any $\delta \in I_c$, there are two key possibilities to consider. A discount factor $\delta \in I_c$ may be *locally isolated*: there may exist a neighborhood of δ (in the space $I(t)$) that does not intersect with $I_c \setminus \{\delta\}$. This would happen when the inverse image of δ is a c-interval that is sandwiched between two i-intervals.[25] Alternatively, δ may be an *accumulation point* of I_c, that is, there is a sequence in $I_c \setminus \{\delta\}$ converging to δ. This would arise if the inverse image of δ lies in a c-interval that is not sandwiched between two i-intervals.

The corresponding two sets defined by

$$\begin{aligned} &(i)\ I_c^{iso} = \{\delta \in I_c \,|\, \delta \text{ is locally isolated}\}, \\ &(ii)\ I_c^a = \{\delta \in I_c \,|\, \delta \text{ is an accumulation point of } I_c\}, \end{aligned}$$

form a partition of $I_c = I_c^{iso} \cup I_c^a$. Observe that points outside I_c may also be accumulation points of I_c. Define

$$\begin{aligned} &(iii)\ I_{nc}^a = \{\delta \in I(t) \setminus I_c \,|\, \delta \text{ is an accumulation point of } I_c\}, \\ &(iv)\ I^i = I(t) \setminus (I_c \cup I_{nc}^a). \end{aligned}$$

The above four sets form a partition of $I(t)$.

For any given $t > 0$, and any discount factor $\delta \in I(t)$ there is a (possibly singleton) subset of R that generates it, in the sense that $D_r(t) = \delta$ for all r in the subset. The next lemma defines a "cognitive marginal cost" $\varphi_t'(\delta)$ of the discount factor δ in terms of the minimum utility level in the subset, and establishes some properties.

[25] It cannot have another c-interval on either side since $D_r(t)$ is continuous in r.

Lemma 1 *Suppose $\succsim$ admits a GDU representation (u, D) that exhibits MDI. Consider the real-valued function $\varphi'_t : I(t) \to \mathbb{R}_+$ defined for each $\delta \in I(t)$ by*

$$\varphi'_t(\delta) = \min\{r \mid \delta = D_r(t)\} \in R. \tag{7}$$

This function satisfies:
(a) $\varphi'_t(\cdot)$ is strictly increasing.
(b) On any open interval $(\gamma_i, \gamma_j) \subset I(t)$ that is the image of an i-interval, $\varphi'_t(\cdot)$ is the inverse of $D_{(\cdot)}(t)$.
(c) $\varphi'_t(\cdot)$ is discontinuous at δ if and only if $\delta \in I_c$.
(d) $\varphi'_t(\cdot)$ is left continuous on $I(t)$.

Proof. By hypothesis, there is a GDU representation (u, D) such that $D_r(t)$ is continuous and weakly increasing in r. Consider the function φ'_t defined in the statement of the lemma. We establish the following.
(a) Show that $\varphi'_t(\cdot)$ is strictly increasing.
This holds because $D_r(t)$ is weakly increasing in r.
(b) On any open interval $(\gamma_i, \gamma_j) \subset I(t)$ that is the image of an i-interval, $\varphi'_t(\cdot)$ is the inverse of $D_{(\cdot)}(t)$, that is, $D_{\varphi'_t(\delta)}(t) = \delta$ for all $\delta \in (\gamma_i, \gamma_j)$.
Since $D_r(t)$ is strictly increasing on (γ_i, γ_j) it must be that, for any point $\delta \in (\gamma_i, \gamma_j)$, the inverse image $r(\delta)$ given by $D_{r(\delta)}(t) = \delta$ is a singleton. Consequently $\varphi'_t(\delta) = r(\delta)$.
(c) Show that $\varphi'_t(\cdot)$ is discontinuous at δ if and only if $\delta \in I_c$.
It is clear from definitions that if $\delta \in I_c$ then δ is a point of discontinuity for $\varphi'_t(\cdot)$. To show the converse, we establish the contrapositive. So take any $\delta \notin I_c$, specifically, $\delta \in I(t) \setminus I_c$. Since we defined $I^i = I(t) \setminus (I_c \cup I^a_{nc})$, note that $I(t) \setminus I_c = I^i \cup I^a_{nc}$. Assume first that $\delta \in I^i$. Since δ is not an accumulation point of I_c, there exists an open interval $(\gamma_i, \gamma_j) \subset I(t)$ that contains δ and does not intersect with I_c. The corresponding open interval in R is an i-interval, and since $D_r(t)$ is strictly increasing and continuous on such an interval, the inverse image of any point in (γ_i, γ_j) is a singleton and correspondingly, $\varphi'_t(\cdot)$ is continuous at δ. Next, assume $\delta \in I^a_{nc}$. Seeking a contradiction, suppose that $\varphi'_t(\cdot)$ is not continuous at δ. Since $\varphi'_t(\cdot)$ is strictly increasing, we have $\lim_{\delta^n \nearrow \delta} \varphi'_t(\delta^n) = s < t = \lim_{\delta^n \searrow \delta} \varphi'_t(\delta^n)$, which means that δ is the inverse image of a c-interval $[s, t]$, a contradiction. Therefore, $\varphi'_t(\cdot)$ is continuous at δ. This completes the proof of only-if part.
(d) Show that $\varphi'_t(\cdot)$ is left continuous.
From (c), we know that $\varphi'_t(\cdot)$ is continuous on $I(t) \setminus I_c$. It is enough to show that $\varphi'_t(\cdot)$ is left continuous on I_c. First take any $\delta \in I^{iso}_c$. Then, the c-interval associated with δ is sandwiched between two i-intervals. This c-interval can be denoted by $[\varphi'_t(\delta), s] \subset R$ in the domain. Then the adjacent i-interval $(t, \varphi'_t(\delta)) \subset R$ on its left has an image $(\gamma, \delta) \subset I(t)$ that is an open interval in the range. Clearly, for any sequence $\gamma^n \nearrow \delta$ that approaches δ from the left, its inverse image must eventually belong to the i-interval $(t, \varphi'_t(\delta))$ in the domain. Consequently, it must be that $\varphi'_t(\gamma^n) \nearrow \varphi'_t(\delta)$. Next, take any $\delta \in I^a_c$. Seeking a contradiction, suppose that $\varphi'_t(\cdot)$ is not left continuous at δ. Since $\varphi'_t(\cdot)$ is strictly increasing, we have $\varphi'_t(\delta) > t = \lim_{\delta^n \nearrow \delta} \varphi'_t(\delta^n)$, which means that the c-interval corresponding to δ

takes a form of $[t, s]$, which contradicts the definition of $\varphi_t'(\delta)$ because $\varphi_t'(\delta)$ is not the minimum in this interval. ■

In Lemma 1 we defined φ_t' on $I(t)$, which is the image of R in $D_{(\cdot)}(t)$. Let $\underline{d}_t = \inf I(t) = \inf_{r\in R} D_r(t)$ and $\overline{d}_t = \sup I(t) = \sup_{r\in R} D_r(t)$. Note that $\underline{d}_t$ is attained in $I(t)$ since R is bounded below by 0. Indeed, $\varphi_t'(\underline{d}_t) = 0$. However $\overline{d}_t$ may not be attained since R may not be bounded above, and $D_r(t)$ may approach $\overline{d}_t$ only asymptotically as $r \to \infty$. In this case, $\lim_{\delta \nearrow \overline{d}_t} \varphi_t'(\delta) = \infty$. Extend $\varphi_t'(\cdot)$ from $I(t)$ to $[0, 1]$ by setting $\varphi_t'(\delta) = 0$ for all $\delta < \underline{d}_t$, $\varphi_t'(\overline{d}_t) = \lim_{\delta \nearrow \overline{d}_t} \varphi_t'(\delta)$, and $\varphi_t'(\delta) = \infty$ for all $\delta > \overline{d}_t$.

We claim that $\varphi_t'(\cdot)$ is continuous at 0. If $\underline{d}_t > 0$, the claim holds because $\varphi_t'(\delta) = 0$ for all sufficiently small $\delta \geq 0$. Now assume $\underline{d}_t = 0$. By seeking a contradiction, suppose that $\varphi_t'(\cdot)$ is not continuous at 0. Since $\varphi_t'(\cdot)$ is strictly increasing, $0 = \varphi_t'(\underline{d}_t) = \varphi_t'(0) < \lim_{\delta\searrow 0} \varphi_t'(\delta)$, which means from (7) that $D_{u(z)}(t) = 0$ for all $0 < u(z) < \lim_{\delta\searrow 0} \varphi_t'(\delta)$. For all such $z \in Z$, let z^t be the stream that pays z at time t and 0 otherwise. Then, $U(z^t) = D_{u(z)}(t)u(z) = 0$, which violates the strict increasingness of the GDU representation.

Now, we define φ_t. Since $\varphi_t'(\cdot)$ is monotone on $[0, 1]$, it is Riemann-integrable. For any $D(t) \in [0, 1]$, define the cognitive cost function by the Riemann integral:

$$\varphi_t(D(t)) = \int_0^{D(t)} \varphi_t'(\delta)\, \mathrm{d}\delta.$$

Then, $\varphi_t(D(t)) = 0$ for any $D(t) \in [0, \underline{d}_t]$ and $\varphi_t(D(t)) = \infty$ for any $D(t) \in (\overline{d}_t, 1]$. If $\varphi_t'(\overline{d}_t) = \infty$ then $\varphi_t(\overline{d}_t) = \infty$.

Since φ_t' is monotone increasingly, φ_t is a convex function on $[0, 1]$. Since $\varphi_t'(\cdot)$ is strictly increasing on $I(t)$, φ_t is strictly convex on $[\underline{d}_t, \overline{d}_t]$. Moreover,

Lemma 2 *$\varphi_t(\cdot)$ is differentiable at δ if $\delta \notin I_c$.*

Proof. By property (c) of $\varphi_t'(\cdot)$ (in Lemma 1), $\bar{\delta} \notin I_c$ if and only if $\varphi_t'(\cdot)$ is continuous at $\bar{\delta}$. By the fundamental theorem of calculus, the derivative exists at $\bar{\delta}$ and satisfies

$$\frac{\mathrm{d}\varphi_t}{\mathrm{d}\delta}(\bar{\delta}) = \lim_{\delta^n \to \bar{\delta}} \frac{\int_{\bar{\delta}}^{\delta^n} \varphi_t'(\delta)\, \mathrm{d}\delta}{\delta^n - \bar{\delta}} = \varphi_t'(\bar{\delta}). \tag{8}$$

Since $\varphi_t'(\cdot)$ is not continuous at all points, $\varphi_t(\cdot)$ is not differentiable at all points. ■

Lemma 3 *$\varphi_t' \leq \varphi_{t+1}'$.*

Proof. Since (u, D) is a GDU representation, $D_r(t)$ is weakly decreasing in t for all fixed r. Thus, for all $\delta \in [0, 1]$, if $D_r(t) = \delta = D_{r'}(t+1)$ holds, then $D_{r'}(t+1) = D_r(t) \geq D_r(t+1)$. Since $D_r(t+1)$ is weakly increasing in r, we must have $r' \geq r$. Hence, by definition of φ_t', $\varphi_t'(\delta) \leq \varphi_{t+1}'(\delta)$, as desired. ■

Lemma 4 *Suppose $\succsim$ admits a GDU representation (u, D). Then $D_r(t)$ is weakly increasing and continuous in r for each t iff there exists a basic representation $(u, \{\varphi_t\}_{t=1}^T)$ such that for each z, t,*

$$D_{u(z)}(t) = \arg \max_{D(t)\in[0,1]} \{D(t)u(z) - \varphi_t(D(t))\}.$$

Proof. First establish sufficiency. Consider φ_t obtained by Lemma 1. Take any $r \in R$ and note that

$$\max_{D(t)\in[0,1]} D(t)r - \varphi_t(D(t)) = \max_{D(t)\in[0,1]} \int_0^{D(t)} (r - \varphi_t'(\delta))\, \mathrm{d}\delta.$$

We show that, since r is constant and $\varphi_t'(\cdot)$ is increasing, the solution to this maximization problem is given by

$$D_r^*(t) = \max\{\delta \,|\, r \geq \varphi_t'(\delta)\}.$$

First we show that the set $I_r = \{\delta \in [0,1] \,|\, r \geq \varphi_t'(\delta)\}$ contains its supremum and that this supremum is finite. Since this set is non-empty and bounded above, it admits a supremum, denoted by δ^*. By definition of the supremum, for all sufficiently large n, there exists $\delta_n \in I_r$ such that $\delta^* - \frac{1}{n} < \delta_n \leq \delta^*$. As $n \to \infty$, $\delta_n \to \delta^*$. Moreover, since $\delta_n \in I_r$, $\varphi_t'(\delta_n) \leq r$. By the left continuity of $\varphi_t'(\cdot)$, we have $\varphi_t'(\delta^*) \leq r$, or $\delta^* \in I_r$. Therefore, the maximum of I_r is attained, and $D_r^*(t) = \delta^*$.

We show next that $D_r(t)$ must equal the solution $D_r^*(t)$ to this maximization problem. If $r \leq \lim_{\delta \searrow \underline{d}_t} \varphi_t'(\delta)$, then the maximizer is uniquely given by $\underline{d}_t$, and so $D_r^*(t) = \underline{d}_t = D_r(t)$. Suppose henceforth that $r > \lim_{\delta \searrow \underline{d}_t} \varphi_t'(\delta)$. If r does not belong to any c-interval, then $\varphi_t'(D_r(t)) = r$ and φ_t is differentiable at $D_r(t)$. By (8) the FOC:

$$r = \varphi_t'(D_r(t)), \tag{9}$$

is satisfied at exactly $D_r^*(t) = \max\{\delta \,|\, r \geq \varphi_t'(\delta)\} = D_r(t)$. On the other hand, if r belongs to a c-interval, then there is a kink in φ_t at $D_r(t)$ with r contained between the left derivative and right derivative of φ_t at $D_r(t)$. Since φ_t is strictly convex on $[\underline{d}_t, \overline{d}_t]$, it must be that for any $\delta \leq D_r(t) < \gamma$,

$$\varphi_t'(\delta) \leq r < \varphi_t'(\gamma).$$

Consequently the solution must be exactly $D_r^*(t) = \max\{\delta \,|\, r \geq \varphi_t'(\delta)\} = D_r(t)$.

Conclude with the proof of necessity: show that if D is obtained from such a maximization problem, then it must be weakly increasing and continuous. Since the objective function is strictly concave, the solution is unique if it exists. First observe that if $\varphi(\overline{d}_t)$ is finite, a maximizer $D_r(t)$ always exists since it is obtained by maximizing a continuous objective function over the compact interval $[0, \overline{d}_t]$. By the Maximum theorem, together with the uniqueness of the solution, $D_r(t)$ is continuous in r. If $\varphi_t(\overline{d}_t) = \infty$, φ_t diverges to infinity as $\delta \to \overline{d}_t$. Since φ_t is strictly increasing and strictly convex on $(\underline{d}_t, \overline{d}_t)$, for all $r > 0$, there exists a unique $\delta(r) \in (\underline{d}_t, \overline{d}_t)$ satisfying $r\delta(r) = \varphi_t(\delta(r))$. As $\delta(r)$ is continuous and the domain of the optimization is effectively restricted to $[0, \delta(r)] = \{\delta | r\delta \geq \varphi_t(\delta)\}$, again, by the Maximum theorem, $D_r(t)$ is continuous in r.

For any r, since φ_t' is strictly increasing, the solution is given by $D_r(t) = \max\{\delta \,|\, r \geq \varphi_t'(\delta)\}$. The maximum exists because φ_t' is left continuous. This solution is weakly increasing in r because φ_t' is strictly increasing. ■

B Appendix: Proof of Theorem 2

Suppose that the preference admits a homogeneous CE representation with corresponding parameters $m, a_t, \overline{d}_t$. We establish a corresponding GDU representation by solving the cognitive optimization problem. Let $\varphi_t(d) = a_t d^m$ on $d \in [0, \overline{d}_t]$. For each x, an optimal discount factor $D_{x_t}(t)$ is determined by $\max_{D(t)\in[0,\overline{d}_t]} D(t)u(x_t) - \varphi_t(D(t))$. Define $\overline{r}_t := \varphi_t'(\overline{d}_t) = ma_t\overline{d}_t^{m-1}$.

Let x_t be such that $u(x_t) \leq \overline{r}_t$. By the FOC, we have $u(x_t) = ma_t D(t)^{m-1}$, or

$$D_{x_t}(t) = \left(\frac{u(x_t)}{ma_t}\right)^{\frac{1}{m-1}} = \kappa_t u(x_t)^\theta,$$

where $\theta := \frac{1}{m-1}$ and $\kappa_t := (ma_t)^{-\theta}$. Thus, for any $r \leq \overline{r}_t$, $D_r(t) = \kappa_t r^\theta$. Since $D_{\alpha r}(t) = \kappa_t(\alpha r)^\theta = \alpha^\theta D_r(t)$ for all $\alpha \in (0,1]$, $D_r(t)$ is homogeneous.

Next suppose $u(x_t) > \overline{r}_t := \varphi_t'(\overline{d}_t)$. Then the boundary constraint $D(t) \leq \overline{d}_t$ is binding and we have $D_{x_t} = \overline{d}_t$. Conclude that

$$D_r(t) = \begin{cases} \kappa_t r^\theta & \text{if } 0 \leq r \leq \overline{r}_t \\ \overline{d}_t & \text{if } r > \overline{r}_t \end{cases}.$$

Note that $D_r(t)$ is continuous in r, and in particular is not discontinuous at $r = \overline{r}_t$: by the preceding definitions (namely, $\overline{r}_t = ma_t\overline{d}_t^{\frac{1}{\theta}}$ and $\kappa_t = (ma_t)^{-\theta}$) it must be that $\overline{d}_t = \kappa_t\overline{r}_t^\theta$, as desired.

For the converse, consider a GDU model (u, D) as in part (b) of the Theorem. We show that D can be written as the solution to a cognitive optimization problem with the desired form for the cost function φ_t.

Lemma 5 *Define $\overline{R}(t) := [0, \overline{r}_t]$. There is $\theta \in \mathbb{R}$ s.t. for any $t > 0$, $r \in \overline{R}(t)$ and $\alpha \in (0,1]$,*

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

Proof. Homogeneity of D_r requires that for any $\alpha \in (0,1]$ there is $h(\alpha)$ s.t. for any $r \in \overline{R}(t)$,

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

Trivially, $h(1) = 1$. By continuity of D in r (required by GDU representations), $h(\alpha)$ is a continuous function defined on $(0,1]$. Moreover, we find that $h(\alpha\gamma)D_r(t) = D_{\alpha\gamma r}(t) = h(\alpha)D_{\gamma r}(t) = h(\alpha)h(\gamma)D_r(t)$. Indeed, h satisfies the multiplicative Cauchy equation:

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

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

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

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

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

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

Lemma 6 *For any $t > 0$, there exists $\kappa_t > 0$ such that for all $r \in \overline{R}(t)$,*

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

Moreover, $\theta > 0$ and κ_t is decreasing in t.

Proof. Take any $r \in \overline{R}(t)$. Then $r \leq \overline{r}_t$. By Lemma 5, $D_r(t) = D_{\frac{r}{\overline{r}_t}\overline{r}_t}(t) = \left(\frac{r}{\overline{r}_t}\right)^\theta D_{\overline{r}_t}(t)$. We obtain the expression $D_r(t) = \kappa_t r^\theta$ by letting $\kappa_t := \left(\frac{1}{\overline{r}_t}\right)^\theta D_{\overline{r}_t}(t)$. Since $\overline{R}(t)$ is a non-trivial interval and $D_r(t)$ is strictly increasing on it, it must be that $\theta > 0$. Since $D_r(t)$ is decreasing in t it must be that κ_t is decreasing in t. ■

Lemma 7 *D is the solution to the cognitive optimization wrt to some φ_t defined by $a_t > 0$, $m > 1$ and $\varphi_t(d) = a_t d^m$ for all $d \leq \overline{d}_t$. Moreover, $a_{t+1} \geq a_t$.*

Proof. By Lemma 6, $D_r(t) = \kappa_t r^\theta$ for all $r \in \overline{R}(t)$ where $\kappa_t > 0$ and $\theta > 0$. Using the FOC (9), define φ_t on $[0, \overline{d}_t]$ as follows. For all $r \in \overline{R}(t)$, let $r = \varphi_t'(\kappa_t r^\theta)$, so that $\varphi_t'(d) = \left(\frac{d}{\kappa_t}\right)^{\frac{1}{\theta}}$. Together with $\varphi_t(0) = 0$, we have

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

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

Since κ_t is decreasing in t (by Lemma 6), we have $\varphi_t(d) \leq \varphi_{t+1}(d)$. Then $a_t d^m = \varphi_t(d) \leq \varphi_{t+1}(d) = a_{t+1}d^m$, implying $a_t \leq a_{t+1}$, as desired. ■

We have established that the cost function has a power form on $[0, \overline{d}_t]$, where $\overline{d}_t = D_{\overline{r}_t}(t) = \kappa_t \overline{r}_t^\theta$. Since $D_r(t)$ is constant beyond $\overline{r}_t$, we can set $\varphi_t(d) = \infty$ for all $d \in (\overline{d}_t, 1]$. Then, by the FOC, $\overline{d}_t$ is optimal for all $r > \overline{r}_t$. Thus, we have a homogeneous CE representation.

C Appendix: Proof of Theorems 3 and 4

Suppose $\succsim$ admits a smooth GDU representation (u, D).

Lemma 8 *If $D_r(t)$ and $u(c)$ are differentiable, then for any stream $x \in X$,*

$$MRS_x(t) = \left(\frac{\partial D_r(t)}{\partial r}|_{r=u(x_t)}\right)\frac{u'(x_t)u(x_t)}{u'(x_0)} + D_{u(x_t)}(t)\frac{u'(x_t)}{u'(x_0)}.$$

Proof. For notational convenience, let $D_r(t)$ denote $d_t(r)$. For any m, use the representation to obtain

$$\begin{aligned}
&(x_0 + \psi_x(m,t), x_1, \cdots, x_t, \cdots, x_T) \sim (x_0, x_1, \cdots, x_t + m, \cdots, x_T)\\
\Longleftrightarrow\ & u(x_0 + \psi_x(m,t)) + d_t(u(x_t))u(x_t) + \sum_{\tau \neq t} d_\tau(u(x_\tau))u(x_\tau)\\
&= u(x_0) + d_t(u(x_t + m))u(x_t + m) + \sum_{\tau \neq t} d_\tau(u(x_\tau)u(x_\tau).
\end{aligned}$$

Differentiate this equality with respect to m to obtain

$$\begin{aligned}
&u'(x_0 + \psi_x(m,t))\frac{\partial \psi_x(m,t)}{\partial m}\\
&= d_t'(u(x_t + m))u'(x_t + m)u(x_t + m) + d_t(u(x_t + m))u'(x_t + m)\\
\Longleftrightarrow\ & \frac{\partial \psi_x(m,t)}{\partial m} = d_t'(u(x_t + m))\frac{u'(x_t + m)u(x_t + m)}{u'(x_0 + \psi_x(m,t))} + d_t(u(x_t + m))\frac{u'(x_t + m)}{u'(x_0 + \psi_x(m,t))}.
\end{aligned}$$

Moreover, apply L'Hospital's rule to the definition $\phi_x(m,t) := \frac{\psi_x(m,t)}{m}$ to obtain:

$$MRS_x(t) := \lim_{m \to 0} \frac{\psi_x(m,t)}{m} = \frac{\partial \psi_x(m,t)}{\partial m}|_{m=0}.$$

Therefore, by evaluating at $m = 0$ (in which case it must be that $\psi_x(m,t) = 0$) we obtain

$$MRS_x(t) = \frac{\partial \psi_x(m,t)}{\partial m}|_{m=0} = d_t'(u(x_t))\frac{u'(x_t)u(x_t)}{u'(x_0)} + d_t(u(x_t))\frac{u'(x_t)}{u'(x_0)}$$

as desired. ■

In particular, Lemma 8 implies that for all steams x on the diagonal,

$$MRS_x(t) = \left(\frac{\partial D_r(t)}{\partial r}|_{r=u(x_t)}\right)u(x_t) + D_{u(x_t)}(t) = [D_r(t)r]'\Big|_{r=u(x_t)}. \tag{10}$$

The result can also be used to provide foundations for the CRRA form for u.[26]

Lemma 9 *$\succsim$ satisfies Constant MRS iff for any $t > 0$, $D_r(t)$ is constant wrt r.*

Proof. If $D_r(t) = D(t)$, (10) implies $MRS_x(t) = D(t)$. Thus, Constant MRS holds.

Conversely, by Constant MRS, for all t and deterministic streams x on the diagonal, we have $MRS_x(t) = MRS_{\alpha x}(t) \geq 0$. Denote $d_t(r) = D_r(t)$. By (10), for all $r, \tilde{r} > 0$, $[d_t(r)r]' = [d_t(\tilde{r})\tilde{r}]' = D(t)$ for some constant $D(t) \geq 0$, which means that $d_t(r)r$ is an affine function. Thus, $d_t(r)r = D(t)r + A_t$ for some $D(t) \geq 0$ and $A_t \in \mathbb{R}$. Since $U_t(c) = D_{u(c)}(t)u(c)$ for all $c > 0$,

$$0 = \lim_{c \to 0} U_t(c) = \lim_{c \to 0} D_{u(c)}(t)u(c) = \lim_{c \to 0}(D(t)u(c) + A_t) = A_t.$$

Therefore, we have $D_r(t) = D(t) + \frac{A_t}{r} = D(t)$, as desired. ∎

Lemma 10 *$\succsim$ satisfies Increasing MRS iff for all t, $D_r(t)r$ is convex in r. Moreover, the latter implies that $D_r(t)$ is increasing in r.*

Proof. By (10), Increasing MRS is equivalent to

$$[D_r(t)r]'\Big|_{r=u(x_t)} \geq [D_r(t)r]'\Big|_{r=u(y_t)}$$

for all $u(x_t) \geq u(y_t) > 0$. Defining $f(r) \equiv D_r(t)r$, the above condition means that $f'(r)$ is increasing, that is, f is a convex function. Moreover, by strict monotonicity of GDU, $f(r)$ is strictly increasing. Finally, since $f(u(c)) = D_{u(c)}(t)u(c) = U_t(c)$ for all $c \in C$,

$$\lim_{r \to 0} f(r) = \lim_{c \to 0} f(u(c)) = \lim_{c \to 0} U_t(c) = 0.$$

Note that we can write $D_r(t) = \frac{f(r)}{r}$ for all $r > 0$. Thus,

$$\frac{\mathrm{d}D_r(t)}{\mathrm{d}r} = \left(\frac{f(r)}{r}\right)' = \frac{f'(r)r - f(r)}{r^2}.$$

By convexity of f, $f(0) - f(r) \geq f'(r)(0 - r)$ for all fixed $r > 0$, that is, $f'(r)r - f(r) \geq 0$. Therefore, $\frac{\mathrm{d}D_r(t)}{\mathrm{d}r} \geq 0$, that is, $D_r(t)$ is increasing, as desired. ∎

[26] When $Z = C$, a CRRA utility index u can be obtained from more basic conditions as follows. Adopt Assumption 1 so that MRS is well-defined. For any stream $x \in X$ let $\alpha^0 \circ x$ define the stream that scales down only immediate consumption, that is, $\alpha^0 \circ x = (\alpha \circ x_0, x_1, ...x_T)$. Then u is CRRA if and only if it satisfies the following MRS-homogeneity condition: for any $x, y \in X$,

$$MRS_{\alpha^0 \circ x}(t) = \beta MRS_x(t) \Longrightarrow MRS_{\alpha^0 \circ y}(t) = \beta MRS_y(t).$$

That is, scaling down immediate consumption by α changes the MRS by a proportion β that is independent of the stream. By Lemma 8, we know that $MRS_x(t) = \frac{1}{u'(x_0)}[\frac{\partial D_{u(x_t)}(t)}{\partial r}u'(x_t)u(x_t) + D_{u(x_t)}(t)u'(x_t)]$, where the term in the square bracket does not depend on x_0. Thus, if MRS is homogeneous in x_0, then $u'(x_0)$ must be homogeneous. By Euler's Theorem, $u(x_0)$ must be homogeneous, as desired.

By Theorem 1 and Lemma 10, $\succsim$ satisfies Increasing MRS if and only if it admits a CE representation with $D_r(t)r$ being convex in r. It is easy to see that φ_t, constructed in Lemma 1 in the proof of Theorem 1, is maximal: since $R = u(C)$, (7) implies $\varphi_t'(\delta) \in u(C)$ for all $\delta \in I(t)$, which in turn implies $\sup_{\delta\in[0,\overline{d}_t]} \varphi_t'(\delta) \leq \sup_{c\in C} u(c)$.

Uniqueness

Consider two maximal CE representations $(u^i, \{\varphi_t^i\})$, $i = 1, 2$, that represent the same preference. Their reduced forms are denoted by $U^i(x) = u^i(x_0) + \sum_{t\geq 1} D^i_{u^i(x_t)}(t)u^i(x_t)$, where $D^i_{u^i(x)}$ is an optimal discount function. Since these are GDU representations that represent the same preference, by the uniqueness shown in Noor and Takeoka [36, Theorem 1], there exists $\lambda > 0$ such that (i) $u^2 = \lambda u^1$, and (ii) for all $c \in C$ and t,

$$D^1_{u^1(c)}(t) = D^2_{u^2(c)}(t). \tag{11}$$

This directly implies the uniqueness of DU representation.

By Lemma 4 in the proof of Theorem 1, for $i = 1, 2$,

$$D^i_{u^i(c)}(t) = \max\{\delta \,|\, u^i(c) \geq (\varphi_t^i)'(\delta)\}. \tag{12}$$

First, we show $\underline{d}_t^1 = \underline{d}_t^2 = \underline{d}_t$ and $\overline{d}_t^1 = \overline{d}_t^2 = \overline{d}_t$. Take any c such that $u^1(c) \leq \lim_{\delta\searrow \underline{d}_t^1}(\varphi_t^1)'(\delta)$. From (12), $D^1_{u^1(c)}(t) = \underline{d}_t^1$. Together with (11), we conclude $\underline{d}_t^1 = \underline{d}_t^2$. Again, from (11),

$$\overline{d}_t^1 = \lim_{c\to\infty} D^1_{u^1(c)}(t) = \lim_{c\to\infty} D^2_{u^2(c)}(t) = \overline{d}_t^2.$$

Next, take any $\delta \in (\underline{d}_t, \overline{d}_t)$. We want to show $(\varphi_t^2)'(\delta) = \lambda(\varphi_t^1)'(\delta)$ for all such δ, which implies $\varphi_t^2 = \lambda\varphi_t^1$. By seeking a contradiction, suppose that there exists δ^* with $(\varphi_t^2)'(\delta^*) > \lambda(\varphi_t^1)'(\delta^*)$. Since $(u^2, \{\varphi_t^2\})$ is maximal and $u_2(C)$ is an interval, there exists $c \in C$ such that $\frac{1}{\lambda}(\varphi_t^2)'(\delta^*) > u^1(c) > (\varphi_t^1)'(\delta^*)$. By the first inequality and

$$D^2_{u^2(c)}(t) = \max\{\delta \,|\, u^1(c) \geq (1/\lambda)(\varphi_t^2)'(\delta)\},$$

we have $D^2_{u^2(c)}(t) < \delta^*$. Since $(\varphi_t^1)'$ is strictly increasing, the above second inequality and $D^1_{u^1(c)}(t) = \max\{\delta \,|\, u^1(c) \geq (\varphi_t^1)'(\delta)\}$ imply $D^1_{u^1(c)}(t) \geq \delta^*$, which contradicts to (11).

D Appendix: Proof of Theorems 5 and 6

We prove the results under a weaker assumption than Assumption 2.

Assumption 3 *$\succsim$ over $X = Z^{T+1}$ admits a GDU representation (u, D) where*
(1) u is unbounded on C,
(2) there is a continuous scaling operation $(\alpha, p) \mapsto \alpha \circ p \in Z$ that satisfies

(i) $\alpha \circ p = 0$ *when* $\alpha = 0$,

(ii) $\gamma \circ (\alpha \circ p) = \gamma\alpha \circ p$.

(3) there exists a strictly increasing continuous function $\beta : [0,1] \to [0,1]$ *satisfying: for any* $p \in Z$ *and* $\alpha \in [0,1]$,

$$u(\alpha \circ p) = \beta_\alpha u(p).$$

If $Z = C$ and u is assumed to be homogeneous of the form $u(c) = \gamma c^\theta$ for some $\theta > 0$, then $\beta_\alpha = \alpha^\theta$. If $Z = \Delta$ and u is assumed to be an expected utility, then $\beta_\alpha = \alpha$.

Some simple implications are:

Lemma 11 *Under Assumption 3,*

(i) $\beta_\alpha = 1$ *when* $\alpha = 1$, *and* $\beta_\alpha = 0$ *when* $\alpha = 0$.

(ii) For any $\beta \in [0,1]$ *there exists a unique* $\alpha \in [0,1]$ *s.t.* $\beta_\alpha = \beta$.

(iii) For any $\alpha \in (0,1]$, $p \succsim p' \iff \alpha \circ p \succsim \alpha \circ p'$.

Proof. Trivially, $\beta_\alpha = 1$ when $\alpha = 1$. For any $u(p) > 0$, we have $0 = u(0 \circ p) = \beta_\alpha u(p)$ and so it must be that $\beta_\alpha = 0$ when $\alpha = 0$. This establishes (i). To establish (ii) observe that by continuity and monotonicity of β_α, it is a homeomorphism. Moreover, given (i), β_α maps $[0,1]$ to $[0,1]$. The assertion follows. Assertion (iii) follows directly from property (3) in the Assumption. ■

The following lemma highlights the implication of Weak Homotheticity and Homotheticity for this GDU representation.

Lemma 12 *Under Assumption 3,* $\succsim$ *satisfies Weak Homotheticity (resp. Homotheticity) if and only if* $D_r(t)$ *is weakly increasing (resp. constant) in* r.

Proof. By the GDU representation, for any $p \in Z$ and $t > 0$, it must be that $U(p^t) = D_{u(p)}(t)u(p)$. Take any $\beta \in (0,1]$. By Lemma 11, there is $\alpha \in (0,1]$ s.t. $\beta = \beta_\alpha$ and in particular $\beta u(q) = u(\alpha \circ q)$ for any $q \in Z$. But then, for the dated reward p^t,

$$\begin{aligned} \beta D_{u(p)}(t)u(p) &= \beta U(p^t) = \beta u(c_{p^t}) = u(\alpha \circ c_{p^t}) \\ &\geq U(\alpha \circ p^t) = D_{u(\alpha \circ p)}(t)u(\alpha \circ p) = \beta D_{\beta u(p)}(t)u(p), \end{aligned}$$

where the inequality holds since Weak Homotheticity requires $\alpha \circ c_x \succsim \alpha \circ x$. Conclude that $D_{u(p)}(t) \geq D_{\beta u(p)}(t)$, that is, D is weakly increasing in utility of magnitude, as desired. Replacing the above inequality with an equality establishes that Homotheticity implies that D is constant. The converse directions of these claims are straightforward to establish. ■

Theorem 5 is established as part of this lemma. Theorem 6 obtains by invoking Theorem 1.

The uniqueness follows from the same proof of Theorems 3 and 4. First of all, under Assumption 2, $u(C)$ is unbounded above, whereby, φ_t is always maximal. If $Z = C$, the proof is exactly the same as before. If $Z = \Delta$, as shown by Noor and Takeoka [36, Theorem 2], GDU representations on Δ^{T+1} admit the desired uniqueness property. The subsequent argument is the same as before.

E Appendix: Proof of Theorem 7

E.1 Intermediate Characterization

Proposition 6 *Under Assumption 3, $\succsim$ over $X = Z^{T+1}$ satisfies X^*-Regularity if and only if it admits a GDU representation such that $D_r(t)$ is strictly increasing up to some threshold and constant thereafter.*

We first show sufficiency. By Theorem 1, since the preference $\succsim$ admits a CE representation, it also admits a GDU representation with $D_r(t)$ weakly increasing in r. Define

$$R^*(t) = \{r \mid r = u(p) \text{ for some } p^t \in X^*\} \subset \mathbb{R}^>, \tag{13}$$

where $\mathbb{R}^>$ denotes the set of strictly positive reals.

Lemma 13 *$R^*(t)$ is an nonempty interval with* $\inf R^*(t) = 0$.

Proof. By X^*-Regularity (i), for any $p^t \in X$ there is α s.t. $\alpha \circ p^t \in X^*$. Therefore X^* and $R^*(t)$ are nonempty. We show that $R^*(t)$ is an interval: Take any $r \in R^*(t)$. There exists $p^t \in X^*$ with $r = u(p)$. By X^*-Regularity (ii), $\alpha \circ p^t \in X^*$ for all $\alpha \in (0,1)$. Then, by Assumption 3, $\beta_\alpha r = u(\alpha \circ p) \in R^*(t)$ all $\alpha \in (0,1)$. In fact, by Lemma 11, $\beta r \in R^*(t)$ for all $\beta \in (0,1)$, and so $R^*(t)$ is an interval with $\inf R^*(t) = 0$. ■

Lemma 14 *$D_r(t)$ is strictly increasing on $R^*(t)$.*

Proof. Take any $r, s \in R^*(t)$ where $s < r$. By definition of $R^*(t)$, there exist $p^t \in X^*$ and $\alpha \in (0,1)$ with $r = u(p)$ and $s = u(\alpha \circ p)$. By definition of X^*, for all $\gamma \in (0,1)$, $\gamma \circ c_{p^t} \succ \gamma \circ p^t$. In particular, $\alpha \circ c_{p^t} \succ \alpha \circ p^t$. But, given Assumption 3, $\alpha \circ c_{p^t} \succ \alpha \circ p^t$ implies

$$\begin{aligned} &u(\alpha \circ c_{p^t}) > D_{u(\alpha \circ p)}(t) u(\alpha \circ p) \implies \beta_\alpha u(c_{p^t}) > \beta_\alpha D_{u(\alpha \circ p)}(t) u(p) \\ \implies &D_{u(p)}(t) u(p) > D_{u(\alpha \circ p)}(t) u(p) \implies D_{u(p)}(t) > D_{u(\alpha \circ p)}(t) \implies D_r(t) > D_s(t), \end{aligned}$$

as desired. ■

Lemma 15 *$D_r(t)$ is constant on $\mathbb{R}^> \setminus R^*(t)$.*

Proof. If $R^*(t) = \mathbb{R}^>$, there is nothing to prove. Thus, assume otherwise. Since $R^*(t)$ is an interval with $\inf R^*(t) = 0$, its complement in $\mathbb{R}^>$ is also an interval that is unbounded above. Take any $s < r$ with this interval. We want to show $D_s(t) = D_r(t)$. We do this in steps, with the first two being preliminary.

Step 1: Fix an arbitrary $p \in Z$. Show that if there exist $\alpha, \gamma \in (0,1)$ such that $\alpha \circ c_{p^t} \sim \alpha \circ p^t$ and $\gamma \circ c_{\alpha \circ p^t} \sim \gamma \circ (\alpha \circ p^t)$, then $\alpha\gamma \circ c_{p^t} \sim \alpha\gamma \circ p^t$.

By definition of $c_{\alpha \circ p^t}$, $c_{\alpha \circ p^t} \sim \alpha \circ p^t$. Thus, $c_{\alpha \circ p^t} \sim \alpha \circ p^t \sim \alpha \circ c_{p^t}$, where the last indifference holds by hypothesis. By Lemma 11, $c_{\alpha \circ p^t} \sim \alpha \circ c_{p^t}$ implies $\gamma \circ c_{\alpha \circ p^t} \sim \gamma\alpha \circ c_{p^t}$.

Finally, we have $\alpha\gamma \circ c_{p^t} \sim \gamma \circ c_{\alpha \circ p^t} \sim \gamma \circ (\alpha \circ p^t) = \alpha\gamma \circ p^t$, where the second indifference holds by hypothesis and the equality holds by the definition of scaling of streams and by Assumption 3. But then $\alpha\gamma \circ c_{p^t} \sim \alpha\gamma \circ p^t$, as desired.

Step 2: Show that for any x, $\alpha \circ c_x \succ \alpha \circ x$ implies $\eta \circ c_x \succ \eta \circ x$ for all $\eta \in (0, \alpha]$.

By hypothesis and by definition of a present equivalent, $\alpha \circ c_x \succ \alpha \circ x \sim c_{\alpha \circ x}$. For any $\gamma \in (0,1)$, let $\eta = \gamma\alpha \in (0, \alpha)$. By Lemma 11, $\alpha \circ c_x \succ c_{\alpha \circ x}$ implies $\gamma\alpha \circ c_x \succ \gamma \circ c_{\alpha \circ x}$. By Weak Homotheticity and Assumption 3,

$$\eta \circ c_x = \gamma\alpha \circ c_x \succ \gamma \circ c_{\alpha \circ x} \succsim \gamma \circ (\alpha \circ x) = \eta \circ x.$$

Recall the number r chosen at the beginning of the proof. Since u is unbounded above, there exists $p \in Z$ such that $r = u(p)$. As $r \in \mathbb{R}^{>} \setminus R^*(t)$, $p^t \notin X^*$. Let $A^{\sim} = \{\alpha \in (0,1) \,|\, \alpha \circ c_{p^t} \sim \alpha \circ p^t\}$ and $A^{\succ} = (0,1) \setminus A^{\sim}$. Since the CE model satisfies Weak Homotheticity, $\alpha \circ c_{p^t} \succsim \alpha \circ p^t$ for all $\alpha \in (0,1]$, and so it must be that $A^{\succ} = \{\alpha \in (0,1) \,|\, \alpha \circ c_{p^t} \succ \alpha \circ p^t\}$.

These sets satisfy the following properties:

(a) $A^{\sim} \neq \emptyset$.

As $p^t \notin X^*$, by definition of X^*, there exists $\alpha \in (0,1)$ with $\alpha \circ c_{p^t} \sim \alpha \circ p^t$. Hence, $A^{\sim} \neq \emptyset$.

(b) $A^{\succ} \neq \emptyset$.

By X^*-Regularity (i), for some $\gamma \in (0,1)$, $\gamma \circ p^t \in X^*$. To prove the claim, we show that $\alpha\gamma \circ c_{p^t} \succ \alpha\gamma \circ p^t$ for some $\alpha \in (0,1)$. Suppose by way of contradiction that $\alpha\gamma \circ c_{p^t} \sim \alpha\gamma \circ p^t$ for all $\alpha \in (0,1)$. It must be that $\alpha \circ c_{\gamma \circ p^t} \succ \alpha\gamma \circ p^t$ since by definition of X^* and Assumption 3, $\gamma \circ p^t \in X^*$ implies $\alpha \circ c_{\gamma \circ p^t} \succ \alpha \circ (\gamma \circ p^t) = \alpha\gamma \circ p^t$. Together then, $\alpha\gamma \circ c_{p^t} \sim \alpha\gamma \circ p^t$ and $\alpha \circ c_{\gamma \circ p^t} \succ \alpha\gamma \circ p^t$ imply $\alpha \circ c_{\gamma \circ p^t} \succ \alpha\gamma \circ c_{p^t}$, which in turn implies $c_{\gamma \circ p^t} \succ \gamma \circ c_{p^t}$ by Lemma 11. However, by Weak Homotheticity, $\gamma \circ c_{p^t} \succsim \gamma \circ p^t$, and so we obtain $c_{\gamma \circ p^t} \succ \gamma \circ c_{p^t} \succsim \gamma \circ p^t$, contradicting the definition of present equivalent, $c_{\gamma \circ p^t} \sim \gamma \circ p^t$.

(c) $A^{\succ}$ is an interval with $\inf A^{\succ} = 0$.

Take any $\alpha \in A^{\succ}$, that is, $\alpha \circ c_{p^t} \succ \alpha \circ p^t$. The claim follows from Step 2.

(d) $A^{\sim}$ is an interval with $\sup A^{\sim} = 1$.

Since $A^{\sim} = (0,1) \setminus A^{\succ}$, the claim follows from (c).

Recall again the numbers $s < r$ chosen at the beginning of the proof. Recall also p chosen above so as to satisfy $r = u(p)$. There exists $\alpha \in (0,1)$ such that $s = \beta_\alpha u(p) = u(\alpha \circ p)$.

Step 3: $\alpha \in A^{\sim}$.

Let $\alpha^* = \inf A^{\sim} > 0$. Take any sequence $\alpha^n \to \alpha^*$ with $\alpha^n \in A^{\sim}$. Since $\alpha^n \circ c_{p^t} \sim \alpha^n \circ p^t$, by the continuity of the scaling operation, $\alpha^* \circ c_{p^t} \sim \alpha^* \circ p^t$, that is, $\alpha^* \in A^{\sim}$. If $\alpha \geq \alpha^*$, we have $\alpha \in A^{\sim}$ by part (d), as desired. Seeking a contradiction, suppose $\alpha < \alpha^*$. Since $\beta_{\alpha^*} u(p) > \beta_\alpha u(p) = s \notin R^*(t)$, Lemma 13 implies $\alpha^* \circ p^t \notin X^*$. By definition of X^*, there exists $\gamma \in (0,1)$ such that $\gamma \circ c_{\alpha^* \circ p^t} \sim \gamma \circ (\alpha^* \circ p^t)$. Since $\alpha^* \circ c_{p^t} \sim \alpha^* \circ p^t$, by Step 1, $\alpha^*\gamma \circ c_{p^t} \sim \alpha^*\gamma \circ p^t$. Since $\alpha^*\gamma < \alpha^*$, this contradicts to $\alpha^* = \inf A^{\sim}$.

Step 4: The result.

Since $\alpha \circ c_{p^t} \sim \alpha \circ p^t$ by Step 3, Assumption 3 and the GDU representation imply that

$$\beta_\alpha D_{u(p)}(t) u(p) = D_{u(\alpha \circ p)}(t) u(\alpha \circ p) \implies D_{u(p)}(t) = D_{u(\alpha \circ p)}(t),$$

which in turn implies $D_r(t) = D_s(t)$, as desired. ■

We turn to necessity of Proposition 6. Assume that $\succsim$ admits a GDU representation where $D_r(t)$ is strictly increasing up to some threshold and constant thereafter. By Lemma 12, $\succsim$ satisfies Weak Homotheticity. To verify X^*-Regularity, take $p^t \in X^*$, so that $\alpha \circ c_{p^t} \succ \alpha \circ p^t$ for all $\alpha \in (0,1)$. Assumption 3 implies $D_{u(p)}(t) > D_{u(\alpha \circ p)}(t)$. In particular, by Lemma 11, $D_r(t)$ is strictly increasing for all $r \leq u(p)$. Thus, For any $\alpha, \gamma \in (0,1)$, $D_{\alpha u(p)}(t) > D_{\gamma \alpha u(p)}(t)$. Let $\widetilde{\alpha}$ and $\widetilde{\gamma} \in (0,1)$ denote their inverse image under β in Assumption 3, that is, $\alpha = \beta_{\widetilde{\alpha}}$ and $\gamma = \beta_{\widetilde{\gamma}}$. Then, the above strict inequality implies $\gamma D_{u(\widetilde{\alpha} \circ p)}(t) u(\widetilde{\alpha} \circ p) > D_{u(\widetilde{\gamma}\widetilde{\alpha} \circ p)}(t) u(\widetilde{\gamma}\widetilde{\alpha} \circ p)$, which means $\widetilde{\gamma} \circ c_{\widetilde{\alpha} \circ p^t} \succ \widetilde{\gamma} \circ (\widetilde{\alpha} \circ p^t)$. Since $\widetilde{\alpha}$ and $\widetilde{\gamma}$ vary through $(0,1)$, $\widetilde{\alpha} \circ p^t \in X^*$, that is, X^*-Regularity (ii) holds. Next, take $p^t \notin X^*$. There exists a sufficiently small $\alpha \in (0,1)$ such that $\alpha u(p)$ is below the threshold up to which $D_r(t)$ is strictly increasing. Since $D_{\alpha u(p)}(t) > D_{\gamma \alpha u(p)}(t)$ for all $\gamma \in (0,1)$, we can show $\widetilde{\alpha} \circ p^t \in X^*$ by the same argument as above. That is, X^*-Regularity (i) holds.

E.2 Result: Sufficiency

By Proposition 6, $\succsim$ admits a GDU representation such that $D_r(t)$ is strictly increasing in r on $R^*(t)$ (defined by (13)) and is constant otherwise. By Theorem 2, it suffices to show that $D_r(t)$ is homogeneous on $R^*(t)$:

Lemma 16 *For each $\alpha \in (0,1]$ there is ζ_α s.t. for any $p^t \in X^*$,*

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

Proof. We first note that for any $x \in X$ such that $x \succ 0$ and $\alpha \in (0,1]$, there exists a unique $\gamma_\alpha(x) \in (0,1]$ such that $\gamma_\alpha(x) \circ c_x \sim \alpha \circ x$: Since $x \succ 0$, Assumption 3 and Weak Homotheticity imply $c_x \succ \alpha \circ c_x \succsim \alpha \circ x$. Consequently by Lemma 11 and by Continuity and Monotonicity of GDU, the desired $\gamma_\alpha(x) \in (0,1]$ exists and is unique.

Take any dated reward $p^t \in X^*$. By definition of X^*, it must be that $p \succ 0$, and by Monotonicity of the GDU representation, $p^t \succ 0$. As noted, there exists $\gamma_\alpha(p^t) \in (0,1]$ such that

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

We make several observations about γ_α :

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

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

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

Since $c_{p^t} \sim p^t$ by definition of present equivalents, and since γ_α is defined by $\gamma_\alpha \circ c_{p^t} \sim \alpha \circ p^t$, it follows trivially that $\gamma_\alpha = 1$ when $\alpha = 1$. Moreover, by Assumption 3 and Monotonicity of GDU, γ_α must be strictly increasing in α, since $\alpha < \alpha'$ implies $\gamma_\alpha \circ c_{p^t} \sim \alpha \circ p^t \prec \alpha' \circ p^t \sim \gamma_{\alpha'} \circ c_{p^t}$. Finally, by the properties of the scaling operation in Assumption 3, $\alpha \to 0$ implies $\alpha \circ p^t = (\alpha \circ p)^t \to 0^t = 0$, and so it must be that $\lim_{\alpha \to 0} \gamma_\alpha = 0$.

(iii) γ_α is continuous in α.

By the representation,

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

that is, $U_t(\alpha \circ p) = \beta_{\gamma_\alpha} U_t(p)$. Since U_t and β are continuous (by Continuity of GDU and Assumption 3), it follows that γ_α is continuous.

(iv) γ_α satisfies

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

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

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

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

By Lemma 11, for any $\lambda \in (0,1)$, there is α_λ s.t. $\lambda u(p) = u(\alpha_\lambda \circ p)$. Defining $\zeta_\lambda = \frac{\beta_{\gamma_{\alpha_\lambda}}}{\beta_{\alpha_\lambda}}$ we obtain the desired expression. ■

E.3 Result: Necessity

By Proposition 6, it is enough to check whether X^*-Homogeneity holds. By Theorem 2, the reduced form of a homogeneous CE representation is given as

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

where

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

Therefore,

$$X^* = \{x \in X \mid \text{there exists some } t \geq 1 \text{ such that } 0 < u(x_t) \leq \overline{d}_t^{m-1} ma_t\}.$$

To show X^*-Homogeneity, take any dated reward $p^t \in X^*$. Then, given Assumption 3, $\gamma \circ c_{p^t} \sim \alpha \circ p^t$ if and only if

$$\beta_\gamma D_{u(p)}(t) u(p) = D_{u(\alpha \circ p)}(t) u(\alpha \circ p)$$

$$\iff \beta_\gamma \left(\frac{u(p)}{ma_t}\right)^{\frac{1}{m-1}} u(p) = \left(\frac{u(\alpha \circ p)}{ma_t}\right)^{\frac{1}{m-1}} u(\alpha \circ p)$$

$$\iff \beta_\gamma \left(\frac{u(p)}{ma_t}\right)^{\frac{1}{m-1}} u(p) = \beta_\alpha^{\frac{m}{m-1}} \left(\frac{u(p)}{ma_t}\right)^{\frac{1}{m-1}} u(p).$$

Thus, $\beta_\gamma = \beta_\alpha^{\frac{m}{m-1}}$. Since β_γ is independent of p^t, we have established X^*-Homogeneity.

E.4 Result: Uniqueness

Assume that there exist two homogeneous CE representations. By the assumption of unbounded range of $u(C)$, φ_t is automatically maximal. From Theorem 4, we have already shown that there exists $\lambda > 0$ such that $u_2 = \lambda u_1$ and $\varphi_t^2 = \lambda \varphi_t^1$. In particular, $\overline{d}_t^1 = \overline{d}_t^2 = \overline{d}_t$. Thus, $a_t^2 d^{m^2} = \lambda a_t^1 d^{m^1}$ for all $d \leq \overline{d}_t$. Note that $d^{m^1 - m^2}$ is constant and equal to $\frac{a_t^2}{\lambda a_t^1}$ for all such d, which happens only when $m^1 = m^2$. Consequently, $a_t^2 = \lambda a_t^1$, as desired.

References

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

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

[3] Andersen, S., G. Harrison, M. Lau and E. Rutström (2008): “Eliciting Risk and Time Preferences”, *Econometrica* 76(3), pp. 583-618.

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

[5] Anderson, L. and S. Stafford (2009): “Individual Decision-making Experiments with Risk and Intertemporal Choice”, *Journal Risk and Uncertainty* 38, pp. 51–72.

[6] Andreoni, J. and C. Sprenger (2012): “Estimating Time Preferences from Convex Budget Sets” *American Economic Review* 102(7), pp. 3333-3356.

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

[8] Baucells, M. and L. Zhao (2021): “Everything in Moderation: Foundations and Applications of the Satiation Model”, *Management Science*, forthcoming.

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

[10] Benartzi, S., A. Previtero and R. Thaler (2011): “Annuitization Puzzles”, *Journal of Economic Perspectives* 25(4): pp. 143–164.

[11] Benhabib, J., A. Bisin and A. Schotter (2010): “Present-bias, Quasi-hyperbolic Discounting and Fixed Costs ”, *Games and Economic Behavior* 69(2), pp. 205–223.

[12] Bordalo, P., N. Gennaioli and A. Shleifer (2012): “Salience Theory of Choice Under Risk”, *Quarterly Journal of Economics* 127 (3), pp. 1243-1285.

[13] Browning, M., and M. D. Collado (2001): "The response of expenditures to anticipated income changes: Panel date estimates," *American Economic Review* 91, pp. 681–692.

[14] Chakraborty, A. (2019): "Present Bias", mimeo.

[15] Dertwinkel-Kalt, M., H. Gerhardt, G. Riener, F. Schwerter and L. Strang (2019): "Concentration Bias in Intertemporal Choice", mimeo.

[16] 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.

[17] 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.

[18] Ericson, K., and J. Noor (2016): "Delay Functions as the Foundations of Time Preference: Testing for Separable Discounted Utility ", mimeo.

[19] 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.

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

[21] Galor, O., and O. Özak (2016): "The Agricultural Origins of Time Preference," *American Economic Review* 106(10), pp. 3064–3103.

[22] Galperti, S., and B. Strulovici (2017): "A Theory of Intergenerational Altruism," *Econometrica* 85, pp. 1175–1218.

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

[24] 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.

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

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

[27] Keren, G. and P. Roelofsma (1995): "Immediacy and Certainty in Intertemporal Choice", *Organizational Behavior and Human Decision Processes* 63(3), pp. 287-297.

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

[29] Koszegi, B. and A. Szeidl (2013): "A Model of Focusing in Economic Choice", *Quarterly Journal of Economics* 128(1), pp 53–104.

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

[31] Loewenstein, G. and D. Prelec (1992): "Anomalies in Intertemporal Choice: Evidence and an Interpretation", *Quarterly Journal of Economics* 57(2), pp. 573–598.

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

[33] Noor, J. (2009): "Decreasing Impatience and the Magnitude Effect Jointly Contradict Exponential Discounting", *Journal of Economic Theory* 144(2), pp. 869–875.

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

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

[36] Noor, J. and N. Takeoka (2021): "Supplementary Appendix to "Optimal Discounting"", Working paper.

[37] Noor, J. and N. Takeoka (2021): "Optimal Discounting with Negative Consumption Utility", Working paper.

[38] Obstfeld, M. (1990), "Intertemporal Dependence, Impatience, and Dynamics," *Journal of Monetary Economics* 26, pp. 45–75.

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

[40] Öncüler, A. (2000): "Intertemporal Choice under Uncertainty: A Behavioral Perspective," INSEAD Working Paper.

[41] 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.

[42] Parfit, D. (1971): "Personal Identity", *The Philosophical Review* 80(1), pp. 3-27.

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

[44] Saez-Marti, M. and J. Weibull (2005): “Discounting and Altruism to Future Decision-makers", *Journal of Economic Theory* 122(2), pp. 254-266.

[45] Scholnick, B. (2013): “Consumption smoothing after the final mortgage payment: Testing the magnitude hypothesis,” *Review of Economics and Statistics* 95, pp. 1444–1449.

[46] Shiv, B. and A. Fedorikhin (1999): “Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making", *Journal of Consumer Research* 26, pp. 278-292.

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

[48] Sun, Y. and S. Li (2010): “The Effect of Risk on Intertemporal Choice,” *Journal of Risk Research* 13, pp. 805–820.

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

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

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

[52] Yaari, M. (1965): “Uncertain Lifetime, Life Insurance, and the Theory of the Consumer”, *Review of Economic Studies* 32(2): pp. 137–150.

[53] Yaari, M. (1977): “A Note on Separability and Quasiconcavity,” *Econometrica* 45, pp. 1183–1186.