Delay Functions as the Foundation of Time Preference: Testing for Separable Discounted Utility*

Keith Marzilli Ericson and Jawwad Noor

April 28, 2015

Abstract

Delay functions, which vary timing of rewards but fix the money dimension, can elicit the form of discount functions with minimal assumptions. We provide conditions to test for separable discounted utility (SDU). We elicit individual delay functions for a range of amounts and time horizons. When we impose SDU assumptions, we classify more than half our analysis sample as exponential discounters. However, we reject SDU assumptions for 68% of the sample in favor of magnitude-dependent discounting with time distortion. The finding has a significant implication for how experimental results are interpreted, since the SDU assumption is necessary for small-stakes behavior in experiments to be informative about large-stakes behavior in the market.

1. Introduction

This paper uses a behavioral object we refer to as "delay functions" as an alternative method for studying time preference. While there is a major empirical debate over whether individuals are exponential or hyperbolic discounters,[1] both exponential and hyperbolic models share the common assumption that the agent's preferences can be represented by a *Separable Discounted Utility* (SDU) function,

$$U(m,t) = D(t) \cdot u(m).$$

Here, separability refers to the fact that the evaluation of time and money are unrelated. While SDU is attractive as a parsimonious model for theoretical analysis, it is a substantive assumption for experimental conclusions: separability implies that the level of discounting shown in small stakes behavior (as often used in experiments) is informative about more important decisions, such as saving for retirement. Our delay functions approach can test the validity of the SDU assumption as well as provide information about the structure of discounting even if SDU assumptions are violated.

We conduct an experiment measuring delay functions on a sample of the American population and find that the SDU assumption is rejected for 68% of our analysis sample.[2] The take-away lessons from this result are that (i) non-separable models may be important for understanding economic behavior, and (ii) experimental results may not be directly valid for real-world decisions.

Most empirical discounting studies work with "present value" data in which the dates of the rewards are fixed and an amount l is elicited that makes the subject indifferent between s at time t_1 and l at time t_2. Researchers then typically assume an SDU model and attempt to estimate D and u. Delay functions instead ask, what delay $\Phi_l(s,t)$ makes the subject indifferent between s at time t and l at time $\Phi_l(s,t)$? Delay functions thus fix the money dimension for both earlier and later rewards, and vary only the time dimension for the later reward. Specifically, for a fixed small reward s and large reward l, a delay function elicits for each t the delay from time zero $\Phi_l(s,t)$ such that

$$(s,t) \sim (l, \Phi_l(s,t)). \tag{1.1}$$

---

[1] See e.g.Andreoni and Sprenger [2012], Augenblick, Niederle, and Sprenger [2013].

[2] This result is consistent with recent research of Ericson et al. [forthcoming], showing that a non-SDU heuristic model outperforms all common SDU models in a cross-validated prediction exercise. Their heuristic model, however, is only defined for binary choices, not multiple price list or multiple delay list elicitation methods.

That is, the subject reveals that s at time t is as good as l at time $\Phi_l(s,t)$. This *delay function* Φ reveals that the 'loss of attractiveness' (discounting) of s due to a delay over any interval $[t, t']$ is equal to that of l over $[\Phi_l(s,t), \Phi_l(s,t')]$, and thus it is an embodiment of the agent's discount function. Delay functions can be obtained in an experimental setting by using a 'multiple delay list' rather than a 'multiple price list' procedure, or alternatively by the Becker-DeGroot-Marschak mechanism.

Delay functions allow researchers to estimate general discounting models that allow subjects to discount different size rewards differently ("magnitude-dependent discounting"). Existing evidence suggests that this may be the case: experiments often find a *magnitude effect* in choice, with subjects seeming to exhibit more patience towards larger rewards (Frederick et al [2002], Noor [2011], Andersen et al. [2013]).[3]

The main theoretical result in this paper delivers an explicit formula for computing the subject's discount function on the basis of the data Φ. This is done for preferences with very little structure: complete, transitive, continuous preferences that satisfy monotonicity and impatience, admitting *General Discounted Utility* representations of the form

$$U(m,t) = D(m,t) \cdot u(m),$$

where $D(\cdot, 0) = 1$. We provide the expression for D in terms of Φ. The general theorem is then specialized for the purpose of applications. In particular, we are interested in tests for the existence of SDU representation. While SDU is an attractive theoretical assumption (parsimony, appealing axioms), its *descriptive validity* is of interest here. Thus, it is valuable to have a simple test and let the data speak on whether separability adequately describes decision-makers' behavior.

Our main specialization restricts attention to a flexible class of delay functions:

$$\Phi(m,t) = \left(a(m) t^\gamma + b(m) \right)^{1/\gamma}.$$

This subsumes the behavior of common SDU models (exponential, generalized hyperbolic), where delay functions are linear and of the specific form $\Phi(m,t) = [1 + kb(m)] t + b(m)$, with $k \geq 0$ and b decreasing in m. The flexible functional form permits non-linearity in the delay function. For this class of delay functions, the test for SDU discounting is simply that there is some $k \geq 0$ such that $a(m) = 1 + kb(m)^\gamma$ for all m.

We prove that subjects who fail the SDU test can nevertheless be attributed the magnitude-dependent exponential discounting with time-distortion: $D(m,t) =$

---

[3]Whether this observed fact rejects SDU is controversial, because experiments must typically estimate both $D(t)$ and the curvature of $u(m)$. The delay function approach does not require us to estimate u.

$e^{-a(m)\cdot t^{\gamma}}$ where a is decreasing. In this model, the subject is more patient towards larger rewards, as suggested by the magnitude effect. Magnitude-dependent discounting could arise from a decision process where the agent has limited cognitive resources and finds it optimal to pay more attention only to higher rewards. It may also relate to models of mental accounting in which larger rewards are assigned to different accounts, or models of self-control and temptation preferences (e.g. Gul and Pesendorfer [2001], Fudenberg and Levine [2006]) in which the degree of self-control implemented may vary with the magnitude of reward at stake. Time distortions are motivated by a literature suggesting non-linear perception of time (e.g. Zauberman et al. [2009]).

We elicit delay functions in an experiment on a representative sample of the American population recruited from a professional sampling service. We conduct all our estimation at the individual level, allowing for heterogeneity across individuals in preferences and models. We first assume that delay functions are linear ($\gamma = 1$ in the above class) and that the agents are SDU ($a(m) = 1 + kb(m)$ for all m). Within this subclass, the individual is exponential if $k = 0$ and hyperbolic if $k > 0$. We find that the median k in our analysis sample is virtually zero, suggesting that more than half of our sample were exponential discounters if we limited consideration to SDU models with linear delay functions only.[4] However, we then directly test the SDU assumption. For 68% of our analysis sample, we reject SDU (with or without time distortion) at $p < 0.05$, finding strong evidence of non-SDU discounting. Using model selection criteria, only 18% of our sample has a best-fit model that is consistent with SDU. These results suggest that if researchers only examine SDU models, they may mistakenly conclude that individuals are exponential discounters.

The rejection of separability has a significant implication for the extrapolation of experimental results. The only way experimental results involving small stakes can be used to make assertions about choices involving large stakes is precisely by assuming separability. Without separability, a theory is needed to understand the validity of such assertions. The results also reveal that there is a possibly significant benefit to researching economic explanations outside the class of SDU models. A worthwhile avenue for future research is to model magnitude-dependent discounting of the type revealed in our experiment and to study its implications in economic settings.

Our theoretical analysis applies to any type of dated reward (money, food, work, etc.). Our experiment uses money, following most of the literature. However, there is

---

[4]Using a related elicitation method, Attema et al [2010] restrict consideration to SDU models and find that 58% of classifiable subjects were exponential discounters. They do not, however, test the SDU assumption.

a dispute over whether money from experiments is consumed when received. While many models assume that individuals will smooth their consumption across time, models of mental accounting or dual-self models (e.g. Fudenberg and Levine [2006]) predict that income from experiments may be consumed when received. Andersen et al. [2008] have data on risk questions and time-money tradeoff questions, and jointly estimate the curvature of utility, discount function, and degree of consumption smoothing. They estimate that payments are consumed when received; see also Booij and van Praag [2009]. However, in recent work, Augenblick, Niederle, and Sprenger [2013] find substantial differences in the degree of preference reversals for money versus real effort tasks, suggesting choices over money may not reveal the discount function. However, if individuals do smooth consumption over time, taking advantage of outside-the-lab borrowing and lending, then subjects' choices should reveal the interest rate they face (Cubitt and Read [2007]). If the interest rate faced is constant in the range of dollar amounts considered (in our experiment, \$50 to \$100), then their choices should appear as though they were SDU discounters. We in fact reject SDU for most subjects. This rejects, at the very least, exponential discounting of income (as opposed to consumption).

The remainder of the paper proceeds as follows. We close the introduction with related literature. Section 2 presents the main theoretical results and Section 3 presents specializations. Section 4 presents our experiment and results and Section 5 concludes. All proofs are contained in appendices.

Related literature

We term the common method of eliciting discount functions as the "present value" approach: subjects choice between dated rewards (m, t) and give indifference points of the form $(s, t) \sim (l, t')$, where, for any dates t, t' either the future reward l is fixed and the subjects' "present value" s is obtained, or the present reward s is fixed and the "future value" l is obtained.[5] In recording how present/future value of a reward changes with t, the data reflects behavior when both the money and time dimensions are changed. This is reflected in how conclusions are drawn about the discount function. Presuming the SDU model, the discount function $D(t)$ is elicited by computing

$$D(t) = \frac{u(s)}{u(l)},$$

and so eliciting the discount function requires an assumption on u. The early literature assumed that u is linear. Since this assumption typically yields implausibly high discount rates, the literature has sought methods of eliciting discount functions

---

[5]See Fredrick et al [2002] for a review of the experimental literature, and later experimental work by Coller and Williams [1999] and Harrison et al [2002].

5

and the curvature of u simultaneously. Andersen et al [2008] replace the linearity assumption with the expected utility assumption, and they use both risk preferences and time preferences to jointly estimate several specifications of u and D. Andreoni and Sprenger [2012] replace the linearity assumption with the assumption that preferences over consumption streams are represented by a time-additive SDU model $\sum D(t)u(m_t)$ with CRRA u. Subjects are asked to choose their allocation of an endowment over two periods for different interest rates and endowments, and thus their intertemporal demand curves are obtained, to which u and D are jointly fit.

The theoretical literatures on multi-attribute utility and conjoint measurement (Fishburn [1967], Krantz et al [1971]) introduce the "sawtooth method" to behaviorally isolate the components of any separable representation, and this is built on the idea of varying only one dimension while fixing others.[6] Attema et al's [2010] study of discount functions and Wakker and Deneffe's [1996] study of probability weighting are experimental implementations of the sawtooth method. The delay function approach presented in this paper differs from the sawtooth method: It is established for preferences that do not necessarily lie in the SDU class. Moreover, the theoretical derivation of the discount function is very different – we find a solution to a functional equation rather than doing a direct construction with the sawtooth method. Finally, we avoid incentive compatibility issues that exist in the experimental application of the sawtooth method, as discussed in Harrison and Ruström [2009].[7]

Relatedly, Laury, McInnes, and Swarthout [2012] also show a procedure for eliciting discount rates using variation in the probability a payment will be made, rather than the amount (as in the present value approach) or the delay (as in our proposed approach). Their approach also assumes SDU, as well as either expected utility theory or a particular probability weighting function. Similarly, Olea and Strzalecki (2014) provide a method using "annuity compensations" to estimate quasi-hyperbolic discounting without estimating the utility function u. Their approach assumes a

---

[6]These papers consider a preference over binary attributes (x, y). Fix any y, y' and x_0 and suppose that x_1 is a quantity such that the agent exhibits $(x_1, y) \sim (x_0, y')$. Furthermore, suppose it is determined that, iteratively for $i = 2, .., n$, that $(x_i, y) \sim (x_{i-1}, y')$. If the preference has a multiplicative representation, $U(x, y) = v(x) \cdot u(y)$, then each indifference point satisfies $\frac{v(x_i)}{v(x_{i-1})} = k$ for all $i = 1, .., n$, for some constant $k := \frac{u(y')}{u(y)}$. Since v is unique up to an affine transformation, $v(x_0)$ and $v(x_1)$ can be normalized, and consequently v is pinned down on $\{x_0, .., x_n\}$. The grid can be made arbitrarily finer. This procedure is referred to as the 'saw-tooth method' (Fishburn [1967]) and the noted sequence is an example of a 'standard sequence' (Krantz et al [1971]).

[7]In the sawtooth method, a subject's answer to one question becomes an input into the next question. Therefore by misstating preferences it is possible for subjects to affect the sequence of questions they face in a way that improves their expect outcome.

quasi- or semi-hyperbolic discounting model– members of the SDU class– while our approach can test the underlying SDU assumption.

2. Theoretical Foundations

We present here the main theoretical result that provides the foundations for our experimental procedure.

The primitive of our analysis is a revealed preference relation $\succsim$ over the set of dated rewards $X = \mathcal{M} \times \mathcal{T}$, where time is continuous and given by $\mathcal{T} = \mathbb{R}_+$, with generic elements t, t', and the set of rewards (e.g. money) is a bounded interval $\mathcal{M} = [0, \overline{m}]$ with generic elements m, m', s, l. Data on such a preference is the minimal and simplest data that may be used to study the basic structure of time preference in a general way.

The choice data needed for our analysis is the *delay function* $\Phi : \mathcal{M} \times \mathcal{T} \to \mathcal{T}$, which is obtained via the indifference:

$$(m, t) \sim (\overline{m}, \Phi(m, t)),$$

for all $0 < m \leq \overline{m}$ and each t. That is, $\Phi(m, t)$ is defined as the date such that m at t is just as good as $\overline{m}$ at $\Phi(m, t)$.[8] Varying t leads to a variation in the desirability of (m, t), and this is measured by variation in the delay $\Phi(m, \cdot)$. The simplest example is a linear delay function, $\Phi(m, t) = a(m)t + b(m)$. The delay function can be obtained in practice by using the Becker-DeGroot-Marschak mechanism or by adapting the Multiple Price List (MPL) popularized by Coller and Williams [1999] and Harrison et al [2002].[9]

Our general result presumes only that the preference $\succsim$ admits a *General Discounted Utility* (GDU) representation:

$$U(m, t) = D(m, t) \cdot u(m),$$

---

[8] Compared to the notation $\Phi_l(s, \cdot)$ in the Introduction for any pair of rewards $s \leq l$, here we fix the largest reward l at $\overline{m}$, and suppress it in the notation.

[9] An MPL asks questions of the form "Do you prefer \$100 now or \$$x$ in 6 months?" where x varies over a grid $x_1, ..., x_{N+1}$ of dollar amounts. The implied interest rate associated with x increases monotonically moving down the list, and the point at which the subject switches from preferring the earlier reward to the later reward determines an interval $[x_i, x_{i+1}]$ within which an indifference point '(\$100,now) $\sim$ (\$$z$,6 months)' lies. A 'Multiple Delay List' asks a sequence of questions of the form "Do you prefer \$50 in 1 month or \$100 in t months?" where t varies over a range of time periods $t_1, .., t_{N+1}$ in a way that the implied interest rate decreases monotonically moving down the list.

where $u : \mathcal{M} \to \mathbb{R}_+$ is a *utility index* (a strictly increasing, continuous function satisfying $u(0) = 0$) and $D : \mathcal{M} \times \mathcal{T} \to (0, 1)$ is a *discount function* (a continuous, strictly decreasing function satisfying $D(m, 0) = 1$ and $\lim_{t \to \infty} D(m, t) = 0$ for all $m > 0$) such that $D(m, t)u(m)$ is strictly increasing in m. We will often refer to the tuple (D, u) as the GDU representation. If a preference admits a GDU representation then we say it is *regular*. Lemma A.1 confirms that regularity corresponds only to very basic properties such completeness, transitivity, monotonicity and impatience.[10]

Any increasing transformation of a GDU representation yields another GDU representation, so there are an infinity of such representations for a given preference.[11] In order to demonstrate the generality of our approach to studying discount functions, we proceed without requiring more structure on the preference or restricting attention to any particular subclass of representations. Later we specialize to prepare for applications.

The exhaustive set of restrictions on the delay function implied by the regularity assumption is given in the following proposition.

Proposition 2.1. Φ *is generated by a regular preference* $\succsim$ *if and only if:*
    (i) $\Phi(m, t)$ *is continuous.*
    (ii) For any t, $\Phi(\cdot, t)$ *is strictly decreasing and* $\lim_{m \to 0} \Phi(m, t) = \infty$.
    (iii) For $m > 0$, $\Phi(m, \cdot)$ *is strictly increasing and* $\Phi(\overline{m}, t) = t$ *for all* t.

Part (ii) reflects monotonicity with respect to money: the delay function increases monotonically to infinity as m decreases to 0. Intuitively, as the sooner-smaller reward m decreases and becomes less desirable, the large reward has to be pushed into the future in order to maintain indifference. In a similar fashion, part (iii) reflects monotonicity with respect to time.

2.1. General Framework

The main result in this paper identifies the *set* of discount functions attributable to the preference $\succsim$, and the utility index u that corresponds to each discount function.

---

[10] The intuitive appeal underlying the connection between delay functions and discount functions does not rely on transitivity, so we expect that our results can be extended to models of intransitive preference, such as Ok and Masatlioglu [2007]. We maintain regularity since it enables a clear exposition.

[11] To see this, first note that for any utility index u there exists a unique representation U for a regular preference $\succsim$ (for this standard result see for instance Fishburn and Rubinstein [1982]). Our assertion then follows from the fact that any representation U can be uniquely written in the form of a GDU representation (D, u) as follows: for any representation U, the utility index u in any GDU functional form is uniquely defined by $u(m) = U(m, 0)$, and D is uniquely defined by $D(m, t) = \frac{U(m,t)}{u(m)}$ for all $m > 0$.

8

Say that a function $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a *restricted transformation* if it is continuous, strictly increasing, unbounded and satisfies $g(0) = 0$.

Theorem 2.2. *Consider a regular preference $\succsim$ and its delay function Φ. Then $\succsim$ admits the GDU representation (D, u) if and only if there is a restricted transformation g and some scalar $u(\overline{m}) > 0$ such that for all $m > 0$ and t,*

$$D(m, t) = e^{-[g(\Phi(m,t)) - g(\Phi(m,0))]},$$

and for all $m \geq 0$,

$$u(m) = e^{-g(\Phi(m,0))} \cdot u(\overline{m}).$$

The result characterizes all the discount functions and corresponding utility indices that can be attributed to the preference $\succsim$. The functional forms involve an increasing transformation g of Φ. Discount functions are defined in terms of the difference $g(\Phi(m, t)) - g(\Phi(m, 0))$, whereas utility indices are defined in terms of $g(\Phi(m, 0))$. Observe that the discount function is completely characterized in terms of the delay function Φ. While u is also characterized in terms of Φ, it essentially only reflects the information contained in present values: by definition, $(m, 0) \sim (\overline{m}, \Phi(m, 0))$. In contrast D requires information on how Φ changes as a function of t. The result reveals that obtaining a functional form for Φ is all that is necessary to obtain all the discount functions attributable to the subject.

The simple idea behind the proof is as follows. Note that the two indifference points

$$(s, 0) \sim (\overline{m}, \Phi(s, 0)) \text{ and } (s, t) \sim (\overline{m}, \Phi(s, t))$$

reveal that the loss of attractiveness (due to discounting) in (s, t) relative to $(s, 0)$ must equal the loss in $(\overline{m}, \Phi(s, t))$ relative to $(\overline{m}, \Phi(s, 0))$. This translates into the statement that any discount function D attributable to the preference $\succsim$ must satisfy the equality $\frac{D(s,t)}{D(s,0)} = \frac{D(\overline{m}, \Phi(s,t))}{D(\overline{m}, \Phi(s,0))}$. By definition, $D(s, 0) = 1$, and so this inequality can be rewritten as:

$$D(s, t) \cdot D(\overline{m}, \Phi(s, 0)) = D(\overline{m}, \Phi(s, t)).$$

But this is a functional equation where D is the unknown function and Φ is the known function. The proof verifies that a discount function D is a solution to this functional equation if and only if there exists a utility index u for which (D, u) is a GDU representation for the preference $\succsim$. The general solution of the functional equation is leads to the statement of the theorem. (In the appendix we prove a more general result by allowing $\mathcal{M}$ to be unbounded).

2.2. The Efficiency of Delay Functions

Theorem 2.2 not withstanding, the set of compatible discount functions can be computed through present value data as well: it is readily seen that (D, u) is a GDU representation for a regular preference $\succsim$ if and only if u is a utility index and the discount function satisfies

$$D(m, t) = \frac{u(p(m, t))}{u(m)},$$

where $p(m, t)$ is the present value of (m, t). The proof of this is trivial, and it provides a characterization of the set of discount functions and corresponding utility indices just as Theorem 2.2 does.

However, in practical settings, where data is necessarily limited, Φ-data better reveals the information in $\succsim$ than present value data tradeoff data. We substantiate this claim here in the context of the SDU model. Fix the set of periods and prizes and order them so that $0 = t_1 < t_2 < ... < t_J$ and $0 < m_1 < ... < m_I$. Write the corresponding finite space of dated rewards as $X_{IJ} := \{m_1, .., m_I\} \times \{0, t_2, .., t_J\}$. Suppose that the present value data is given by p_{ij} such that

$$(p_{ij}, 0) \sim (m_i, t_j) \text{ for all } m_i \text{ and all } t_j > 0.$$

Assume that the analyst is interested in SDU representations. Say that the (magnitude-independent) discount function D is *attributable* to the present value data if there exists a utility index u such that $u(p_{ij}) = D(t_j)u(m_i)$ for all $(m_i, t_j) \in X_{IJ}$. Let the set of attributable D be denoted by $\mathcal{D}_I^p$. This is indexed by I since we will be varying I below.

Given X_{IJ}, let $p^*(= p_{I1})$ denote the present value of (m_I, t_1), the largest reward at the earliest future period. Suppose that Φ-data is obtained by determining τ_j such that

$$(p^*, \tau_j) \sim (m_I, t_j) \text{ for all } t_j > 0.$$

That is, we determine τ_j such that $\Phi(p^*, \tau_j) = t_j$. Say that the discount function D is attributable to the Φ-data if $D(t)D(\Phi(p^*, t_1)) = D(\Phi(p^*, t))$ for all these periods $\tau_1, .., \tau_J$.[12] Denote the set of attributable D by $\mathcal{D}^\Phi$.

The present value and Φ-data are related by a common time horizon t_J and also the indifference point $(p_{I1}, 0) \sim (m_I, t_1)$ which defines both p_{I1} and $\Phi(p_{I1}, 0)$. The following theorem reveals that the $J - 1$ data points for Φ are more discerning than the $I \cdot (J - 1)$ data points for present values, *regardless of the number I of rewards.*

---

[12]This is equivalent to requiring that there are utilities $0 < u(p^*) < u(m_I)$ such that $D(t)u(p^*) = D(\Phi(p^*, t))u(m_I)$ for these t's. Observe that $u(p^*) = D(\Phi(p^*, t_1))u(m_I)$ must hold and so the utilities can be substituted out, yielding the original definition.

Theorem 2.3. *Suppose that $\succsim$ admits some SDU representation. Then for all I,*

$$\mathcal{D}^{\Phi} \subset \mathcal{D}_I^p.$$

The proof of the theorem is based on the following insight: limited present value data will at best put bounds on the participant's true delay function Φ and this is the *only* extent to which it restricts the range of possible D's. The remainder of the data, no matter how rich, will only help determine what u goes with any such D (observe that in Theorem 2.2 the utility index is determined by $\Phi(\cdot, 0)$, which essentially comes from money-time trade-off data). Limited direct data on Φ will speak more than data that just puts bounds on Φ.

The 'true' D is in both $\mathcal{D}^{\Phi}$ and $\mathcal{D}_I^p$. The theorem therefore tells us that there is greater efficiency achieved by using Φ, in that we can get closer to the true D with fewer data points. Stated differently, the theorem reveals that the degree of potential misidentification is greater with present value data than it is with Φ-data. In this sense, limited Φ data provides a better picture of the agent's entire preference than does present value, which is the claim we set out to establish. This theorem thus provides further validation for our claim that Φ serves as a behavioral definition of discount functions.

3. Specializations[13]

We proved the result for a class as general as the GDU class so as to enable a wide variety of applications. Practical interest will typically lie in subclasses with much more structure, such as the SDU class or its generalizations. We consider a class smaller than GDU that subsumes SDU. In the sequel we further specialize within this class.

The literature discusses two dimensions along which discounting can deviate from exponential discounting. The first is time distortion (non-linear perception of time, as in Zauberman et al. [2009]). The second is magnitude-dependence (greater patience towards larger rewards, as in the magnitude effect discussed in Frederick et al [2002]); this has not received much attention in theoretical literature. A functional form that allows for both is:

$$D(m, t) = e^{-a(m) \cdot g(t)}. \tag{3.1}$$

for some $a : \mathcal{M} \to \mathbb{R}_+$ and $g : \mathbb{R}_+ \to \mathbb{R}_+$. Note that the SDU model obtains if $a(\cdot) = 1$.

---

[13]This subsection can be skipped without loss of continuity.

The proposition highlights some of the *ordinal* content of the representation.

Proposition 3.1. *Consider a regular preference that admits a discount function $D(m,t) = e^{-a(m) \cdot g(t)}$. Then:*

(a) $a(\cdot)$ is a continuous decreasing function satisfying $a(\cdot) > 0$ and $a(\overline{m}) = 1$.

(b) $g(\cdot)$ is continuous, strictly increasing, unbounded and satisfies $g(0) = 0$. A twice differentiable g is linear if and only if the delay function Φ is linear in t.

The proposition reveals the following. First, given regularity, a must be decreasing function of m. Intuitively, if a larger reward is discounted more heavily, then the agent may prefer \$50 over \$100 at some large enough t, violating monotonicity. Therefore the ordinal restriction of regularity is consistent with only one kind of magnitude-dependence, namely that of greater patience towards larger rewards. Second, the existence of time distortion is an ordinal property: it represents nonlinearity in the delay function in t.

To restrict attention to this class is to implicitly make an assumption on the agent's delay function. For completeness we clarify the technical restriction:[14]

Proposition 3.2. *Consider a regular preference $\succsim$ and its delay function Φ, and take any a, g that satisfy the properties in Proposition 3.1. Then preference $\succsim$ admits a GDU representation with a discount function of the form $D(m,t) = e^{-a(m) \cdot g(t)}$ if and only if Φ satisfies,*

$$g(\Phi(m,t)) - g(\Phi(m,0)) = a(m) \cdot g(t), \tag{3.2}$$

for all m, t.

The restriction on the delay function is not a transparent one, but what matters for our purposes is that it is lax enough to allow for a variety of functional forms some of which, as we will see below are adequate for a general analysis. Some examples are presented in the table.

---

[14]The proof of Proposition 3.2 notes that a discount function $D(m,t) = e^{-[g(\Phi(m,t))-g(\Phi(m,0))]}$ in our main theorem takes the desired form if and only if the power term $g(\Phi(m,t)) - g(\Phi(m,0))$ equals $a(m)f(t)$ for some function f. Observing that for $m = \overline{m}$ the power term becomes $g(t) - 0$ it follows that $f = g$. Hence the condition (3.2).

For the proof for Proposition 3.1, write $\delta(m) = e^{-a(m)}$, suppose $s < l$ and observe that Monotonicity (ie, for any t, larger rewards are preferred) requires that $\delta(l)^{f(t)} \cdot u(l) > \delta(s)^{f(t)} \cdot u(s)$ and thus $\frac{u(l)}{u(s)} > \left(\frac{\delta(s)}{\delta(l)}\right)^{f(t)}$ for all t. If $\frac{\delta(s)}{\delta(l)} > 1$, the unboundedness of f implies that this inequality cannot hold for all t, a contradiction. Thus $\delta(s) \leq \delta(l)$, and $a(\cdot)$ must be decreasing. The remaining properties follow readily from regularity.

Φ-Function	Discount Function D	Generated by transformation:
$\Phi(m,t) = g^{-1}(g(t) + g(\Phi(m,0)))$	$D(t) = e^{-rg(t)}$	$a(m) = 1$, any g
$\Phi(m,t) = (1 + \alpha\Phi(m,0))t + \Phi(m,0)$	$D(t) = (1 + \alpha t)^{-1}$	$a(m) = 1, \ g(t) = \ln(1 + \alpha t)$
$\Phi(m,t) = a(m)t + \Phi(m,0)$	$D(m,t) = e^{-ra(m)\cdot t}$	$g(t) = t$
$\Phi(m,t) = [a(m)\cdot t^\alpha + \Phi(m,0)^\alpha]^{\frac{1}{\alpha}}$	$D(m,t) = e^{-ra(m)\cdot t^\alpha}$	$g(t) = t^\alpha$
$\Phi(m,t) = \frac{[(1+\alpha t)^{a(s)}(1+\alpha\Phi(m,0))-1]}{\alpha}$	$D(m,t) = (1 + \alpha t)^{-\varphi(m)}$	$g(t) = \ln(1 + \alpha t)$

Table 1: Φ-functions and associated D.

Unless the analyst uses richer data in order to pin down the representation, or further restricts the representation (for instance we take $g(t) = t^\gamma$ in our application), there may be more than one pair of a and g satisfying (3.2) for a given Φ, and for each such pair a different representation for the preference may be obtained. While such non-uniqueness of a representation is a nuisance of sorts in theoretical work, in the current context it is a strength: different representations may provide different intuitive explanations for a given behavior, which may be relevant for interpretation and applications. The usefulness of such nonuniqueness will be demonstrated in the sequel, where we showed that behavior arising from hyperbolic discounting can be replicated by magnitude-dependent discounting.

3.1. A Tractable Class of Delay Functions

To prepare for our experimental application of Theorem 2.2, we introduce a class of delay functions that lies within the restricted framework of the previous subsection. This is a functional form that accommodate familiar behaviors like Stationarity and Preference Reversals but is flexible enough to accommodate nonlinearities in the data, and more besides:

$$\Phi(m,t) = (a(m)t^\gamma + b(m)^\gamma)^{\frac{1}{\gamma}}, \tag{3.3}$$

for $\gamma > 0$. Given Proposition 3.1, regularity requires that $a(\cdot)$ is weakly decreasing and satisfies $a(\overline{m}) = 1$, and $b(\cdot)$ is strictly decreasing and satisfies $b(\overline{m}) = 0$.[15] Thus, the curves $\Phi(m,\cdot)$ are upward sloping, non-intersecting, and the curves for lower m lie strictly above those for higher m. Time distortion (curvature of the delay function

---

[15]To see that $a(\cdot)$ must be weakly decreasing, suppose $m > m'$ and $a(m) > a(m')$ then for large t we would obtain $\Phi(m,t) > \Phi(m',t)$, a contradiction.

in t) is captured by γ. The delay function is linear (resp concave, convex) if $\gamma = 1$ (resp $\gamma > 1$, $\gamma < 1$).[16]

We first identify conditions under which there exists an SDU representation. Refer to magnitude-*in*dependent discount functions $D(t)$ as *separable* discount functions.

Proposition 3.3. *Consider a regular preference $\succsim$ and with a delay function Φ that has the nonlinear form (3.3). A separable discount function $D(t)$ can be attributed if and only if there exists $k \geq 0$ such that, for all m,*

$$a(m) = 1 + kb(m)^\gamma. \tag{3.4}$$

If $k = 0$, then the only attributable separable discount function is exponential discounting with time distortion,

$$D(t) = e^{-rt^\gamma}, \qquad r > 0.$$

If $k > 0$, then the only attributable separable discount function is hyperbolic discounting with time distortion,

$$D(t) = (1 + kt^\gamma)^{-r}, \qquad r > 0.$$

Thus, the test for the existence of an SDU representation is simply that the slopes $a(m)$ must be a linear function of $b(m)^\gamma$ in (3.3). The slope $k \geq 0$ of this function determines the shape of the separable discount function, which can either be exponential or hyperbolic, but with time distortion. If the delay function is linear ($\gamma = 1$) then there is no time distortion and we obtain the corresponding standard models.

If the SDU test fails we can nevertheless still identify the non-SDU representation that exists:[17]

Proposition 3.4. *The delay function Φ has the form (3.3) if and only if the following general exponential discount function can be attributed to it:*

$$D(m, t) = e^{-ra(m) \cdot t^\gamma}, \qquad r > 0$$

where $a(\cdot)$ is weakly decreasing in the size of the reward.

---

[16]The expression for the second derivative of $\Phi(m, t)$ wrt to t is $(1 - \gamma)(a(m)t^\gamma + b(m)^\gamma)^{\frac{1-2\gamma}{\gamma}} t^{2(\gamma-1)} [1 - (\frac{\Phi(m,t)}{t})^\gamma]$, and the term in the square brackets is always negative since $\frac{\Phi(m,t)}{t} > 1$.

[17]To prove this, simply apply Prop 3.2 with the transformation function $g(r) = r^\gamma$.

As before, the direction of the magnitude dependence in $a(\cdot)$ is an ordinal restriction, and strongly reminiscent of the magnitude effect discussed in the literature (Frederick et al [2002]). The free parameter r needs additional data to pin down, but for our purposes below it will suffice to restrict attention to the canonical representation obtained by setting $r = 1$. An attractive feature of this canonical representation is that all its parameters are fixed by Φ.

A noteworthy observation is that hyperbolic discounting is *behaviorally a special case* of the general exponential discount function on the domain of dated rewards: all subjects with linear Φ are representable by general exponential discounting but only a subset are representable by hyperbolic discounting. Therefore the analysis reveals that magnitude-dependent discounting is an alternative explanation for preference reversals attributed to hyperbolic discounting. The two forms of discounting are substantially different in spirit. Hyperbolic discounting is suggestive of a self-control problem, whereas the magnitude effect is suggestive of bounded rationality: the former suggests a passion for the present [1997] whereas the latter suggests that subjects pay greater attention to larger rewards [2011].

4. Empirical Application

4.1. Experiment Design

For an empirical application of the method, we recruited 100 participants aged 18-65 from an online sampling service (Qualtrics) designed to produce an approximately representative sample based on age, gender, and income.[18] Prior to collecting this data, we ran a pilot using participants recruited from an online labor market (Amazon MTurk); results are quite similar.[19]

Participants were paid a flat participation fee and received incentives to respond truthfully, as 10% of participants were randomly selected to be paid for one of their choices. This level of incentives is in line with previous work, e.g. Andersen et al.

---

[18]The sampling service screens out inattentive participants who fail basic attention checks, such as answering questions too quickly or answering illogically or inconsistently. 11 such participants were screened out. Additionally, 20 subjects began the experiment but did not complete it. These participants do not count toward our 100 completed participants.

[19]The Mturk pilot was unincentivized. The procedure was similar, but differed in the set of smaller, sooner magnitudes considered: $m \in \{\$25, \$50, \$75, \$90\}$. The pilot analysis sample (dropping non-monotonic choices and subjects who ever have a list on which they always take the larger-later option) was 44 of 118 participants; (many subjects always chose \$100 over \$25). In pilot analysis sample, we find more evidence of decreasing impatience. We get similar results testing model restrictions.

2008 pay 10% of their sample.

Participants faced a series of multiple delay list decisions. On each list, they were asked to choose between (m, t)– smaller amount m at time t– and a series of options $(100, t + \Delta)$, with $\Delta \in \{1, 2, 3, 5, 7, 9, 13, 17, 22, 28, 35, 43, 68\}$. Participants saw lists for 5 smaller-sooner amounts $m \in \{\$50, \$60, \$70, \$80, \$90\}$ available at 6 different time horizons for the smaller-sooner amount $t \in \{0, 1, 3, 5, 12, 24\}$ weeks, for a total of 30 lists. As required by the theory, the larger-later amount was constant on all lists (here, \$100). Thus, on one particular list, participants choose between \$50 at time 0 versus \$100 at time 1, then \$50 at time 0 versus \$100 at time 2, etc.

Our objective is to find each participant's delay function $\Phi_i(m, t)$ s.t. $(m, t) \sim (\overline{m}, \Phi_i(m, t))$. For each participant i on each multiple delay list, we determine the indifference point from the point at which the participant switches from choosing the larger-later reward to the smaller-sooner reward. We place the indifference point at the midpoint of the interval: if the participant chooses larger-later at $t + \Delta_j$ and smaller-sooner at $t + \Delta_{j+1}$ then $\Phi_i(m, t) = \frac{1}{2}(\Delta_j + \Delta_{j+1})$. If a participant always chooses the earlier option on a list, $\Phi_i(m, t) = \frac{1}{2}(1 + 0) = \frac{1}{2}$.

So, for example, a participant who chose the larger-later option in "\$50 in 1 week v. \$100 in 2 weeks", but the smaller-sooner option in "\$50 in 1 week v. \$100 in 3 weeks", has revealed they are willing to wait 2 weeks but not 3 weeks for \$100. Hence, $\Phi_i(25, 1) = 2.5$. If a participant always chooses the larger-later option on a particular list,[20] they were too patient for us to estimate their delay function. High levels of patience do not allow us to observe the characteristics of their discount function: e.g. very patient exponential discounting looks very similar to very patient hyperbolic discounting. This issue is not unique to our method; similar issues apply to multiple price lists.

To construct our Analysis Sample, we drop the 26/100 subjects who ever make a non-monotonic choice (they do not have a unique switching point on some list). We then limit to the 40/74 who never have a multiple delay list on which they always choose the larger-later option. Of these excluded subjects, 26% make the more patient choice on every question asked, and 59% always make the more patient choice on the majority of multiple delay lists. Participants who are very patient do not allow us to observe the characteristics of their discount function. In robustness

---

[20]The maximum available delay offered was constrained for the purposes of payment reliability and feasibility. The latest available payment in this experiment was 92 weeks later, or almost 2 years. The preferences of patient participants could be better captured with lists that use higher smaller-sooner amounts (e.g. $m = 99$). However, we wanted to limit the total number of decisions made to avoid taxing participants' attention and cognitive ability. Incentive-compatible dynamic designs would be complicated to explain and implement. Pre-testing indicated that our chosen range of magnitudes and delays captured the preferences of the largest fraction of participants.

checks, we broaden our sample to include those participants who always choose the larger-later option on less than 10 m, t pairs and impute a value of the delay function for those choices; the fraction of subjects consistent with an SDU representation is quite similar.

4.2. The Data

Our Analysis Sample has broad demographic coverage and is similar to the U.S. population. It is 57% male, with a median age of 48, and 50% are married. By comparison, the 18-65 U.S. population has a median age of 43 and a marriage rate of about 50% (2011-2013 American Community Survey). Our sample has a range of income levels: 33% have household incomes below $35,000 and 30% have household incomes above $75,000. By comparison, about 35% of U.S. households have incomes below $35,000 and about 33% have incomes about 75,000. The full sample also has very similar characteristics.

For each participant i, magnitude m of the smaller-sooner payment, and time horizon t to the earlier payment, we calculate $\Delta t_{i,m,t} = \Phi_i(m, t) - t$, which is the maximum *additional* delay the participant is willing to accept while still choosing the larger-later payment. The larger $\Delta t_{i,m,t}$ is, the more patient the choice. Figure 1 plots the average $\Delta t_{i,m,t}$ across participants, broken out separately by magnitude of and time to the smaller-sooner payment. In the upper left panel, the average maximum additional delay participants are willing to take to choose $100 over $50 is 10.7 weeks when the smaller-sooner reward was available today, and 13.1 weeks when the smaller-sooner payment was available at 24 weeks. In the lower right panel, we find that on average, participants are only willing to wait 3-4 additional weeks for $100 over $90, regardless of when the $90 became available.

Constant impatience (i.e. exponential discounting) implies that the maximum additional delay $\Delta t_{i,m,t}$ should not vary by the horizon t to the smaller-sooner reward, while decreasing impatience (e.g. hyperbolic discounting) implies that $\Delta t_{i,m,t}$ should increase with t. Visually, Figure 1 shows only very limited evidence of decreasing impatience at the aggregate level– only the $50 v. $100 choice displays a noticeable increasing pattern.[21] But individual heterogeneity is important. We regress the maximum additional delay for each participant i, amount m, horizon t against participant-specific fixed effects for smaller-sooner amount and for horizon

---

[21]This is consistent with a stream of literature that has not found aggregate decreasing impatience for monetary rewards, including McClure et al. [2004], who use the same payment method as we do, and Andreoni and Sprenger [2012], who take care to remove many confounds in their estimation of preferences. Other work, such as Olea and Strzalecki (2014), has found evidence of both decreasing impatience and increasing impatience for monetary rewards.
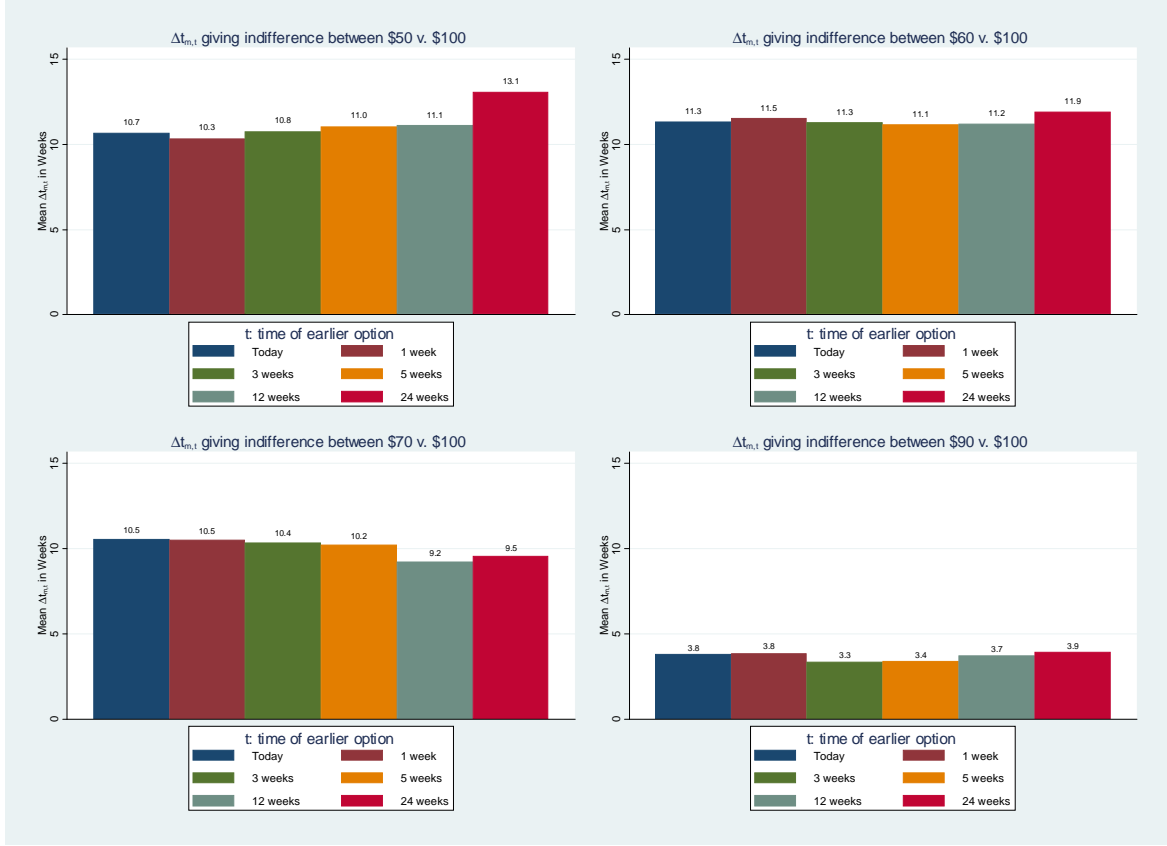
17

Figure 1: Average maximum *additional* delay accepted for larger-later reward. Sample: Analysis Sample.

to the smaller-sooner payment: $\Delta t_{i,m,t} = \theta_{im} + \xi_{it}$, where θ_{im} is a vector of indicator variables for the smaller-sooner reward magnitudes interacted with participant, and ξ_{it} is a vector of indicator variables for the different t, also interacted with participant. While $\Delta t_{i,m,t}$ can vary across participants and amounts in an arbitrary way (captured in θ_{im}), constant impatience requires that $\Delta t_{i,m,t}$ not depend on the horizon t to the smaller-sooner reward. Thus, we test the restriction that all $\xi_{it} = 0$. We reject this restriction (and thus constant impatience) via a likelihood ratio test at $p < 0.001$.[22]

---

[22]We can also test separately, for each participant, whether $\xi_{it} = 0$ for that particular i, but doing so has much less power. For 25% of the participants in the Analysis Sample, we reject that $\xi_{it} = 0$ at $p < 0.05$.

4.3. Results

Having calculated the values of the delay function $\Phi_i(m,t)$ from participants' switching points, we then seek to characterize its structure. Given the values of $\Phi_i(m,t)$ for each m,t pair, we estimate four delay functions via non-linear least squares. These delay functions vary in whether they enforce SDU and/or no time distortion (i.e. linear delay function) structure, and are as follows:

- SDU-Linear: $\Phi_i(m,t) = \beta_{im} + (1 + k_i\beta_{im})\,t$, where β_{im} varies across participants and magnitudes, and k_i varies across participants only and is common across magnitudes. Following Proposition 3.4, we constrain $k \geq 0$.

- SDU-NonLinear: $\Phi_i(m,t) = [\beta_{im}^{\gamma_i} + (1 + k_i\beta_{im}^{\gamma_i})\,t^{\gamma_i}]^{\frac{1}{\gamma_i}}$, which adds the time distortion parameter γ_i that can vary across participants; we still require $k_i \geq 0$.

- NonSDU-Linear: $\Phi_i(m,t) = \beta_{im} + \alpha_{im}t$, where α_{im}, β_{im} vary across participants and magnitudes.

- NonSDU-NonLinear: $\Phi_i(m,t) = [\beta_{im}^{\gamma_i} + \alpha_{im}t^{\gamma_i}]^{1/\gamma_i}$, which adds the time distortion parameter γ_i that can vary across participants.

The SDU-Linear model is of particular interest, since most economic applications of discounting assume SDU and do not allow for time distortion. When we estimate the SDU-Linear model, we find the distribution of k_i seen in Figure 2: the mean k_i is 0.028, and the median k_i is virtually zero (2.13×10^{-16}), suggesting the the majority of our sample are exponential discounters. In fact, for 75% of our sample, we cannot reject the restriction that $k_i = 0$ at the $p < 0.05$ level. (We find similarly small median k_i in the SDU-NonLinear model as well, as shown in Table 2)

However, what appears to be evidence of exponential discounting is instead an artifact of imposing the SDU model. We reject the SDU restrictions, indicating participants cannot be exponential (SDU) discounters. Because each model is more general than the previous ones, we can use a likelihood ratio test to assess the restrictions implied SDU and linearity (no time distortion). The SDU model is simply the NonSDU-NonLinear model with the constraints that $\gamma_{im} = 1$ and $\alpha_{im} = 1 + k_i\beta_{im}$. The parameter estimates in the Non-Linear Non-SDU then tell us how choices deviate from the SDU model.

Table 2 shows that we strongly reject the restrictions implied by the SDU model using our likelihood ratio tests. For 68% of the sample we can reject at $p < 0.05$ the restrictions implied by the SDU-Linear model in favor of the NonSDU-NonLinear
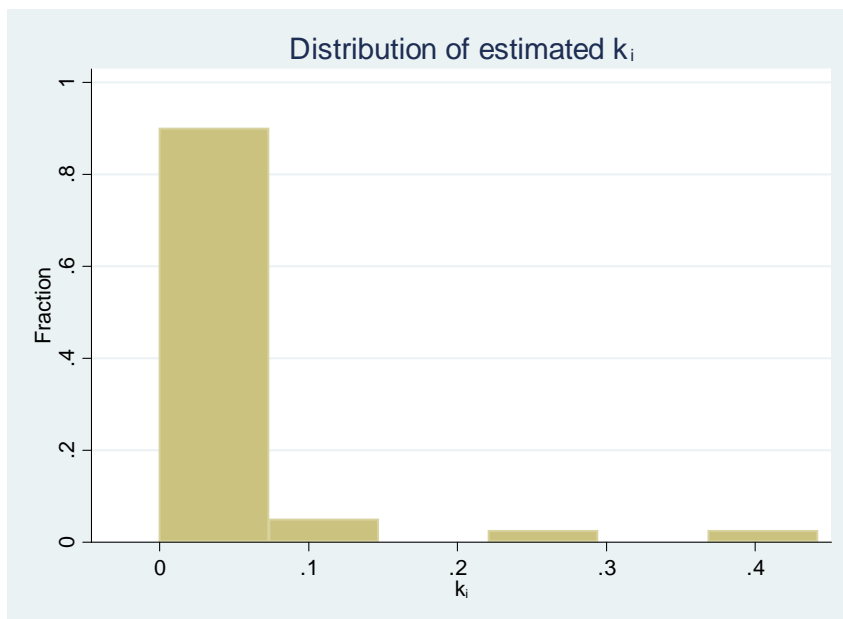
19

Figure 2: Distribution of Estimated k_i Under the Assumption of the SDU-Linear Model. Sample: Analysis Sample.

model; we reject the SDU-NonLinear in favor of the NonSDU-NonLinear model for 70% of the sample.[23]

The non-linearity of the delay function (i.e. time distortion) is not crucial to the results. We can only reject the NonSDU-Linear model in favor of NonSDU-NonLinear for 38% of the sample. In additional to the results in the table, we have also compared the SDU-Linear model directly to the NonSDU-Linear model. Doing so simply fixes $\gamma_i = 1$ and tests the restriction that $\alpha_{im} = 1 + k_i\beta_{im.}$ For 58% of the sample, we can reject at $p < 0.05$ the restriction implied by the SDU model in favor of the Linear Non-SDU model.

We then use a commonly used model selection criterion, the Akaike Information Criterion (AIC), to determine the "best-fit" model, without privileging any model

---

[23]While it is slightly surprising that we reject the more general SDU model for a larger fraction of the population, this can occur because the null model being tested against the unrestricted model differs. For the participant for which we reject the SDU-NonLinear model but not the SDU-Linear model at p<0.05, we have verified that while the SDU-NonLinear fits slightly better than the SDU-Linear model, the test rejects SDU-NonLinear at p=0.032 and SDU-Linear at p=0.054 because the degrees of freedom for the SDU-NonLinear likelihood ratio test are 4 instead of 5.

as the null hypothesis. Of course, a model with more parameters will of course be able to capture more variation in the data. However, the AIC trades off a penalty for additional parameters against the improved fit to select among models. Table 2 shows that one of the non-SDU models are preferred for 82% of participants. The most frequently selected model is the NonSDU-NonLinear model, chosen for 53% of participants. In addition to the results in the table, we have also removed consideration of non-linearity and compared the SDU-Linear model to the NonSDU-Linear model. In this case, the AIC prefers the NonSDU-Linear model 73% of the time over the SDU-Linear model.

Table 2: Model Selection and Delay Function Parameters

Model	% of Sample For Which Model is Chosen by AIC:	Rejected by LR Test:	Parameters: Mean, [Median], (Std. Dev.)					
			$\beta_{m=50}$	$\beta_{m=90}$	$\alpha_{m=50}$	$\alpha_{m=90}$	k_i	γ_i
SDU-Linear	13%	68%	10.41 [7.55] (9.84)	3.43 [2.46] (3.53)			2.84E-02 [2.13E-16] (0.08229)	
SDU-NonLinear	5%	70%	10.72 [7.42] (10.22)	3.90 [3.19] (3.65)			4.95E-02 [5.41E-12] (0.11284)	1.03 [1.01] (0.06)
NonSDU-Linear	30%	38%	10.40 [6.76] (9.59)	3.58 [3.13] (3.22)	1.10 [1.04] (0.38)	1.01 [1.00] (0.19)		
NonSDU-NonLinear	53%	NA	9.80 [7.39] (8.66)	3.54 [3.19] (3.39)	1.36 [1.05] (1.42)	1.05 [1.01] (0.25)		1.35 [1.14] (1.19)

Sample: Analysis Sample. Median parameters displayed. Mean parameters displayed, with medians in brackets and standard deviations in parentheses below. For the NonSDU-NonLinear model, the means and standard deviations exclude one extreme outlier subject. Likelihood ratio tests conducted relative to NonSDU-NonLinear model.

For each model, the row in Table 2 displays the mean and median estimated parameters for the all participants. Because each participant is described by a vector of parameters (e.g. β_{im} for $m = 50, 60, ..., 90$) we display a subset of the parameters–those for the lowest and highest m. The various $\beta_{m=50}$ parameters imply that for a decision between \$50 today and \$100 sometime in the future, the median participant would be willing to wait 6.76 weeks (as predicted by the NonSDU-Linear model) to 7.55 weeks (as predicted by the SDU-Linear model).[24] The lower $\beta_{m=90}$ parameter indicates that the median participant is willing to wait is only 2.5-3.2 weeks for \$100 over \$90 today. An mean $\alpha_{m=50}$ value of 1.36 means that for each additional week the smaller-sooner payment is delayed ($t = 1, 2, ...$) , on average participants would be willing to wait an additional 0.36 weeks (=1.36-1) after the time of the smaller-sooner payment. (Recall, the delay function is measured in absolute time, not time relative to the smaller-sooner payment.) While the median subject has an α near 1, close to constant impatience, there can is substantial variation around the median. Finally, γ captures the non-linearity in the delay function; it is more clearly described in the figure.

To illustrate the different delay functions, we examine one particular participant in Figure 3; the population average parameters mask substantial individual heterogeneity. The figure displays the participant's estimated delay functions for two different smaller-sooner amounts m, for the SDU-Linear, NonSDU-Linear and NonSDU-NonLinear models (we omit the SDU-NonLinear model for readability). Each panel also includes a 45° line, which is the delay function that would be produced by SDU exponential discounting. First, note that on the right panel ($m = 90$), the lines virtually overlap and are close to the 45° line, showing the models don't make very different predictions for the choice between \$90 and \$100. However, the models substantially differ in the left panel, which describes the choice between \$50 and \$100. Note that the NonSDU-Linear model and SDU-Linear model lines cross– we find the SDU-Linear model predicts more patience when time to the smaller-sooner payment $t \approx 0$, but less patience when t is above 10. The NonSDU-NonLinear model predicts a similar willingness-to-wait to the other models when the smaller-sooner payment is available at short time horizons (t near zero). Yet it quickly diverges when the smaller-sooner payment is available at a longer delay. At about 1 month to the smaller-sooner payment ($t = 4$), the NonSDU-NonLinear delay function shows much more patience than the SDU-Linear model: a willingness-to-wait of about 36 weeks, 12 weeks more than the approximately 25 weeks predicted by the other models.

---

[24]Note that while Figure 1 shows the mean additional delay accepted for \$50 v. \$100 when $t = 0$ is about 10 weeks, the median additional delay is approximately 8 weeks, near to what is implied by the $\beta_{m=50}$ estimates.
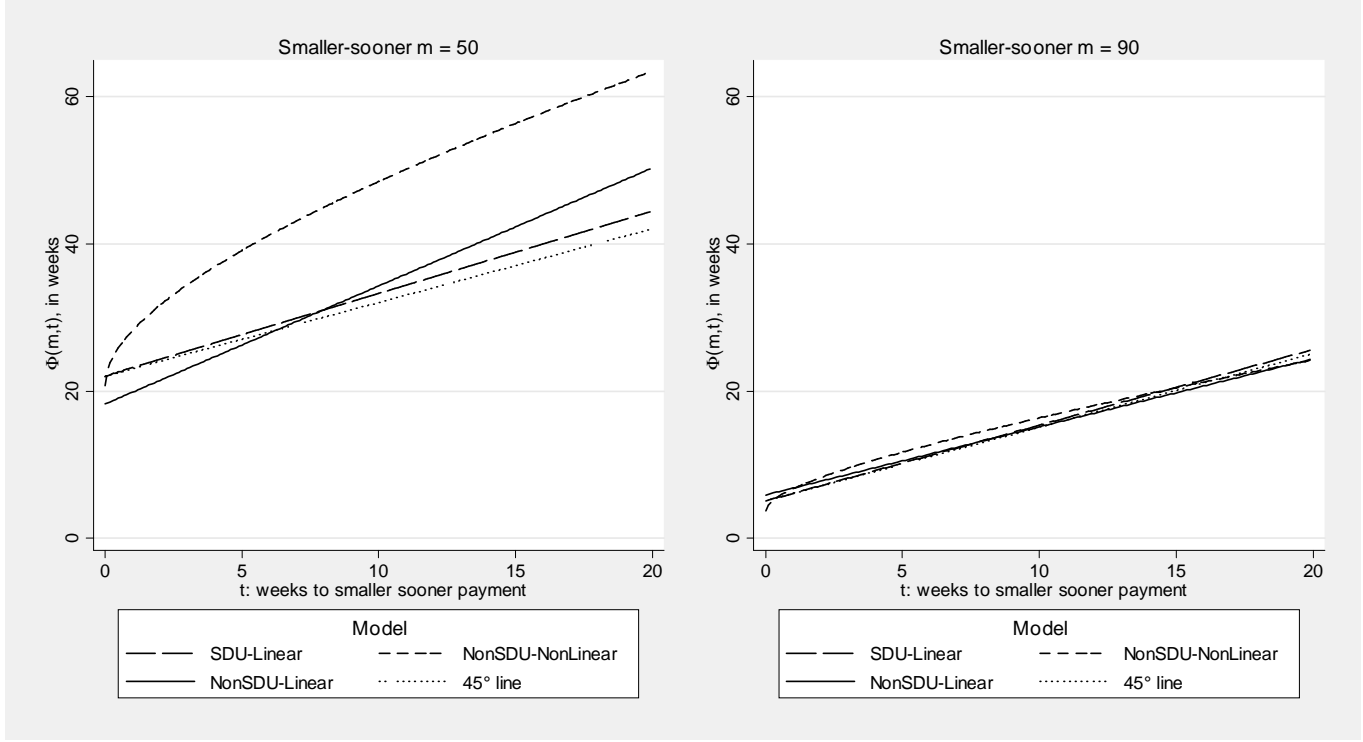
Figure 3: Delay Function Models Plotted Using an Example Participant's Estimates. Note: Parameters are as follow. SDU-Linear model: $k_i = 0.0055$, $\beta_{50} = 22.0$, $\beta_{90} = 5.1$. NonSDU-Linear: $\beta_{50} = 18.2$, $\alpha_{50} = 1.60$, $\beta_{90} = 5.9$, $\alpha_{90} = 0.92$. NonSDU-NonLinear: $\gamma = 0.50$, $\beta_{50} = 20.7$, $\alpha_{50} = 0.77$, $\beta_{90} = 3.7$, $\alpha_{90} = 0.67$.

Finally, a quick glance at Table 2 suggests that the results satisfy the conditions that α_{im} and β_{im} decrease in m, as required by regularity. We more rigorously verify this for the most general NonSDU-NonLinear model by estimating a linear relationship between α_{im} and m and between β_{im} and m. Pooling all participants to address noise,[25] we find the expected statistically significant negative trend on both cases: $\alpha_{im} = -0.009m$, $\beta_{im} = -0.159m$, rejecting a zero coefficient with $p < 0.10$ in both cases. (While our participant-specific estimates are subject to more noise, 68% of the participant-specific trends are negative, and only 1 out of 80 trends is positive and statistically significant.)

---

[25]We drop the one outlier participant with extreme parameters values (e.g. $\gamma = 27$, $\alpha_{50} = 4.97e+09$.) Including this participant still gives aggregate negative trends, significant for β but not for α.

4.4. Robustness

The common SDU discount functions (exponential, hyperbolic, generalized hyperbolic) produce a linear delay function and are thus appropriately tested for in our analysis above. However, some SDU discount functions treat $t = 0$ specially and produce a nonlinear delay function that is not of the form considered in Proposition 3.3. For instance, the quasi-hyperbolic $\beta - \delta$ model produces a linear delay function when no option is available at $t = 0$, but has a nonlinearity at $t = 0$. Similarly, the Benhabib, Bisin, and Schotter (2010) fixed-cost of delay model, in which payments delayed from $t = 0$ to $t > 0$ incur a fixed-cost, has a nonlinearity at $t = 0$; nonetheless, it produces a linear delay function when $t > 0$. Additionally, immediate $t = 0$ payments may be treated differently due to confounds, such as trust, uncertainty, or transactions costs.

To account for special treatment of $t = 0$, we repeat the above analysis excluding all choices in which the smaller-sooner option is available at $t = 0$. We compare the SDU-Linear model (which would be produced by the quasi-hyperbolic model and fixed-cost of delay model) to the NonSDU models. Results are very similar:[26] with a likelihood ratio test, the SDU-Linear model is rejected in favor of the NonSDU-Linear model for 64% of participants and rejected in favor of the NonSDU-NonLinear model for 69% of participants. Moreover, the AIC best-fit criterion only chooses the SDU-Linear model for 22% of subjects.

We also explore robustness to sample selection. We examine an expanded sample that drops only participants who always choose the larger-later option for more than 10 of the m, t pairs or who ever make a non-monotonic choice.[27] For the resulting 51 participants, the AIC model selection criteria prefers the SDU-Linear model for only 24% of the sample, and we still reject the SDU-Linear model for 63% of the sample in favor of the Non-Linear Non-SDU model via a likelihood ratio test.

5. Concluding Remarks

This paper approaches the practical question of how to design intertemporal choice experiments from the perspective of decision theory: the behavioral foundations for

---

[26]Dropping the $t = 0$ choices, we have difficult fitting the model for 4 participants, who we then exclude from this analysis.

[27]In this expanded sample, we need to impute the indifference point for lists on which the participant always chose the larger-later option. We impute their choices as though they choose the earlier option when the later option was available at an additional delay of 93 weeks, a 25 week increment over the last option they actually faced. We explored alternative imputation strategies, which gave similar results.

the object of interest (discount functions) is studied by writing a behavioral definition and characterizing the connection between the two. The general mapping is obtained by behaviorally defining the discount function by the delay function Φ, and second by noting that functional equations provide a means of characterizing the connection between Φ and the representations (D, u). This result suggests a procedure for experimentally testing theories and eliciting discount functions.

While separability is an assumption made for the sake of parsimony in economic models, we evaluate whether there is an empirical price for assuming it. In both our main experiment and pilot experiment, we find that more than half our participants are not well-characterized by separable discounted utility. More than simply rejecting SDU, we show that the form of our estimated delay functions are consistent with a magnitude-dependent discount function. The general framework we examine invites further exploration. Our results suggest that exploring non-SDU preferences will be fruitful, and that empirical applications might benefit from searching for explanations for observed behavior that do not rely on SDU.

On the extent of magnitude dependence, our experiment shows that subjects treat \$50 differently enough than \$100 to give rise to a highly significant rejection of separability. Take the canonical representation $D(m, t) = e^{-a(m) \cdot t^\gamma}$ where all parameters are fixed by the data, $a(50) = 1.36$ from Table 2 and $a(100) = 1$ by regularity. With a one week delay $t = 1$ (in order to shut down time distortion), we see that

$$\frac{D(50, 1)}{D(100, 1)} = e^{t^\gamma (a(100) - a(50))} = 0.70,$$

that is, the agent discounts \$50 much heavier than she does \$100. It seems intuitive that \$50 and \$100 might be treated very differently, just as it is also intuitive that for larger stakes a \$50 difference may not lead to a very pronounced difference.

While our delay function approach is defined for the analysis of time preference, it can be adapted easily to other domains as well. For instance, experiments on risk often offer subjects lotteries that have one nonzero payoff. Such lotteries can be written as (m, p), where p is the probability of the nonzero outcome. By defining 'time' as $t = \frac{1}{p} - 1$ our procedure becomes immediately applicable to the study of risk preference, where the general representation takes the form $U(m, p) = f(m, p)u(m)$ and where f is the decision weight.

A. Appendix: Regularity and Proof of Proposition 2.1

Say that a preference $\succsim$ over X is *regular* if it satisfies the following basic restrictions:[28]

1- **Order**: $\succsim$ is complete and transitive.

2- **Continuity**: For each (m,t), the sets $\{(m',t') : (m',t') \succsim (m,t)\}$ and $\{(m',t') : (m,t) \succsim (m',t')\}$ are closed.

3- **Impatience**:

(i) For all $m > 0$ and $t < t'$, $(0,t) \sim (0,t')$ and $(m,t) \succ (m,t')$.

(ii) For each m, m' such that $m' > m > 0$, there is t such that $(m,0) \succ (m',t)$.

4- **Monotonicity**: For all t, if $m < m'$ then $(m',t) \succ (m,t)$.

We establish some basic results on regularity which are used later, though not always explicitly.

Lemma A.1. *For any continuous increasing $u : \mathcal{M} \to \mathbb{R}$, a regular preference $\succsim$ admits a representation $U : \mathcal{M} \times \mathcal{T} \to \mathbb{R}$ such that $U(\cdot, t)$ is continuous and strictly increasing, $U(m, \cdot)$ is continuous and strictly decreasing if $m > 0$ and constant if $m = 0$, and $U(m, 0) = u(m)$. Conversely, any preference that admits such a representation is regular. By defining $D(m,t) = \frac{U(m,t)}{u(m)}$ for any $m > 0$, any such representation can be written as a GDU representation (D, u).*

Proof. The first claim is established in [1982, Thm 1]. The remaining are trivial. ∎

Lemma A.2. *If $\succsim$ is regular then*

(a) For every m, t and d there exists $m' \leq m$ such that $(m', t) \sim (m, t + d)$. Moreover, for every m, t and $m' \leq m$ there exists d such that $(m', t) \sim (m, t + d)$.

(b) For any $s \leq l$ and τ such that $(s, 0) \sim (l, \tau)$, and for every $t' \geq \tau$ there exists t such that $(s, t) \sim (l, t')$. Moreover, when $s > 0$ then for any $t \geq 0$ there is a unique $T \geq t$ such that $(s, t) \sim (l, T)$.

(c) For each (m, t) there exists a unique 'present value' $\psi(m, t)$ satisfying

$$(\psi(m, t), 0) \sim (m, t).$$

Moreover, $\psi(0, \cdot) = 0$, $\psi(m, \cdot)$ is strictly decreasing for any $m > 0$, $\lim_{t \to \infty} \psi(m, t) = 0$ for all m, and $\psi(m, \cdot)$ is continuous.

(d) If $(s, 0) \sim (l, \tau)$ and $(s, t) \sim (l, T + \tau)$, then $T + \tau \geq t$.

---

[28]Continuity presumes that $\mathbb{R}_+$ has the euclidean topology, any subset of $\mathbb{R}_+$ has the subspace topology, and X has the product topology.

Proof. Part (a) follows from Impatience, Monotonicity and Continuity; we omit the proof. The t in part (b) exists by Impatience, Monotonicity and Continuity: By Monotonicity, $(s, t') \precsim (l, t')$. By Impatience and the fact that $(s, 0) \sim (l, \tau)$ and $t' \geq \tau$, it follows that $(s, 0) \succsim (l, t')$. Thus, by Continuity, $(s, 0) \succsim (l, t') \succsim (s, t')$ implies that there is t such that $(s, t) \sim (l, t')$, as desired. For the second claim in (b), the existence of T is established in a similar way. Impatience guarantees uniqueness when $s > 0$.

Turning to part (c): part (a) establishes the existence of present values, and Impatience implies that $\psi(m, \cdot)$ is strictly decreasing for any $m > 0$. To see that $\lim_{t \to \infty} \psi(m, t) = 0$ for all m, suppose not. Then there exists m and $s > 0$ such that $(s, 0) \prec (\psi(m, t), 0) \sim (m, t)$ for all t. But this contradicts Impatience. Finally, to see that $\psi(m, \cdot)$ must be continuous, take any strictly increasing homeomorphism and consider the representation U delivered in Lemma A.1. Since $u(\psi(m, t)) = U(m, t)$ and in particular, $\psi(m, t) = u^{-1}(U(m, t))$, continuity of u^{-1} implies that of $\psi(m, \cdot)$.

For part (d), note that if $T + \tau < t$ then $(s, T + \tau) \succ (l, T + \tau)$ by Impatience, which then violates Monotonicity. ∎

The next lemma (stated as Proposition 2.1 in the text) characterizes regularity in terms of properties of Φ. Say that $\Phi : \mathcal{M} \times \mathcal{T} \to \mathcal{T}$ is *generated* by $\succsim$ if for any $0 < m \leq \overline{m}$ and each t, $(m, t) \sim (\overline{m}, \Phi(m, t))$.

Lemma A.3. Φ *is generated by a regular preference* $\succsim$ *if and only if:*
 (i) $\Phi(m, t)$ *is continuous.*
 (ii) For any t, $\Phi(\cdot, t)$ *is strictly decreasing and* $\lim_{m \to 0} \Phi(m, t) = \infty$.
 (iii) For $m > 0$, $\Phi(m, \cdot)$ *is strictly increasing and* $\Phi(\overline{m}, t) = t$ *for all* t.

Proof. Prove the 'if' part. Let $\Phi_0^{-1}(t)$ be defined by $\Phi(\Phi_0^{-1}(t), 0) = t$. Define a function $U(m, t) = \Phi_0^{-1}(\Phi(m, t))$, where the inverse exists and is continuous by the monotonicity and continuity properties in (i)-(ii). Intuitively, $U(m, t)$ is the present value of (m, t), that is, if there was a regular preference generating Φ then $(x, 0) \sim (m, t) \sim (\overline{m}, \Phi(m, t))$ and $(x, 0) \sim (\overline{m}, \Phi(x, 0))$ would hold. Thus $\Phi(x, 0) = \Phi(m, t)$, and in turn, $U(m, t) = x = \Phi_0^{-1}(\Phi(m, t))$.

We first verify that U represents a regular preference. By (ii), for fixed t, since $\Phi(m, t)$ strictly decreases in m and Φ_0^{-1} is also strictly decreasing, it follows that $\Phi_0^{-1}(\Phi(m, t))$ is strictly increasing in m. Therefore $U(m, t)$ is strictly increasing in m. Similarly, $U(m, t)$ is strictly decreasing in t if $m > 0$. By continuity of U and by (ii), $U(0, t) = \lim_{m \to 0} U(m, t) = \lim_{m \to 0} \Phi_0^{-1}(\Phi(m, t)) = 0$ for any t. The other Impatience property follows from the fact that by (ii) and (iii), $0 \leq$

$\lim_{t\to\infty} U(m,t) = \lim_{t\to\infty} \Phi_0^{-1}(\Phi(m,t)) \leq \lim_{t\to\infty} \Phi_0^{-1}(\Phi(\overline{m},t)) = \lim_{t\to\infty} \Phi_0^{-1}(t) = 0$, that is, $\lim_{t\to\infty} U(m,t) = 0$. As already noted, U is continuous.

Finally, we check that Φ is generated by the preference $\succsim$ represented by U, that is, $U(m,t) = U(\overline{m}, \Phi(m,t))$. Note that by definition and by (iii), $U(\overline{m},t) = \Phi_0^{-1}(\Phi(\overline{m},t)) = \Phi_0^{-1}(t)$. Thus, $U(\overline{m}, \Phi(m,t)) = \Phi_0^{-1}(\Phi(m,t)) = U(m,t)$, as desired.
∎

B. Appendix: Proof of Theorem 2.2

B.1. Proof

We formally prove the result by taking $\mathcal{M} = \mathbb{R}_+$ and noting that the same argument establishes Theorem 2.2 as a corollary when $\mathcal{M} = [0, \overline{m}]$.

For a given preference $\succsim$ and any rewards $0 < s \leq l$, define the function $\Phi_{s,l}(\cdot)$ by the indifference:

$$(s,t) \sim (l, \Phi_{s,l}(t)). \tag{B.1}$$

For $s = 0 < l$, let $\Phi_{s,l}(t) := \infty$.

We first clarify the exhaustive implications of regularity on Φ.

Lemma B.1. Φ is generated by a regular $\succsim$ if and only if:

(i) $\Phi(s,l,t)$ is continuous,

(ii) $\Phi(s,\cdot,t)$ is strictly increasing and $\Phi(\cdot,l,t)$ is strictly decreasing in s, and moreover $\lim_{s\to 0} \Phi(s,l,t) = \infty$ when $l > 0$,

(iii) $\Phi(s,l,\cdot)$ is strictly increasing if $s,l > 0$, and $\Phi(m,m,t) = t$ for all t,

(iv) $\Phi_{m_1,m_2}(\Phi_{m_0,m_1}(t)) = \Phi_{m_0,m_2}(t)$ for all t and $m_0 \leq m_1 \leq m_2$.

Proof. Prove the 'if' part. Define $\Phi_{(l,0)}^{-1}(r)$ by $\Phi(\Phi_{(l,0)}^{-1}(r),l,0) = r$. Let $U(m,t) := \Phi_{m,0}^{-1}(t)$, where the inverse exists by the monotonicity and continuity properties in (i)-(ii). Intuitively, $U(m,t)$ is the present value of (m,t), that is, it is a small reward $s = U(m,t)$ that satisfies,

$$\Phi(s,m,0) = t.$$

We first verify that U represents a regular preference.

To see that U is continuous, suppose $m_n \to m$, $t_n \to t$ and to ease notation write $s_n := U(m_n,t_n)$, that is, $\Phi(s_n,m_n,0) = t_n$. We show that s_n converges. Since $m_n \to m$, there is some M and N such that $m_n \leq M$ for all $n \geq N$ (wlog let $N = 1$). Define $T_n := \Phi(m_n,M,t_n)$. By (i), T_n converges. Observe that $\Phi(s_n,M,0) = T_n$ by (iv). Since $\Phi(\cdot,M,0)$ is strictly monotone and continuous, it follows that $\Phi_{M,0}^{-1}(\cdot)$

is continuous. Therefore, since T_n converges to $T := \Phi(m, M, t)$, it must be that $s_n = \Phi_{M,0}^{-1}(T_n)$ converges to $s := \Phi_{M,0}^{-1}(T) = \Phi_{M,0}^{-1}(\Phi(m, M, t))$, and in particular,

$$\Phi(s, M, 0) = \Phi(m, M, t).$$

It remains to show that $\Phi(s, m, 0) = t$. By the displayed equality and (iv), $\Phi(m, M, \Phi(s, m, 0)) = \Phi(s, M, 0) = \Phi(m, M, t)$ and so by (iii), $\Phi(s, m, 0) = t$, as desired. Thus U is continuous.

Now show the remaining regularity properties. By (ii), for fixed t, the equation $\Phi(s, m, 0) = t$ implies that as m increases, s must also increase. Therefore $U(m, t)$ is strictly increasing in m. Similarly, $U(m, t)$ is strictly decreasing in t if $m > 0$. To show the second Impatience property, take any m, m' such that $m' > m > 0$. By (ii), there is a small enough $s' > 0$ s.t. $\Phi(s', m', 0) > 0 = \Phi(m, m, 0)$. Define $t := \Phi(s', m', 0)$. Then by (ii), $U(m', t) < U(m, 0)$, as desired. Finally we show the first Impatience property, that is, $U(0, t) = 0$. By (ii) and (iii), since $\Phi(m, m, t) = t$, it must be that for s that satisfies $\Phi(s, m, 0) = t$ it must be that $s \leq m$. That is, $0 \leq U(m, t) \leq m$. Then by continuity of U, $U(0, t) = \lim_{m \to 0} U(m, t) = 0$, as desired.

To conclude, we check that Φ is generated by the preference $\succsim$ represented by U. By definition, for any $s \leq l$, $s = U(l, \Phi(s, l, 0))$. Take any t and suppose $s'' = U(l, \Phi(s, l, t))$, that is,

$$\Phi(s'', l, 0) = \Phi(s, l, t).$$

Since (iii) implies $\Phi(s, l, t) \geq \Phi(s, l, 0)$, it follows that $s'' = U(l, \Phi(s, l, t)) \leq U(l, \Phi(s, l, 0)) = s$. That is, $s'' \leq s \leq l$. By (iv), $\Phi(s, l, \Phi(s'', s, 0)) = \Phi(s'', l, 0)$ and so, by the displayed equality, $\Phi(s, l, \Phi(s'', s, 0)) = \Phi(s, l, t)$. By (iii), $\Phi(s'', s, 0) = t$, and this implies that $U(s, t) = s'' = U(l, \Phi(s, l, t))$, and thus Φ is generated by U, as desired. ∎

Lemma B.2. *If D solves the functional equation (FE) below, then for any $0 < m_1 \leq m_2 \leq m_3$,*

$$D(m_2, \Phi_{m_1,m_2}(0)) \cdot D(m_3, \Phi_{m_2,m_3}(0)) = D(m_3, \Phi_{m_1,m_3}(0))$$

Proof. Suppose $m_1 \leq m_2 \leq m_3$ then the functional equation implies

$$D(m_2, \Phi_{m_1,m_2}(0)) \cdot D(m_3, \Phi_{m_2,m_3}(0)) = D(m_3, \Phi_{m_2,m_3}(\Phi_{m_1,m_2}(0))).$$

But transitivity of $\succsim$ implies $\Phi_{m_2,m_3}(\Phi_{m_1,m_2}(0)) = \Phi_{m_1,m_3}(0)$. The assertion follows. ∎

Lemma B.3. *The following statements hold:*

(a) Consider any regular preference $\succsim$ and its Φ-function. Then D can be attributed to $\succsim$ if and only if D solves the functional equation:

$$D(s,t) \cdot D(l, \Phi_{s,l}(0)) = D(l, \Phi_{s,l}(t)), \tag{FE}$$

for all $0 < s \leq l$ and t.

(b) Suppose $\succsim$ is a regular preference with the function Φ, and that D is a solution to (FE). Then (D, u) represents $\succsim$ if and only if u is given by

$$u(m) = \begin{cases} D(\overline{m}, \Phi_{m,\overline{m}}(0)) \cdot u(\overline{m}) & \text{if } m \leq \overline{m} \\ [D(m, \Phi_{\overline{m},m}(0))]^{-1} \cdot u(\overline{m}) & \text{otherwise} \end{cases}, \quad \text{for all } m,$$

where $\overline{m} > 0$ and $u(\overline{m}) > 0$ are arbitrary.

Proof. We prove (a), and part (b) follows as a corollary of the proof.

First show that any attributable D must satisfy the functional equation. By regularity, $\succsim$ admits a representation U. Any representation can be written as a GDU model with some D and u. By definition of the Φ-function (1.1), it must be that for all $s, l > 0$ and t, both $u(s) = D(l, \Phi_{s,l}(0))u(l)$ and $D(s,t)u(s) = D(l, \Phi_{s,l}(t))u(l)$ hold. Rearranging yields the functional equation. Observe that we have also determined that a solution must always exist if Φ comes from a regular preference $\succsim$.

For the converse, suppose D is a solution. Take any $\overline{m} > 0$ and assign it any utility $u(\overline{m}) > 0$. Define

$$u(m) = \begin{cases} D(\overline{m}, \Phi_{m,\overline{m}}(0))u(\overline{m}) & \text{if } m \leq \overline{m} \\ D(m, \Phi_{\overline{m},m}(0))^{-1}u(\overline{m}) & \text{otherwise} \end{cases}, \quad \text{for all } m.$$

By continuity of D, the utility u is continuous as well (monotonicity will be determined shortly). Next we show that, given transitivity of $\succsim$, the utility u is consistent with D in the sense that it satisfies

$$u(s) = D(l, \Phi_{s,l}(0))u(l) \tag{B.2}$$

for all s, l s.t. $s \leq l$. To see this, consider the following cases:

Case 1- $s, l \leq \overline{m}$.

Then $u(s) = D(\overline{m}, \Phi_{s,\overline{m}}(0))u(\overline{m})$ and $u(l) = D(\overline{m}, \Phi_{l,\overline{m}}(0))u(\overline{m})$, which implies

$$u(s) = \frac{D(\overline{m}, \Phi_{s,\overline{m}}(0))}{D(\overline{m}, \Phi_{l,\overline{m}}(0))}u(l).$$

31

By the Lemma, $D(l, \Phi_{s,l}(0)) \cdot D(\overline{m}, \Phi_{l,M}(0)) = D(\overline{m}, \Phi_{s,M}(0))$, that is, $\frac{D(\overline{m}, \Phi_{s,\overline{m}}(0))}{D(\overline{m}, \Phi_{l,\overline{m}}(0))} = D(l, \Phi_{s,l}(0))$. It follows that (B.2) holds.

Case 2- $s \le \overline{m} \le l$

Then $u(s) = D(\overline{m}, \Phi_{s,\overline{m}}(0))u(\overline{m})$ and $u(l) = \frac{u(\overline{m})}{D(l,\Phi_{\overline{m},l}(0))}$, which implies

$$u(s) = D(\overline{m}, \Phi_{s,\overline{m}}(0))D(l, \Phi_{\overline{m},1}(0))u(l).$$

By the Lemma, $D(\overline{m}, \Phi_{s,\overline{m}}(0)) \cdot D(l, \Phi_{\overline{m},l}(0)) = D(l, \Phi_{s,l}(0))$, and (B.2) follows.

Case 3- $\overline{m} \le s \le l$.

Then $u(s) = \frac{u(\overline{m})}{D(s,\Phi_{\overline{m},s}(0))}$ and $u(l) = \frac{u(\overline{m})}{D(l,\Phi_{\overline{m},l}(0))}$, which implies

$$u(s) = \frac{D(l, \Phi_{\overline{m},l}(0))}{D(s, \Phi_{\overline{m},s}(0))}u(l).$$

By the Lemma, $D(s, \Phi_{\overline{m},s}(0)) \cdot D(l, \Phi_{s,l}(0)) = D(l, \Phi_{\overline{m},l}(0))$, and (B.2) follows.

Thus u is consistent with D in the sense of (B.2). Observe that the equality also assures us that u must be strictly increasing: D is strictly increasing in its second argument and by Monotonicity and Impatience $\Phi_{s,l}(0)$ must be strictly increasing in s. To show that there is a GDU representation with D, define $U(m, t) := u(p(m, t))$, where $p(m, t)$ is the present value of (m, t). Since $p(m, t)$ is a representation for $\succsim$ and u is strictly increasing, it follows that $U(m, t)$ represents $\succsim$. But then $U(m, t) = u(p(m, t)) = D(m, t)u(m)$, as desired. ∎

The next lemma determines how to check if D solves (FE) on the basis of information on $\Phi_{m,\overline{m}}$, $\Phi_{\overline{m},m}$ and the present value of (m, t) for all $m > \overline{m}$ and $t < \Phi_{\overline{m},m}(0)$. Write p_{mt} for the present value of (m, t), that is, $(p_{mt}, 0) \sim (m, t)$.

Lemma B.4. *Fix any $\overline{m} > 0$. Then D solves (FE) for all $0 < s \le l$ and t if and only if:*

i) D solves (FE) for all s, l s.t. $0 < s \le l$ for $s = \overline{m}$ or $l = \overline{m}$, and all t; and

ii) $D(m, t) = \frac{D(m, \Phi_{\overline{m},m}(0))}{D(p_{mt}, \Phi_{\overline{m},p_{mt}}(0))}$ for all $m > \overline{m}$ and $t < \Phi_{\overline{m},m}(0)$.

Proof. The 'if' part is straightforward – note that part (ii) follows from lemma B.2. Turn to the 'only if' part. Suppose the hypothesis holds. Take any $0 < s \le l$ and t. Consider the following cases. We make frequent use of the fact that if $m_1 \le m_2 \le m_3$ then transitivity implies $\Phi_{m_2,m_3}(\Phi_{m_1,m_2}(t)) = \Phi_{m_1,m_3}(t)$ for any t.

Case 1- $s, l \le \overline{m}$.

By hypothesis, $D(s, t) \cdot D(\overline{m}, \Phi_{s,\overline{m}}(0)) = D(\overline{m}, \Phi_{s,\overline{m}}(t))$ and $D(l, t) \cdot D(\overline{m}, \Phi_{l,\overline{m}}(0)) = D(\overline{m}, \Phi_{l,\overline{m}}(t))$. Moreover, by transitivity, $\Phi_{l,\overline{m}}(\Phi_{s,l}(t)) = \Phi_{s,\overline{m}}(t)$. Observe that:

$D(l, \Phi_{s,l}(t))$

$= D(l, \Phi_{l,\overline{m}}^{-1}(\Phi_{s,\overline{m}}(t)))$ by transitivity

$= \frac{D(\overline{m}, \Phi_{l,\overline{m}}[\Phi_{l,\overline{m}}^{-1}(\Phi_{s,\overline{m}}(t))])}{D(\overline{m}, \Phi_{l,\overline{m}}(0))}$ by hypothesis

$= \frac{D(\overline{m}, \Phi_{s,\overline{m}}(t))}{D(\overline{m}, \Phi_{l,\overline{m}}(0))}$

$= \frac{D(s,t)D(\overline{m}, \Phi_{s,\overline{m}}(0))}{D(\overline{m}, \Phi_{l,\overline{m}}(0))}$ by hypothesis

$= D(s,t)[\frac{D(\overline{m}, \Phi_{s,\overline{m}}(0))}{D(\overline{m}, \Phi_{l,\overline{m}}(0))}]$. We are done if we show that $\frac{D(\overline{m}, \Phi_{s,\overline{m}}(0))}{D(\overline{m}, \Phi_{l,\overline{m}}(0))} = D(l, \Phi_{s,l}(0))$. But this follows

since

$D(l, \Phi_{s,l}(0))D(\overline{m}, \Phi_{l,\overline{m}}(0))$

$= D(\overline{m}, \Phi_{l,\overline{m}}(\Phi_{s,l}(0)))$ by hypothesis

$= D(\overline{m}, \Phi_{s,\overline{m}}(0))$ by transitivity. This completes the argument.

Case 2- $s \leq \overline{m} \leq l$.

By hypothesis $D(s,t) \cdot D(\overline{m}, \Phi_{s,\overline{m}}(0)) = D(\overline{m}, \Phi_{s,\overline{m}}(t))$ and $D(\overline{m}, t) \cdot D(l, \Phi_{\overline{m},l}(0)) = D(l, \Phi_{\overline{m},l}(t))$ and by transitivity, $\Phi_{\overline{m},l}(\Phi_{s,\overline{m}}(t)) = \Phi_{s,l}(t)$. Observe that

$D(l, \Phi_{s,l}(t))$

$= D(l, \Phi_{\overline{m},l}(\Phi_{s,\overline{m}}(t)))$ by transitivity

$= D(\overline{m}, \Phi_{s,\overline{m}}(t)) \cdot D(l, \Phi_{\overline{m},l}(0))$ by hypothesis

$= [D(s,t) \cdot D(\overline{m}, \Phi_{s,\overline{m}}(0))] \cdot D(l, \Phi_{\overline{m},l}(0))$ by hypothesis

$= D(s,t) \cdot [D(\overline{m}, \Phi_{s,\overline{m}}(0)) \cdot D(l, \Phi_{\overline{m},l}(0))]$. We are done if $D(\overline{m}, \Phi_{s,\overline{m}}(0)) \cdot D(l, \Phi_{\overline{m},l}(0)) = D(l, \Phi_{s,l}(0))$. But this follows since

$D(\overline{m}, \Phi_{s,\overline{m}}(0)) \cdot D(l, \Phi_{\overline{m},l}(0))$

$= D(\overline{m}, \Phi_{\overline{m},l}(\Phi_{s,\overline{m}}(0)))$ by hypothesis

$= D(\overline{m}, \Phi_{s,l}(0))$ by transitivity. This completes the argument.

Case 3(i)- $\overline{m} < s \leq l$ and $t \geq \Phi_{\overline{m},s}(0)$.

By hypothesis $D(\overline{m}, t) \cdot D(s, \Phi_{\overline{m},s}(0)) = D(s, \Phi_{\overline{m},s}(t))$ and $D(\overline{m}, t) \cdot D(l, \Phi_{\overline{m},l}(0)) = D(l, \Phi_{\overline{m},l}(t))$ and by transitivity, $\Phi_{s,l}(\Phi_{\overline{m},s}(t)) = \Phi_{\overline{m},l}(t)$. The restriction $t \geq \Phi_{\overline{m},s}(0)$ implies that $\Phi_{\overline{m},s}^{-1}(t)$ exists. Observe that:

$D(l, \Phi_{s,l}(t))$

$= D(l, \Phi_{\overline{m},l}(\Phi_{\overline{m},s}^{-1}(t)))$ by transitivity

$= D(\overline{m}, \Phi_{\overline{m},s}^{-1}(t)) \cdot D(l, \Phi_{\overline{m},l}(0))$

$= \frac{D(s, \Phi_{\overline{m},s}(\Phi_{\overline{m},s}^{-1}(t)))}{D(s, \Phi_{\overline{m},s}(0))} \cdot D(l, \Phi_{\overline{m},l}(0))$

$= D(s,t) \cdot \frac{D(l, \Phi_{\overline{m},l}(0))}{D(s, \Phi_{\overline{m},s}(0))}$. We are done if we show that $\frac{D(l, \Phi_{\overline{m},l}(0))}{D(s, \Phi_{\overline{m},s}(0))} = D(l, \Phi_{s,l}(0))$: observe that by hypothesis and transitivity, $D(s, \Phi_{\overline{m},s}(0))D(l, \Phi_{s,l}(0)) = D(l, \Phi_{s,l}(\Phi_{\overline{m},s}(0))) = D(l, \Phi_{\overline{m},l}(0))$, as desired.

Case 3(ii)- $\overline{m} < s \leq l$ and $t < \Phi_{\overline{m},s}(0)$.

By hypothesis, $D(m,t) = \frac{D(m,\Phi_{\overline{m},m}(0))}{D(p_{mt},\Phi_{\overline{m},p_{mt}}(0))}$ where p_{mt} satisfies $(p_{mt},0) \sim (m,t)$. Since $t < \Phi_{\overline{m},s}(0)$, it must be that $p_{mt} \geq \overline{m}$.

$D(s,t) \cdot D(l, \Phi_{s,l}(0))$

$= \frac{D(s,\Phi_{\overline{m},s}(0))}{D(p,\Phi_{\overline{m},p}(0))} \cdot \frac{D(l,\Phi_{\overline{m}l}(0))}{D(s,\Phi_{\overline{m},s}(0))}$ by hypothesis, where p is the present value of (s,t) (s is the present value of $(l, \Phi_{s,l}(0))$ by definition)

$= \frac{D(l,\Phi_{\overline{m}l}(0))}{D(p,\Phi_{\overline{m}p}(0))}$

$= D(l, \Phi_{s,l}(t))$ by hypothesis since by transitivity p must be the present value of $(l, \Phi_{s,l}(t))$ as well.

This completes the proof. ∎

Fix any $\overline{m} > 0$. For any $m > 0$ and $t \geq 0$, define $\Phi(m,t)$ by:

$$(m,t) \sim (\overline{m}, \Phi(m,t)) \quad \text{if } m \leq \overline{m}$$
$$(\overline{m}, t) \sim (m, \Phi(m,t)) \quad \text{otherwise.}$$

Lemma B.5. D *is attributable iff there is a continuous, strictly increasing and unbounded function g satisfying $g(0) = 0$ such that*

$$D(m,t) = \begin{cases} e^{-[g(\Phi(m,t))-g(\Phi(m,0))]} & \text{if } m \leq \overline{m} \\ e^{-[g(\Phi_m^{-1}(t))-g(\Phi(m,0))]} & \text{if } m \geq \overline{m} \text{ and } t \geq \Phi_{\overline{m},m}(0) \\ e^{-[g(m,0)-g(\Phi(p(m,t),0))]} & \text{if } m > \overline{m} \text{ and } t < \Phi_{\overline{m},m}(0) \end{cases} .$$

Proof. By lemma B.3, D is attributable if and only if it solves (FE). Suppose D solves (FE). Then by lemma B.3, (D, u) represents $\succsim$ for some u. Wlog, let $u(\overline{m}) = 1$. By regularity, $\Phi(m,0)$ is continuous and strictly decreasing in m for $m < \overline{m}$ and strictly increasing for $m > \overline{m}$. Since u is strictly increasing, there is a continuous strictly increasing function g satisfying $g(0) = 0$ such that[29]

$$u(m) = \begin{cases} e^{-g(\Phi(m,0))} & \text{if } m \leq \overline{m} \\ e^{g(\Phi(m,0))} & \text{if } m \geq \overline{m} \end{cases} .$$

We will see shortly that g must be unbounded.[30] Given that (D, u) represents $\succsim$ and $u(\overline{m}) = 1$, the definition of Φ implies $u(m) = D(\overline{m}, \Phi(m,0))$ for $m \leq \overline{m}$, and $u(m) = D(m, \Phi(m,0))^{-1}$ for $m > \overline{m}$. Therefore,

$$D(\overline{m}, \Phi(m,0)) = e^{-g(\Phi(m,0))} \quad \text{for } m \leq \overline{m}$$
$$D(m, \Phi(m,0)) = e^{-g(\Phi(m,0))} \quad \text{for } m \geq \overline{m} \quad .$$

---

[29] By regularity $\Phi(\overline{m},0) = 0$ and so $u(\overline{m}) = e^{-g(\Phi(\overline{m},0))} = e^{-g(\Phi(\overline{m},0))} = 1$, consistent with our assumption that $u(\overline{m}) = 1$.

[30] This does not imply that u is unbounded: though $u(m) = e^{g(\Phi(m,0))}$ for $m \geq \overline{m}$ for unbounded g, regularity does not require $\Phi(\cdot, 0)$ to be unbounded.

We use this observation below. Another observation is that by regularity $\Phi(m, 0)$ ranges from 0 to ∞ as m varies over $[0, \overline{m}]$, and so we have

$$D(\overline{m}, t) = e^{-g(t)}.$$

Moreover, since D is a discount function it must satisfy the property $\lim_{t \to \infty} D(\overline{m}, t) = 0$, which implies that g must be unbounded.

To find the general solution of (FE), we first show that D has the desired form for $0 < m \leq \overline{m}$. By the previous lemma, D solves the functional equation for s, l s.t. $[0 < s \leq l = \overline{m}]$ and all t, and in particular, it solves

$$D(m, t) \cdot D(\overline{m}, \Phi(m, 0)) = D(\overline{m}, \Phi(m, t))$$

for any $0 < m \leq \overline{m}$ and all t. Since we have determined that $D(\overline{m}, t) = e^{-g(t)}$, this functional equation therefore implies

$$D(m, t) = e^{-[g(\Phi(m,t)) - g(\Phi(m,0))]} \quad \text{for all } m \leq \overline{m} \text{ and } t,$$

as desired.

Next consider $m \geq \overline{m}$. By the previous lemma, D must satisfy

$$D(\overline{m}, t) \cdot D(m, \Phi(m, 0)) = D(m, \Phi(m, t)).$$

Then, given the earlier observations, $D(m, \Phi(m, t)) = D(\overline{m}, t) \cdot D(m, \Phi(m, 0)) = e^{-[g(t) + g(\Phi(m,0))]}$. Therefore,

$$D(m, t) = e^{-[g(\Phi_m^{-1}(t)) + g(\Phi(m,0))]} \text{ for all } m > \overline{m} \text{ and } t \geq \Phi_{\overline{m},m}(0).$$

Finally, to consider the case $[m > \overline{m} \text{ and } t < \Phi_{\overline{m},m}(0)]$, we note that by the previous lemma $D(m, t) = \frac{D(m, \Phi_{\overline{m},m}(0))}{D(p_{mt}, \Phi_{\overline{m},p_{mt}}(0))}$ and therefore by our earlier observations,

$$D(m, t) = e^{-[g(\Phi(m,0)) - g(\Phi(p(m,t),0))]} \text{ for } m > \overline{m} \text{ and } t < \Phi_{\overline{m},m}(0).$$

Thus, we have shown that if D is attributable to the preference then it must have the desired form.

To complete the proof, we need to check that the discount function solves (FE). This is straightforward to establish in light of the previous lemma. For instance, for the case where $l = \overline{m}$, we see that

$D(m, t) \cdot D(\overline{m}, \Phi(m, 0)) = D(\overline{m}, \Phi(m, t))$
$\iff e^{-[g(\Phi(m,t)) - g(\Phi(m,0))]} \cdot e^{-[g(\Phi(\overline{m},\Phi(m,0))) - g(\Phi(\overline{m},0))]} = e^{-[g(\Phi(\overline{m},\Phi(m,t))) - g(\Phi(\overline{m},0))]}$
$\iff e^{-[g(\Phi(m,t)) - g(\Phi(m,0)) + g(\Phi(\overline{m},\Phi(m,0))) - g(\Phi(\overline{m},0))]} = e^{-[g(\Phi(\overline{m},\Phi(m,t))) - g(\Phi(\overline{m},0))]}$
$\iff g(\Phi(m, t)) - g(\Phi(m, 0)) + g(\Phi(\overline{m}, \Phi(m, 0))) = g(\Phi(\overline{m}, \Phi(m, t)))$. But $\Phi(\overline{m}, x) := \Phi_{\overline{m},\overline{m}}(x) = x$, and thus the last equation is an identity. $\blacksquare$

C. Appendix: Proof of Theorem 2.3

It follows from the definition of admissibility that $D(t) = e^{-g(t)}$ is admissible for the data $\{\Phi(p_{I1}, t) : t = \tau_2, .., \tau_J\}$ if and only if g solves the functional equation

$$g(\Phi(p_{I1}, t)) = g(t) + g(\Phi(p_{I1}, 0)), \qquad \text{for all } t = \tau_2, .., \tau_J.$$

The set of admissible $D(t) = e^{-g(t)}$ is nonempty (the 'true' one is in the set). Take any admissible D, and corresponding g.

Below we extend the data $\{\Phi(p_{I1}, t) : t = \tau_2, .., \tau_J\}$ to some function Φ on a subset of $X = \mathbb{R}_+^2$ in a way that is consistent with the present value data, and then proceed to prove the theorem. Specifically, we inductively define Φ on $\{p_{ij}\} \times \mathbb{R}_+$. It will be convenient to define, for each $1 \le \iota \le I$, the set $S_\iota \subset \{p_{ij}\}$ of all observed present values of rewards $m_\iota, .., m_I$, that is, $S_\iota := \{p_{ij}: \iota \le i \le I \text{ and } j = 0, .., J\}$. Note that by regularity, $m_i = p_{i0}$.

First consider $\iota = I$. Define $\Phi(m_I, t) = t$ for all t. For all j, define $\Phi(p_{Ij}, 0) = t_j$ and moreover, $\Phi(p_{Ij}, t) = g^{-1}(g(t) + g(\Phi(p_{Ij}, 0)))$.

Next suppose that, for $1 < \iota \le I$,

(a) Φ is defined for $m \in S_\iota$ and all t,

(b) $\Phi(\cdot, 0)$ is strictly increasing on S_ι,

(c) for all $p_{ij} \in S_\iota$,

$$\Phi(p_{ij}, 0) = \Phi(m_i, t_j),$$

(d) for all $m_i \in S_\iota$,

$$g(t_j) = g(\Phi(m_i, t_j)) - g(\Phi(m_i, 0))$$

Observe that this is satisfied for the case $\iota = I$ that we just defined. We now extend Φ to $S_{\iota-1}$ and all t such that these conditions are satisfied.

If $m_{\iota-1}(= p_{\iota-1,0})$ equals some $m \in S_\iota$ then define $\Phi(m_{\iota-1}, t) = \Phi(m, t)$ for all t. If $m_{\iota-1} < S_\iota$ then define $\Phi(m_{\iota-1}, 0)$ by taking any arbitrary number in $(\max_{m \in S_\iota} \Phi(m, 0), \infty)$ and let $\Phi(m_{\iota-1}, t) = g^{-1}(g(t) + g(\Phi(m_{\iota-1}, 0)))$ for all t. If neither of these cases hold, then there exist $m^*, m_* \in S_\iota$ such that m^* is the smallest element in S_ι that is greater than m_ι and m_* is the largest element smaller than it. Define $\Phi(m_{\iota-1}, 0)$ by taking any number in the interval $(\Phi(m^*, 0), \Phi(m_*, 0))$ (regularity and the construction ensures that the interval is nonempty) and let $\Phi(m_{\iota-1}, t) = g^{-1}(g(t) + g(\Phi(m_{\iota-1}, 0)))$ for all t. Next for all j, define $\Phi(p_{\iota-1,j}, 0) = \Phi(m_{\iota-1}, t_j)$ and moreover, $\Phi(p_{\iota-1,j}, t) = g^{-1}(g(t) + g(\Phi(p_{\iota-1,j}, 0)))$. Then Φ is defined on $S_{\iota-1}$ and all t, and moreover, the analogues of (a)-(d) hold by construction. Continue this construction till we obtain Φ on $\{p_{ij}\} \times \mathbb{R}_+$. This satisfies the analogues of (a)-(d) for $\iota = 1$.

To prove the theorem, take the admissible discount function $D(t) = e^{-g(t)}$. By property (b), we can extend $\Phi(\cdot, 0)$ continuous and monotonically to all of $\mathbb{R}_+$ and define the utility index,

$$u(m) = e^{-g(\Phi(m,0))}.$$

Given properties (c) and (d), determine that, for all i, j,

$$u(p_{ij}) = e^{-g(\Phi(p_{ij},0))}$$
$$= e^{-g(\Phi(m_i,t_j))} = e^{-[g(\Phi(m_i,t_j))-g(\Phi(m_i,0))]} \cdot e^{-g(\Phi(m_i,0))}$$
$$= e^{-g(t_j)}u(m_i) = D(t_j)u(m_i).$$

Thus, we have shown that any D that is attributable to Φ data is also attributable to the present value data.

D. Appendix: Proof of Proposition 3.1

Most of the assertions follow from Proposition 3.2. Regularity implies that a must be weakly decreasing: if $m > m'$ then by monotonicity $(m, t) \succ (m', t)$ for all t. The representation implies

$$(m, t) \succ (m', t) \iff e^{-g(t)[a(m)-a(m')]} > \frac{u(m')}{u(m)},$$

but if $a(m) > a(m')$ then, although the inequality holds for $t = 0$ (since $\frac{u(m')}{u(m)} < 1$), the unboundedness of g implies that the above inequality cannot hold for all t, which violates monotonicity. Therefore a is weakly decreasing.

To see that the linearity of g corresponds to that of Φ in t, recall that by Proposition 3.2,

$$g(\Phi(m, t)) - g(\Phi(m, 0)) = a(m) \cdot g(t),$$

for all m, t. If g is linear then $\Phi(m, t) = a(m) \cdot t + \Phi(m, 0)$, a linear function. Conversely if Φ is a linear function $\Phi(m, t) = \alpha(m) \cdot t + \beta(m)$ then clearly linear g and $a(\cdot) = \alpha(\cdot)$ is a solution to the functional equation. We want to show there are no other solutions. Suppose there is a solution g that is twice differentiable. Then using the functional equation we see that

$$g'(\Phi(m, t)) \cdot \Phi''(m, t) = a(m) \left[g''(t) - g''(\Phi(m, t)) \cdot \frac{g'(t)}{g'(\Phi(m, t))} \right].$$

If Φ is linear then $\Phi''(m, t)$ equals 0 and so does the term in the square brackets, that is, for all m, t,

$$\frac{g''(t)}{g'(t)} = \frac{g''(\Phi(m, t))}{g'(\Phi(m, t))}.$$

By varying m between 0 and $\overline{m}$ we see that $\frac{g''(t)}{g'(t)}$ must be a constant, in which case it can either be linear or of the CARA form $g(r) = e^{\kappa \cdot r} - 1$, where $\kappa > 0$ (since g must be unbounded and satisfy $g(0) = 0$). However the CARA form does not solve the functional equation and so g must be linear.

E. Appendix: Proof of Proposition 3.3

We exploit proposition 3.2 which yields that a separable discount function $D(t) = \delta^{g(t)}$ can be attributed if and only if g satisfies:

$$g((a(s)t^\gamma + b(s)^\gamma)^{\frac{1}{\gamma}}) = g(t) + g(b(s)).$$

First suppose a g that satisfies this equation exists. We show that $a(s)$ and $b(s)$ must be linearly related. Take any s'. Letting $t = \Phi(s', 0) = b(s')$ we see that

$$g((a(s)b(s')^\gamma + b(s)^\gamma)^{\frac{1}{\gamma}}) = g(b(s')) + g(b(s)) = g((a(s')b(s)^\gamma + b(s')^\gamma)^{\frac{1}{\gamma}}),$$

and since g is strictly increasing, $a(s)b(s')^\gamma + b(s)^\gamma = a(s')b(s)^\gamma + b(s')^\gamma$, which implies

$$\frac{a(s) - 1}{b(s)^\gamma} = \frac{a(s') - 1}{b(s')^\gamma}.$$

Thus if an g exists, then the ratio $\frac{a(s)-1}{b(s)^\gamma}$ must be a constant k for all s, and so the equation $a(s) = 1 + kb(s)^\gamma$ must hold, as desired. To see that $k \geq 0$, note when $\frac{a(s)-1}{b(s)^\gamma} = k$ for all s, then we have a functional equation $g(((1 + kb(s)^\gamma)t^\gamma + b(s)^\gamma)^{\frac{1}{\gamma}}) = g(t) + g(b(s))$. Denoting $x = t$ and $y = b(s)$ we can write this as:

$$g((x^\gamma + y^\gamma + kx^\gamma y^\gamma)^{\frac{1}{\gamma}}) = g(x) + g(y).$$

Suppose by way of contradiction that $k < 0$. Write $g((x^\gamma + y^\gamma(1 - |k| x^\gamma))^{\frac{1}{\gamma}}) = g(x) + g(y)$ and take x such that $(1 - |k| x^\gamma) < 0$. Then as y increases the LHS decreases and the RHS increases (since g is strictly increasing), a contradiction. Thus $k \geq 0$ must hold.

Conversely, suppose $a(s) = 1 + kb(s)^\gamma$ holds with $k \geq 0$. Consider the above displayed functional equation. Consider various cases:

(i) $k = 0$.
A solution is $g(x) = cx^\gamma$, $c > 0$. Then $D(t) = \delta^t$ is attributable for any $\delta \in (0, 1)$.
(ii) $k > 0$.

38

Then it is easily verified that $g(x) = \ln(1 + kx^\gamma)$ is a solution, and so for any $r > 0$, an attributable discount function is $D(t) = e^{-r \ln(1 + kt^\gamma)} = (1 + kt^\gamma)^{-r}$.

Finally, the uniqueness properties of the SDU model (Fishburn and Rubinstein [1982]) confirm that in each of the above cases there are no other attributable discount functions within the SDU class. ∎

References

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

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

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

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2013. "Discounting Behaviour and the Magnitude Effect: Evidence from a Field Experiment in Denmark." *Economica* 80 (320): 670–97. doi:10.1111/ecca.12028.

Andreoni, J. and C. Sprenger (2012): 'Estimating Time Preferences from Convex Budget Sets', *American Economic Review*, 102(7): 3333-56.

Attema, A., H. Bleichrodt, K. Rohde and P. Wakker (2010): 'Time-Tradeoff Sequences for Analyzing Discounting and Time Inconsistency', *Management Science* 56:11, pp 2015-2030.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2013. "Working Over Time: Dynamic Inconsistency in Real Effort Tasks." NBER Working Paper 18734.

Benhabib, Jess, Alberto Bisin, and Andrew Schotter. 2010. "Present-Bias, Quasi-Hyperbolic Discounting, and Fixed Costs." *Games and Economic Behavior* 69 (2): 205–23. doi:10.1016/j.geb.2009.11.003.

Booij, Adam S., and Bernard M. S. van Praag. 2009. "A Simultaneous Approach to the Estimation of Risk Aversion and the Subjective Time Discount Rate." *Journal of Economic Behavior & Organization* 70 (1–2): 374–88. doi:10.1016/j.jebo.2009.01.005.

Coller, M. and M. Williams (1999): 'Eliciting Individual Discount Rates,' *Experimental Economics* 2, pp 107-127.

Cubitt, Robin P., and Daniel Read. 2007. "Can Intertemporal Choice Experiments Elicit Time Preferences for Consumption?" *Experimental Economics* 10 (4): 369–89. doi:10.1007/s10683-006-9140-2.

Ericson, Keith M. Marzilli, John Myles White, David Laibson, and Jonathan D. Cohen. Forthcoming. "Money Earlier or Later? Simple Heuristics Explain Intertemporal Choices Better than Delay Discounting." *Psychological Science.*

Fishburn, P. (1967): 'Methods of Estimating Additive Utilities', *Management Science* 13(7), pp 435-453.

Fishburn, P. and A. Rubinstein (1982): 'Time Preference', *International Economic Review* 23:3, pp 677-94.

Frederick, 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.

Fudenberg, Drew, and David K. Levine. 2006. "A Dual-Self Model of Impulse Control." The American Economic Review 96 (5): 1449–76.

Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." Econometrica 69 (6): 1403–35. doi:10.1111/1468-0262.00252.

Harrison, G., M. Lau, and M. Williams (2002): 'Estimating Individual Discount Rates in Denmark: A Field experiment,' *American Economic Review* 92(5), pp 1606-1617.

Harrison, G. and E. Rutström (2009): 'Risk Aversion in Experiments', *Research in Experimental Economics* 12, pp 41–196.

Krantz, D.H., R. D. Luce, P. Suppes and A. Tversky (1971): *Foundations of Measurement, Vol. I: Additive and Polynomial Representations*, NY Academic Press.

Laibson, D. (1997): 'Golden Eggs and Hyperbolic Discounting', *Quarterly Journal of Economics* 112, pp 443-77.

Laury, Susan K., Melayne Morgan McInnes, and J. Todd Swarthout. 2012. "Avoiding the Curves: Direct Elicitation of Time Preferences." *Journal of Risk and Uncertainty* 44 (3): 181–217. doi:10.1007/s11166-012-9144-6.

Mazur, J. E. (1987): 'An Adjustment Procedure for Studying Delayed Reinforcement' in M. L. Commons, J. E. Mazur, J. A. Nevin and H. Rachlin (Eds), *Quantitative Analyses of Behavior: Vol. 5. The Effect of Delay and Intervening Events on Reinforcement Value*, pp 55-73, Hillsdale, NJ: Erlbaum.

McClure, Samuel M., David I. Laibson, George Loewenstein, and Jonathan D. Cohen. (2004). "Separate Neural Systems Value Immediate and Delayed Monetary Rewards." Science 306 (5695): 503–7. doi:10.1126/science.1100907.

Noor, J. (2011): 'Intertemporal Choice and the Magnitude Effect', *Games and Economic Behavior* 72(1), pp 255-270.

Olea, J. and T. Strzalecki. (2014). "Axiomatization and measurement of Quasi-hyperbolic Discounting," *Quarterly Journal of Economics* 129 (3), pp. 1449-1499.

Ok, E. and Y. Masatlioglu (2007): 'A Theory of (Relative) Discounting', *Journal of Economic Theory* 137(1), pp 214-245.

Rohde, K. (2008): 'The Hyperbolic Factor: A Measure of Time-Inconsistency', mimeo.

Takeuchi, K. (2010): 'Nonparametric Test of Time-Inconsistency: Present Bias and Future Bias', *Games and Economic Behavior* 71(2), pp 456–478.

Wakker, P. and D. Deneffe (1996): 'Eliciting von Neumann-Morgenstern Utilities when Probabilities are Distorted or Unknown', *Management Science* 42, pp 1131-1150.

Zauberman, Gal, B. Kyu Kim, Selin A Malkoc, and James R Bettman. 2009. "Discounting Time and Time Discounting: Subjective Time Perception and Intertemporal Preferences." *Journal of Marketing Research* 46 (4): 543–56. doi:10.1509/jmkr.46.4.543.