ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

REVEALED TIME-PREFERENCE

Pawel Dziewulski

Number 748
June 2015

Manor Road Building, Oxford OX1 3UQ

Revealed time-preference*

Paweł Dziewulski†

June 2015

Abstract

In this paper we concentrate on the observable implications of the discounted utility model of time-preference. We consider a framework in which subjects are allowed to choose between pairs consisting of a reward and a time-delay at which the prize is delivered. Given a finite set of observations, we are interested in conditions under which choices of an agent can be rationalised by a discounted utility maximisation. We develop an axiomatic characterisation of time-preference with various forms of discounting, including weakly present-biased, quasi-hyperbolic, and exponential, and determine the testable restrictions for each specification. Moreover, we discuss possible identification issues that may arise in this class of tests. Finally, we apply the methods to study the impact of substance abuse on time-preference.

Keywords: revealed preference, testable restrictions, rationalisation, time-preference, discounted utility, hyperbolic discounting, exponential discounting
JEL Classification: C14, C60, C61, D11, D12

1 Introduction

Consider the following experiment. In each trial an agent is asked to choose a pair (x, t), consisting of a reward $x \in \mathbb{R}^\ell_+$ delivered at a time-delay $t \in \mathbb{N}$, from a finite set of options A. Suppose that, in each attempt, both the set A and the actual choice are observable.

*I am especially grateful to John Quah for his invaluable help during writing this paper. Moreover, I would like to thank Alan Beggs, Vince Crawford, Glenn Harrison, Jean-Jacques Herings, Max Kwiek, Josh Lanier, Yusufcan Masatlioglu, Sujoy Mukerji, Drazen Prelec, Krzysztof Pytka, Collin Raymond, Kota Saito, Koji Shirai, and Tomasz Strzalecki for many insightful comments and suggestions.

†Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford OX1 3UQ, United Kingdom. Email: pawel.dziewulski@economics.ox.ac.uk.

Given a finite number of repetitions of the experiment, under what conditions imposed on the set of observations it is possible to rationalise the choices of the agent by a utility maximisation? That is, when there is a function $v : \mathbb{R}^\ell_+ \times \mathbb{N} \to \mathbb{R}$ such that, for any set of feasible options A and the corresponding choice (x, t), we have

$$v(x, t) \geq v(y, s), \text{ for all } (y, s) \in A?$$

Clearly, without any additional restrictions imposed on v, the above problem is trivial, since any constant function satisfies the above condition. For this reason, given our setting, we focus on a class of functions that are strictly increasing with respect to rewards and strictly decreasing in time-delays. In particular, we are interested in preferences that are separable with respect to the two variables. That is, we discuss conditions under which the observable choices of agents can be supported by a *discounted utility* function $v(x, t) := u(x)\gamma(t)$, where $u : \mathbb{R}^\ell_+ \to \mathbb{R}_+$ is an instantaneous utility function, while $\gamma : \mathbb{N} \to (0, 1]$ is a discounting function.

The separable specification of time-preference seems to be particularly important from the economic point of view. The discounted utility model plays a crucial role throughout the economic analysis and is widely accepted as a valid normative standard for public policies, as well as a descriptively accurate representation of the actual behaviour of economic agents. However, in the recent years an important question was raised concerning the form of the discounting function that reflects the actual time-preferences of consumers. In particular, alternative specifications of hyperbolic and quasi-hyperbolic discounting were proposed, which could explain various observations anomalous in the model of exponential discounted utility, formerly dominant in economics. See Frederick, Loewenstein, and O'Donoghue (2002) for a detailed discussion concerning this topic.

We propose an axiomatic characterisation of time-preference in a framework where the domain of choices is restricted to pairs of rewards and time-delays. Our prize-time set-up is similar to the one discussed in Fishburn and Rubinstein (1982), Ok and Masatlioglu (2007), Bleichrodt, Rohde, and Wakker (2009), or Noor (2011). However, unlike those papers, we do not take the preference relation of an agent as a primitive; rather, we assume that the observer can monitor only a finite number of choices made by the consumer. This restriction significantly affects conditions characterising time-preference. Since the observable choices induce only an incomplete preference ordering over the space of prize-time pairs (x, t), the question is how to extend the relation in a way that is consistent with a certain type of a utility function. Whether this is possible or not determines if a given data set can be rationalised by a specific form of time-preference.

The main motivation of this paper is to establish the testable restrictions for various models of inter-temporal choice. In particular, we are interested in conditions that would allow us to demarcate between different specifications of the discounted utility model, including its hyperbolic, quasi-hyperbolic, and exponential representations.

We consider our framework to be relevant from the perspective of empirical applications. There are numerous examples of experiments in which subjects are asked to choose between monetary payments delivered with various time-delays. This includes an extensive list of studies discussed by Frederick, Loewenstein, and O'Donoghue (2002, Table 1), as well as the experiments performed by Kirby and Petry (2004), Chabris, Laibson, Morris, Schuldt, and Taubinsky (2008, 2009), Andersen, Harrison, Lau, and Rutström (2008, 2013, 2014), Benhabib, Bisin, and Schotter (2010), Coller, Harrison, and Rutström (2012), or Dohmen, Falk, Huffman, and Sunde (2012). The design of these studies allows us to apply our results directly to the observations they generate. In fact, in Section 5 we implement our tests to the data from the experiment by Kirby and Petry (2004), in order to study the impact of substance abuse on time-preference.

We begin our discussion in Section 2, where we introduce the notation as well as some preliminary results. Throughout the paper our axiomatic characterisation is imposed on the directly revealed preference relation induced by the set of observations. We say that a pair (x, t) is *directly revealed preferred* to (y, s), whenever there exists at least one observation of the experiment such that both options were available, i.e., they both belonged to the some feasible set A, and (x, t) was chosen. We shall denote $(x, t)\,\mathcal{R}^*(y, s)$. The main difficulty in our problem is to determine a preorder that extends the directly revealed preference relation to the whole domain $\mathbb{R}^\ell_+ \times \mathbb{N}$ of the prize-time pairs. In addition, we need to guarantee that the ordering can be represented by a utility function that possesses the desirable properties, in particular, monotonicity and separability.

First, we present restrictions for a general specification of time-preference. By evoking a spacial case of the *generalised cyclical consistency* axiom discussed in Nishimura, Ok, and Quah (2013), we determine necessary and sufficient conditions under which the set of observations can be rationalised by a utility function $v : \mathbb{R}^\ell_+ \times \mathbb{N} \to \mathbb{R}$, that is strictly increasing with respect to rewards x and strictly decreasing in time-delays t. We show that the set of observations can be rationalised in the above sense, whenever the revealed preference relation combined with the imposed partial ordering on $\mathbb{R}^\ell_+ \times \mathbb{N}$ does not induce any strict cycles over the observable options. Hence, our result is a direct application of the Rationalisability Theorem II, presented in the aforementioned paper.

The first main theorem of this paper is introduced in Section 3, where we provide an axiomatic characterisation of the *discounted utility* model. Since this form of time-preference is a special case of the representation discussed in the previous paragraph, cyclical consistency remains a necessary condition, but it is no longer sufficient. We introduce an alternative restriction called *dominance* axiom. Roughly speaking, the axiom requires that there exists no collection of directly revealed preference relations $(x, t)\,\mathcal{R}^*(y, s)$ in which the distribution of rewards y, appearing in the inferior options, first order stochastically dominates the distribution of prizes x in the preferred pairs, while the distribution of time-delays t first order stochastically dominates the distribution of s.

Our approach to characterisation of time-preference via the notion of stochastic dominance is novel. However, the tools we use to show the necessity and sufficiency of our axioms are similar to those applied in the classical literature on intuitive probability and additive plausibility (see, e.g., Kraft, Pratt, and Seidenberg, 1959 or Scott, 1964). In particular, dominance axiom has a similar flavour to the *cancellation law* used extensively in this area of research. Moreover, our restriction describing the separable formulation of time-preference resembles the condition characterising the expected utility hypothesis, introduced by Border (1992). In his paper, Border concentrates on observable choices over sets of lotteries, and discusses conditions under which they can be rationalised by the expected utility model. The restriction of *ex-ante dominance*, that he proposes, hinges on a specific form of the first order stochastic dominance between the observable and an alternative, hypothetical choice function. Even thought the question we consider, the framework we specify, as well as the tools we apply are substantially different from those used by Border, the main line of our argument is similar.

The second main theorem of this paper, discussed in Section 4, focuses on conditions characterising the *weakly present-biased* specification of the discounting function γ, for which ratio $\gamma(t)/\gamma(t+1)$ is a decreasing function of t. Therefore, under our formulation, the relative discounting between any two subsequent dates diminishes as they become more distant in the future. We consider this class of time-preference to be particularly relevant, as it admits the well-known hyperbolic, quasi-hyperbolic, and exponential specifications of the discounting function.

The condition characterising this class of time-preference is summarised by the *cumulative dominance* axiom. Our restriction requires that there exists no collection of directly revealed relations $(x, t)\,\mathcal{R}^*(y, s)$ such that the distribution of prizes y in the inferior options first order order stochastically dominates the distribution of rewards x, while the distribution of time-delays t *second* order stochastically dominates the distribution of s. The condition is similar to the dominance axiom. However, unlike in the general case of the discounted utility characterisation, it is necessary to consider samples in which the distributions of time-delays are ordered with respect to the second order stochastic dominance. Since the first order stochastic dominance implies the second, the cumulative dominance axiom is more restrictive, as it is requires to verify a larger class of collections from the directly revealed preference relation.

In the second part of Section 4, we draw our attention to an axiomatic characterisations of two specific examples of weakly present-biased discounting functions: quasi-hyperbolic and exponential. The testable implications of the two specifications are similar, however, distinguishable. The essence of the two restrictions is summarised in the *strong cumulative dominance* axiom. Its formulation and intuition is similar to the one of the dominance and cumulative dominance axioms. Analogously to the previous conditions, it focuses on distributions of rewards and time-delays of arbitrary collections drawn from the directly revealed preference set, whilst requiring an even stronger form of stochastic dominance between the distributions.

We are not the first to axiomatise time-preference in a setting with a finite number of observations. Echenique, Imai, and Saito (2014) characterise various forms of the time-separable model of inter-temporal choice in a framework where agents acquire streams of a one-dimensional consumption good. In their setting, an observation consists of a consumption path selected by the subject and prices at which the choice was made. Hence, the set-up resembles the standard Afriat-type revealed preference test for consumption-price data. The authors specify both necessary and sufficient conditions under which the set of observations can be rationalised by different forms of time-separable preference. What is crucial to their result, is the assumption concerning concavity of the instantaneous utility function. Therefore, the test allows to test for only a subclass of discounted utility models. On the other hand, our framework concentrates on the core implications of the discounted utility theory. We dispense the assumptions that are not crucial to the hypothesis and characterise these observable restrictions which are pivotal to this class of models. Nevertheless, as our set-up differs substantially from the one adopted by Echenique, Imai, and Saito, our results are not comparable.

2 Preliminaries

We begin the analysis with a formal specification of our framework. We define the *domain* over which the agents determine their choices by $X \times T$, where $X := \mathbb{R}^{\ell}_{+}$ and $T := \mathbb{N}$. Each element (x, t) of the set consists of a vector of rewards $x \in X$ and a time-delay

$t \in T$ at which the prizes are delivered.

An *experiment* $\mathcal{E}$ is a collection of finite subsets A of $X \times T$. In each trial a subject is asked to choose a single element (x, t) from a set of feasible options A. Therefore, an *observation* is an ordered pair $\big(A, (x, t)\big)$ consisting of a set $A \in \mathcal{E}$ and an option $(x, t) \in A$ that was selected by the agent. Denote the set of observations by $\mathcal{O}$.[1]

Let $\mathcal{A}$ be the set of *observable options*, i.e., of all possible pairs of rewards and time-delays that were offered to the agent at least once during the experiment. Hence,

$$\mathcal{A} := \bigcup_{A \in \mathcal{E}} A.$$

Note that, by construction, $\mathcal{A}$ is a finite subset of $X \times T$. In addition, let the set of *observable rewards* be given by

$$\mathcal{X} := \big\{x \in X : (x, t) \in \mathcal{A}, \text{ for some } t\big\}.$$

Therefore, $\mathcal{X}$ denotes the set of all prizes that were available to the subject at least once during the experiment. Analogously, the set of *observable time-delays* is defined by

$$\mathcal{T} := \big\{t \in T : (x, t) \in \mathcal{A}, \text{ for some } x\big\}.$$

In the following subsection we discuss the key properties of the revealed preference relation induced by the set of observations $\mathcal{O}$.

2.1 Revealed preference relations and mixed-monotonicity

For any elements (x, t) and (y, s) in $\mathcal{A}$, we say that (x, t) is *directly revealed preferred* to (y, s), if in at least one trial of the experiment both options (x, t) and (y, s) were feasible but the agent decided to choose (x, t) rather than (y, s). Formally, we say that the ordered pair $\big((x, t), (y, s)\big)$ belongs to set $\mathcal{R}^*$ defined by

$$\mathcal{R}^* := \Big\{\big((x, t), (y, s)\big) \in \mathcal{A} \times \mathcal{A} : \text{ there exists some } A \in \mathcal{E} \\ \text{such that } (x, t), (y, s) \in A \text{ and } \big(A, (x, t)\big) \in \mathcal{O}\Big\}.$$

[1] Our framework allows agents to make multiple choices from a single set A. However, each such choice has to be considered as a separate trial "with replacement". In fact, our framework does *not* allow for consumers to choose several options simultaneously from a set of feasible options, as the sequence in which the elements are chosen matters. Suppose that an agents chooses two elements (x, t) and (y, s) from some set A "without replacement". Then the corresponding observations are either $\big(A, (x, t)\big)$ and $\big(A \backslash \{(x, t)\}, (y, s)\big)$, or $\big(A, (y, s)\big)$ and $\big(A \backslash \{(y, s)\}, (x, t)\big)$, depending on which option was chosen first.

For convenience, we denote $(x,t)\,\mathcal{R}^*(y,s)$ instead of $\big((x,t),(y,s)\big) \in \mathcal{R}^*$. Moreover, whenever $(x,t)\,\mathcal{R}^*(y,s)$ and $(y,s)\,\mathcal{R}^*(x,t)$, we say that (x,t) is *directly revealed indifferent* to (y,s). Nevertheless, we do *not* define the strict counterpart of $\mathcal{R}^*$.

A *sample* of the directly revealed preference relation is a finite, indexed collection $\big\{\big((x_i,t_i),(y_i,s_i)\big)\big\}_{i\in I}$ of elements in $\mathcal{R}^*$. In particular, we allow for samples to be generated "with replacement". That is, a single element of $\mathcal{R}^*$ may appear more than once in a sample. In addition, for an arbitrary sample $\big\{\big((x_i,t_i),(y_i,s_i)\big)\big\}_{i\in I}$, we define probability distributions μ and ν of rewards in the superior and inferior options by

$$\mu(z) := \tfrac{1}{|I|}\big|\{i\in I : x_i = z\}\big| \quad\text{and}\quad \nu(z) := \tfrac{1}{|I|}\big|\{i\in I : y_i = z\}\big|. \tag{1}$$

Analogously, the corresponding distributions τ and σ of time-delays are given by

$$\tau(z) := \tfrac{1}{|I|}\big|\{i\in I : t_i = z\}\big| \quad\text{and}\quad \sigma(z) := \tfrac{1}{|I|}\big|\{i\in I : s_i = z\}\big|. \tag{2}$$

By construction, the supports of probability distributions μ and ν are contained in $\mathcal{X}$, while distributions τ and σ are defined over $\mathcal{T}$.

The purpose of our analysis is to establish conditions under which the set of observations $\mathcal{O}$ can be rationalised by a specific form of time-preference. Clearly, one of the necessary conditions for rationalisation is existence of a transitive closure of $\mathcal{R}^*$ over $\mathcal{A}$. We say that a complete, transitive, and reflexive preorder $\mathcal{R}\subseteq\mathcal{A}\times\mathcal{A}$ is *consistent* with relation $\mathcal{R}^*$ if for any two (x,t), $(y,s)\in\mathcal{A}$, whenever $(x,t)\,\mathcal{R}^*(y,s)$ then $(x,t)\,\mathcal{R}\,(y,s)$, or equivalently $\mathcal{R}^*\subseteq\mathcal{R}$.[2] We denote the strict counterpart of $\mathcal{R}$ by $\mathcal{P}$. Hence, we write $(x,t)\,\mathcal{P}\,(y,s)$ whenever $(x,t)\,\mathcal{R}\,(y,s)$ and not $(y,s)\,\mathcal{R}\,(x,t)$.

For the purposes of this paper, we concentrate on a specific class of consistent preorders. Let $\succeq$ denote a partial order on $X\times T$ such that, for any of its elements (x,t) and (y,s), we have $(x,t)\succeq(y,s)$ whenever $x\geq y$ and $t\leq s$.[3] Moreover, the relation is strict and denoted by $x\succ y$, if at least one of the above inequalities is strict. Loosely speaking, we say that option (x,t) is greater than (y,s) with respect to $\succeq$, if it offers a greater reward at a shorter delay. A preorder $\mathcal{R}$ is *mixed-monotone*, whenever for any

[2]Recall that, for any set S, a *preorder* (or quasiorder) $\mathcal{R}\subseteq S\times S$ is a binary relation that is both reflexive and transitive. In particular, for any x in S, we have $x\,\mathcal{R}\,x$, while for any x, y, and z belonging to S, whenever $x\,\mathcal{R}\,y$ and $y\,\mathcal{R}\,z$ then $x\,\mathcal{R}\,z$.

[3]For an arbitrary set S, a *partial order* $\succeq\subseteq S\times S$ is a preorder satisfying antisymmetry. That is, for any x, y in S, whenever $x\succeq y$ and $y\succeq x$ then $x=y$. Throughout the paper assume that $\mathbb{R}^\ell_+$ is endowed with the coordinate-wise partial order $\geq$. In particular, for any x, y in $\mathbb{R}^\ell_+$ we say that the relation is strict, i.e., $x>y$, whenever $x\geq y$ and $x\neq y$.

two elements (x,t) and (y,s) in $\mathcal{A}$, $(x,t) \succeq (y,s)$ implies $(x,t)\,\mathcal{R}\,(y,s)$, while $x \succ y$ implies $(x,t)\,\mathcal{P}\,(y,s)$. The definition suggests that whenever an agent is presented with two options such that one of them has a (weakly) greater reward and a (weakly) shorter delay than the other one, then the former option should be preferred. Clearly, not every set of observations admits a consistent mixed-monotone preorder. In the following section we discuss conditions under which there exists such an extension.

2.2 Mixed-monotone rationalisation

Set $\mathcal{O}$ is *rationalisable* if there exists a function $v : X \times T \to \mathbb{R}$, strictly increasing with respect to $\succeq$,[4] such that for any $\big(A, (x,t)\big) \in \mathcal{O}$, we have

$$v(x,t) \geq v(y,s), \text{ for all } (y,s) \in A.$$

In the remainder of the paper we focus on conditions under which the set of observations can be rationalised by the above class of time-preference.

Axiom 1 (Cyclical consistency). *For an arbitrary sequence* $\big\{(x_i,t_i)\big\}_{i=1}^{n}$ *belonging to* $\mathcal{A}$ *such that* $(x_{i+1},t_{i+1})\,\mathcal{R}^*(x_i,t_i)$ *or* $(x_{i+1},t_{i+1}) \succeq (x_i,t_i)$*, for all* $i = 1,\dots,n-1$*, and* $(x_1,t_1) \succeq (x_n,t_n)$*, we have* $(x_1,t_1) = (x_n,t_n)$*.*

The above axiom is a special case of the *generalised cyclical consistency* condition formulated by Nishimura, Ok, and Quah (2013). It requires that, for any sequence of observable options such that every subsequent element is directly revealed preferred to or greater (with respect to $\succeq$) than the preceding one, its first element cannot be strictly greater than the ultimate one. Clearly, a violation of this condition implies the existence of a strict cycle of revealed preference, which cannot be reconciled with a mixed-monotone preorder on $\mathcal{A}$. Hence, the requirement stated in the axiom is necessary for rationalisation. Moreover, as we show in the proposition below, cyclical consistency is also sufficient.

Proposition 1. *The following statements are equivalent.*

(i) *Set* $\mathcal{O}$ *is rationalisable.*

(ii) *Set* $\mathcal{O}$ *obeys cyclical consistency.*

(iii) *There exists a mixed-monotone preorder* $\mathcal{R}$ *on* $\mathcal{A}$ *consistent with* $\mathcal{R}^*$*.*

(iv) *There exists a sequence* $\{v_{(x,t)}\}_{(x,t)\in\mathcal{A}}$ *of real numbers such that* $(x,t)\,\mathcal{R}^*\,(y,s)$ *implies* $v_{(x,t)} \geq v_{(y,s)}$*, while* $(x,t) \succ (y,s)$ *implies* $v_{(x,t)} > v_{(y,s)}$*.*

[4] That is, for any (x,t), $(y,s) \in X \times T$, whenever $(x,t) \succ (y,s)$ then $v(x,t) > v(y,s)$.

Proof. Implication (i) ⇒ (ii) ⇒ (iii) ⇒ (iv) follows directly from Nishimura, Ok, and Quah (2013, Rationalizability Theorem II). To show (iv) ⇒ (i), take any sequence $\{v_{(x,t)}\}_{(x,t)\in\mathcal{A}}$ satisfying the property stated in the proposition and define function $v : \mathcal{A} \to \mathbb{R}$ by $v(x,t) := v_{(x,t)}$. Clearly, it strictly increasing with respect to $\succeq$, while $(x,t)\,\mathcal{R}^*(y,s)$ implies $v(x,t) \geq v(y,s)$. Finally, by Husseinov (2010, Corollary 3), there exists a strictly increasing extension of v to $X \times T$ that rationalises $\mathcal{O}$. □

Proposition 1 is a special case of the result by Nishimura, Ok, and Quah (2013, Rationalisability Theorem II), who establish the necessity and sufficiency of the generalised cyclical consistency condition for the existence of a utility function rationalising the choice data in a general class of partially ordered spaces. In particular, the above proposition implies that any set of observations satisfying the cyclical consistency axiom is rationalisable. In addition, statement (iii) guarantees that in order to verify the condition, it is both necessary and sufficient to determine the existence of a mixed-monotone preorder consistent with the directly revealed preference relation.

Example 1. Let $X = \mathbb{R}_+$ and $T = \mathbb{N}$. Suppose that the set of directly revealed preference relations induced by the set of observations contains

$$(20,3)\,\mathcal{R}^*(10,0),\ \ (5,0)\,\mathcal{R}^*(15,3),\ \ (15,0)\,\mathcal{R}^*(20,1),\ \text{ and }\ (10,2)\,\mathcal{R}^*(5,0).$$

Given Proposition 1, in order to verify whether the above set of observations satisfies the cyclical consistency axiom, it is both necessary and sufficient to propose a consistent, mixed-monotone relation $\mathcal{R}$ defined over the set of observable options. For example,

$$(15,0)\;\mathcal{R}\;(20,1)\;\mathcal{P}\;(20,3)\;\mathcal{R}\;(10,0)\;\mathcal{P}\;(10,2)\;\mathcal{R}\;(5,0)\;\mathcal{R}\;(15,3).$$

Clearly, the above relation is both consistent with $\mathcal{R}^*$ and mixed-monotone.

In general, referring to the definition of cyclical consistency in order to determine whether a set of observations is rationalisable is inconvenient for empirical applications. However, in Proposition 1(iv) we propose an equivalent condition which reduces the test to determining the existence of a solution to a system of linear inequalities. We exploit this observation in Section 5, where we implement our methods to the real-life data.

3 Discounted utility rationalisation

In this section we present the first main theorem of this paper. We say that set $\mathcal{O}$ is rationalisable by a *discounted utility* function whenever there is a strictly increasing instantaneous utility function $u : X \to \mathbb{R}_+$ and a strictly decreasing discounting function $\gamma : T \to (0, 1]$, where $\gamma(0) = 1$, such that $v(x, t) := u(x)\gamma(t)$ rationalises the set of observations $\mathcal{O}$. Clearly, cyclical consistency is a necessary condition for this form of representation. However, it is no longer sufficient.

3.1 Dominance axiom

This subsection is devoted to determining both necessary and sufficient conditions on the set of observations that allow for a discounted utility representation. Before stating our first axiom, recall that for any sample $\big\{\big((x_i, t_i), (y_i, s_i)\big)\big\}_{i\in I}$ of $\mathcal{R}^*$, the distributions of rewards in the superior and inferior options are denoted respectively by μ and ν, while the distributions of time-delays are given by τ and σ, as in (1) and (2) in Section 2.1.

Axiom 2 (Dominance)**.** *For any sample of $\mathcal{R}^*$, if distribution ν first order stochastically dominates μ, while τ first order stochastically dominates σ, then $\mu = \nu$ and $\tau = \sigma$.*[5]

The above axiom requires that whenever there exists a sample $\big\{\big((x_i, t_i), (y_i, s_i)\big)\big\}_{i\in I}$ such that the distribution of rewards y_i in the inferior options first order stochastically dominates the distribution of prizes x_i in the preferred options, while the distribution of time-delays t_i appearing on the left hand side of $\mathcal{R}^*$ first order stochastically dominates the distribution of s_i, then both distributions of monetary payments and time-delays have to be equal. In other words, it excludes the existence of a sample in which the distribution of y_i stochastically dominates the distribution of x_i, the distribution of t_i stochastically dominates the distribution of s_i, and at least one of the two relations is strict. To make our presentation more transparent, we discuss the following example.

Example 2. Consider the directly revealed preference relation analysed in Example 1. We claim that the set of observations inducing the relation fails to satisfy the dominance axiom. In order to show this, take $\mathcal{R}^*$. Clearly, the set is a sample of itself. Note that, given the support $\big\{5, 10, 15, 20\big\}$, the corresponding distributions of rewards are $\mu = \nu = \big(1/4, 1/4, 1/4, 1/4\big)$. Similarly, given the support $\big\{0, 1, 2, 3\big\}$, the distributions of time-delays are $\tau = \big(1/2, 0, 1/4, 1/4\big)$ and $\sigma = \big(1/2, 1/4, 0, 1/4\big)$. Therefore, there is a sample of $\mathcal{R}^*$ such the distributions of prizes μ and ν are equal, while τ *strictly* first order stochastically dominates σ. Hence, the set of observations $\mathcal{O}$ violates the axiom.

[5] Let Δ_S denote the set of probability measures over an arbitrary partially ordered set S. For any distributions μ and ν in Δ_S we say that μ first order stochastically dominates ν, if for any increasing function $f : S \to \mathbb{R}$, we have $\int_S f d\mu \geq \int_S f d\nu$. See also Section A of the appendix.

The above example indicates that dominance is a stronger condition than cyclical consistency, as there exists a set of observations satisfying the latter but failing the former. Interestingly, the dominance axiom is both a necessary and sufficient condition for the set of observations $\mathcal{O}$ to be rationalisable by a discounted utility function.

Theorem 1. *Set of observations $\mathcal{O}$ is rationalisable by a discounted utility function if and only if it obeys the dominance axiom.*

In order to show the necessity part of the above theorem, suppose that $\mathcal{O}$ is rationalisable by a discounted utility function $v(x,t) := u(x)\gamma(t)$. This implies that the set is at the same time rationalisable by function $w(x,t) := \phi(x) + \varphi(t)$, where $\phi := \log(u)$ and $\varphi := \log(\gamma)$. Moreover, under this transformation functions ϕ and φ preserve the strict monotonicity of u and γ respectively.

Take any sample $\big\{\big((x_i,t_i),(y_i,s_i)\big)\big\}_{i\in I}$ of the directly revealed preference relation $\mathcal{R}^*$. For any element $\big((x_i,t_i),(y_i,s_i)\big)$ of the sample, we have $\phi(x_i)+\varphi(t_i) \geq \phi(y_i)+\varphi(s_i)$. In particular, once we sum up all the inequalities with respect to $i \in I$, this implies that

$$\sum_{x\in\mathcal{X}} \phi(x)\mu(x) + \sum_{t\in\mathcal{T}} \varphi(t)\tau(t) \geq \sum_{x\in\mathcal{X}} \phi(x)\nu(x) + \sum_{t\in\mathcal{T}} \varphi(t)\sigma(t), \tag{3}$$

where distributions μ, ν, τ, and σ are defined as in (1) and (2) in Section 2.1. Observe that, given strict monotonicity of ϕ, whenever ν first order stochastically dominates μ, then $\sum_{x\in\mathcal{X}}\phi(x)\mu(x) \leq \sum_{x\in\mathcal{X}}\phi(x)\nu(x)$. Analogously, as function φ is strictly decreasing, distribution τ dominates σ only if $\sum_{t\in\mathcal{T}}\varphi(t)\tau(t) \leq \sum_{t\in\mathcal{T}}\varphi(t)\sigma(t)$. However, by Corollary A.1 (see Section A of the appendix) the two inequalities are consistent with condition (3) only if the corresponding distributions of rewards and time-delays are equivalent.

The above argument highlights the form of consistency that is expected from a discounted utility maximiser. Notice that, for any sample $\big\{\big((x_i,t_i),(y_i,s_i)\big)\big\}_{i\in I}$ of $\mathcal{R}^*$, it is possible to construct two lotteries over the observable rewards and time-delays, where the first one assigns probability $\mu(x)\tau(t)$ to a pair (x,t), while the second one assigns probability $\nu(x)\sigma(t)$ to the corresponding option. Therefore, in both lotteries, the rewards and time-delays are drawn independently from the appropriate distributions.

Whenever set $\mathcal{O}$ is rationalisable by the discounted utility model, and hence by some additively separable function $w(x,t) := \phi(x) + \varphi(t)$, condition (3) requires that the expected utility of the first lottery is greater than the expected utility of the second. In other words, an expected utility maximiser with the Bernoulli utility function w would always prefer the lottery defined over the preferred options. On the other hand, due to monotonicity of the utility function, any violation of the dominance axiom would imply that the consumer would rather choose the second gamble. However, such behaviour could not be reconciled with the discounted utility maximisation. Finally, the fact that in each lottery the rewards and time-delays are draw from independent distributions highlights the form of separability that is required by this form of rationalisation.

Proving that the dominance axiom is a sufficient condition for the discounted utility rationalisation is more demanding. Hence, we place the argument in Section B of the appendix. Nevertheless, below we present the main observation used in our argument.

Proposition 2. *Set $\mathcal{O}$ obeys the dominance axiom if and only if there exist numbers $\{\phi_x\}_{x\in\mathcal{X}}$ and $\{\varphi_t\}_{t\in\mathcal{T}}$ such that (i) $(x,t)\,\mathcal{R}^*(y,s)$ implies $\phi_x + \varphi_t \geq \phi_y + \varphi_s$, (ii) $x > y$ implies $\phi_x > \phi_y$, and (iii) $t > s$ implies $\varphi_t < \varphi_s$.*

The above result is a direct implication of Lemmas B.1 and B.2 in Section B of the appendix, as well as the necessity of the dominance axiom for rationalisation by a discounted utility function. To support the above proposition we apply a variation of Farkas' Lemma, commonly known as *Motzkin's Rational Transposition*. Using the result, we show that the system of inequalities implied by the directly revealed preference relation fails to have a solution if and only if the set of observations violates the dominance axiom.

It is worth pointing out the importance of Proposition 2 for the applicability of Theorem 1. The result presents an alternative way of verifying whether the set of observations obeys the dominance axiom. In fact, the proposition states that the axiom is equivalent to the existence of a solution to a system of linear inequalities. Since such systems are in general solvable, i.e., there exist finite algorithms which allow to determine whether an arbitrary system has a solution or not, we find the alternative method of verifying the axiom to be much more convenient. In fact, we exploit the result in Section 5.

Theorem 1 establishes both the testable implications of a mixed-monotone and separable preferences. As it was already shown in Proposition 1, mixed-monotonicity is a testable property. Therefore, given our framework, an important question is whether it is also possible to test *solely* for separability of the utility function. Unfortunately, given our framework, the answer is negative. This follows directly from the fact that in the discussed class of experiments the set of observations does not induce the *strict* directly revealed preference relations. Therefore, any separable function $v(x,t) := u(x)\gamma(t)$, where both u and γ are constant, would trivially rationalises any set of observations.

In the remaining sections of this paper we determine conditions under which the class of preferences rationalising the data can be characterised more precisely. In particular, we focus on finer restrictions imposed on the form of the discounting function γ. Before we proceed with our analysis, we discuss a family of experiments for which a narrower specification of time-preference is *never* possible.

3.2 Anchored experiments and identification

Throughout this subsection we assume that $X = \mathbb{R}_+$. We say that an experiment $\mathcal{E}$ is *anchored*, whenever it consists solely of *doubleton* sets of feasible options A, and there is some (x^*, t^*) in $X \times T$ such that $(x^*, t^*) \in A$, for all $A \in \mathcal{E}$. In other words, in each trial of an anchored experiment the subjects are asked to choose between one fixed option (x^*, t^*) and an alternative element in $X \times T$.

There are several notable examples of anchored experiments that were performed in the literature, including Pender (1996),[6] Coller and Williams (1999), Harrison, Lau, and Williams (2002), or Coller, Harrison, and Rutström (2012). Therefore, we consider this class of experiments to be relevant from the empirical point of view. Moreover, a significant advantage of these tests is that the choices the subjects face are relatively simple, which minimises the chance of errors made by the agents. Nevertheless, as we show in the following result, the simplicity of the experiment substantially reduces the informativeness of observations it generates.

Proposition 3 (Indeterminacy)**.** *For any set of observations $\mathcal{O}$ generated in an anchored experiment, the following statements are equivalent.*

(i) Set $\mathcal{O}$ is rationalisable.

(ii) Set $\mathcal{O}$ is rationalisable by a discounted utility function.

(iii) For any discounting function $\gamma : T \to (0, 1]$, there is a strictly increasing function $u : X \to \mathbb{R}_+$ such that $v(x, t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$.

We present the proof of Proposition 3 in Section B of the appendix. The above result states that, given observations from an anchored experiment, one can only determine if the choices of the subject are rationalisable according to the definition in Section 2.2. Therefore, by design, anchored experiments do not allow to verify whether the observable choices can be rationalised by a narrower class of preferences. In particular, Proposition 3(iii) implies that once the set of observations is rationalisable, it can also be rationalised by virtually *any* form of discounting. Hence, we consider this class of experiments to be

[6]In case of Pender (1996), we refer to the follow-up study conducted in 1991.

rather weak, as the data they produce do not allow for a conclusive specification of the class of time-preference explaining the observable choices.[7]

The above result fails to hold once we allow for a larger variation in prizes and/or time-delays between different options. In particular, as shown in Example 2, whenever each set A consists of two elements (x, t^*) and (y, s), where t^* is fixed across all the sets, while x, y, and s may vary, there exist sets of observations which are rationalisable by a mixed-monotone function, but not by the discounted utility model. Moreover, as shown in Section 5, the larger variation across options allows for a more conclusive demarcation between different forms of the discounted utility model.

Finally, note that the result in Proposition 3(iii) states that for any discounting function γ there exists a *strictly increasing* function u such that $u(x)\gamma(t)$ rationalises the set of observations. In fact the result is no longer true if we restrict our attention to a narrower class of instantaneous utility functions, e.g., concave.

4 Weakly present-biased rationalisation

In the previous section we established the necessary and sufficient conditions under which the set of observations can be rationalised by a discounted utility function. We devote the remainder of the paper to determine restrictions that allow for a narrower characterisation of the discounting function γ. In particular, we focus on a class of preferences that exhibit some degree of present-bias. We say that a strictly decreasing discounting function $\gamma : T \to (0, 1]$, with $\gamma(0) = 1$, is *weakly present-biased*, whenever function $\vartheta : T \to \mathbb{R}_{++}$,

$$\vartheta(t) := \frac{\gamma(t)}{\gamma(t+1)},$$

is decreasing. In other words, we require that the relative discounting between any two subsequent periods decreases as the two dates are further away in the future. Equivalently, this is to say that there exists a *log-convex* extension of function γ to the domain of the real numbers. Our interest in the above class of functions is primarily justified by the fact

[7]It worth pointing that experiments by Andersen, Harrison, Lau, and Rutström (2008, 2014) show some resemblance to an anchored experiment. However, there is one significant variation which makes them distinct from the class of tests discussed in this section. In particular, the experiments consist of two parts. First, the subjects are asked to choose between different lotteries. Then, in the second stage, the agents participate in a standard anchored experiment, described as above. Interestingly, depending on restrictions imposed on preferences over risky outcomes, the design of the experiments allows for a more precise elicitation of time-preference, as long as we consider the observations from both stages of the experiment jointly. Hence, Proposition 3 does not apply to this class of tests.

that it contains the most commonly used forms of discounting, including the hyperbolic, quasi-hyperbolic, and exponential specifications.

4.1 Cumulative dominance axiom

In this section we present the second main result of this paper. Set $\mathcal{O}$ is rationalisable by a *weakly present-biased discounted* utility function, whenever there exists a strictly increasing instantaneous utility function $u : X \to \mathbb{R}_+$ and a weakly present-biased discounting function $\gamma : T \to (0, 1]$ such that $v(x, t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$. Before we state the next axiom, recall that for any sample $\big\{\big((x_i, t_i), (y_i, s_i)\big)\big\}_{i\in I}$ of $\mathcal{R}^*$, we define probability distributions μ, ν, τ, and σ as in (1) and (2) in Section 2.1.

Axiom 3 (Cumulative dominance)**.** *For any sample of $\mathcal{R}^*$, if probability distribution ν first order stochastically dominates μ, while τ second order stochastically dominates σ, then $\mu = \nu$ and σ is a mean-preserving spread of τ.*[8]

The cumulative dominance axiom requires that for any a sample $\big\{\big((x_i, t_i), (y_i, s_i)\big)\big\}_{i\in I}$ for which the distribution of rewards y_i in the inferior options first order stochastically dominates the distribution of prizes x_i in the preferred prize-time pairs, while the distribution of time-delays t_i appearing on the left hand side of $\mathcal{R}^*$ second order stochastically dominates the distribution of s_i, then the distributions of monetary payments have to be equal, while the distribution of the delays in the inferior options must be a mean-preserving spread of its counterpart over the preferred pairs.

Note that the cumulative dominance axiom is a stronger requirement than the dominance axiom. Suppose that the set of observations satisfies the former condition. Then, there exists no sample of the directly revealed preference relation such that the rewards in the inferior options first order stochastically dominate the prizes in the preferred pairs, while the superior time-delays second order stochastically dominate the delays appearing on the right hand side of the relation. Since first order stochastic dominance implies the second, at the same time there exists no sample such the latter relation is preserved under the first order stochastic dominance. Hence, the set of observations must also satisfy the dominance axiom. However, the opposite implication does not hold, as there might exist a collection of elements in $\mathcal{R}^*$ such that the corresponding distributions of time-delays

[8] Let Δ_S be a set of probability measures over an arbitrary set $S \subseteq \mathbb{R}^\ell$. For any μ, ν in Δ_S, we say that ν is a *mean-preserving spread* of μ, if for any concave function $f : \mathbb{R}^\ell \to \mathbb{R}$, we have $\int_S f d\mu \geq \int_S f d\nu$. Moreover, whenever the above condition holds for any concave and increasing function f, we say that μ *second order stochastically dominates* ν. See Section A of the appendix.

are not ordered with respect to the first order stochastic dominance, but are ordered with respect to the second. In other words, the cumulative dominance axiom requires verifying a larger number of samples than the dominance axiom.

Theorem 2. *Set $\mathcal{O}$ is rationalisable by a weakly present-biased discounted utility function if and only if it obeys the cumulative dominance axiom.*

In order to show the necessity of the cumulative dominance axiom for the more stringent form of rationalisation, suppose that choices of an agent can be explained by some function $v(x,t) := u(x)\gamma(t)$, where u is strictly increasing, while γ is strictly decreasing and weakly present-biased. Clearly, one can rationalise the set of observations by function $w(x,t) := \phi(x) + \varphi(t)$, with $\phi := \log(u)$ and $\varphi := \log(\gamma)$. Moreover, the transformation preserves monotonicity of the two functions, while φ has a convex extension to $\mathbb{R}_+$.

By construction, for any element of a sample $\big\{\big((x_i,t_i),(y_i,s_i)\big)\big\}_{i\in I}$ of $\mathcal{R}^*$, we have $\phi(x_i) + \varphi(t_i) \geq \phi(y_i) + \varphi(s_i)$. In particular, once we sum up all the conditions with respect to $i \in I$, this implies that

$$\sum_{x\in\mathcal{X}} \phi(x)\mu(x) + \sum_{t\in\mathcal{T}} \varphi(t)\tau(t) \geq \sum_{x\in\mathcal{X}} \phi(x)\nu(x) + \sum_{t\in\mathcal{T}} \varphi(t)\sigma(t),$$

where distributions μ, ν, τ, and σ are defined as in (1) and (2) in Section 2.1. Suppose that the sample is specified as in the definition of the cumulative dominance axiom. As ϕ is strictly increasing, we have $\sum_{x\in\mathcal{X}} \phi(x)\mu(x) \leq \sum_{x\in\mathcal{X}} \phi(x)\nu(x)$. Moreover, since the distribution of time-delays t second order stochastically dominates the distribution of s, the existence of a convex and strictly decreasing extension of φ to $\mathbb{R}_+$ implies that $\sum_{t\in\mathcal{T}} \varphi(t)\tau(t) \leq \sum_{t\in\mathcal{T}} \varphi(t)\sigma(t)$. However, the two inequalities can be consistent with the initial condition only if they are satisfied with equality. By Corollaries A.1 and A.2 (see Section A of the appendix), this requires for the distributions of rewards to be equivalent, while σ has to be a mean-preserving spread of τ.

The form of consistency that has to be satisfied by a weakly present-biased discounted utility maximiser is similar to the one discussed in Section 3. Take any sample $\big\{\big((x_i,t_i),(y_i,s_i)\big)\big\}_{i\in I}$ of $\mathcal{R}^*$ and construct two lotteries over the observable rewards and time-delays, where the first one assigns probability $\mu(x)\tau(t)$ to a pair (x,t), while the second one assigns probability $\nu(x)\sigma(t)$ to the corresponding option. Observe that, any expected utility maximiser with a separable and mixed-monotone Bernoulli utility function $w(x,t) := \phi(x) + \varphi(t)$ should prefer the former gamble to the latter. However, given that φ is additionally "convex", whenever the cumulative axiom is violated, it would be possible to construct a sample such that the lottery over the inferior options would be preferred to the gamble over the superior prize-time pairs, which would violate the form of consistency required by the weakly present-biased discounted utility maximisation.

The sufficiency part of the proof of Theorem 2 is more demanding. Similarly as in the case of the discounted utility model, our argument consists of two steps. First, in Lemma B.4 (see Section B of the appendix) we show that once the set of observations satisfies the cumulative dominance axiom, there exists a solution to a specific system of linear inequalities. In the second step, see Lemma B.5, we use the solution to construct an instantaneous utility function u and a weakly present-biased discounting function γ that rationalise the data. We present our key observation in the following proposition.

Proposition 4. *Set of observations $\mathcal{O}$ obeys the cumulative dominance axiom if and only if there exist numbers $\{\phi_x\}_{x\in\mathcal{X}}$, $\{\varphi_t\}_{t\in\mathcal{T}}$, and $\{v_t\}_{t\in\mathcal{T}}$ such that (i) $(x,t)\mathcal{R}^*(y,s)$ implies $\phi_x+\varphi_t \geq \phi_y+\varphi_s$, (ii) $x>y$ implies $\phi_x>\phi_y$, (iii) $v_t<0$, for all $t\in\mathcal{T}$, and (iv) for all $t\in\mathcal{T}$, we have $\varphi_t+v_t(s-t)\leq\varphi_s$, for all $s\in\mathcal{T}$.*

Once again, the above proposition is relevant for the applicability of our main result, as it presents an alternative method of verifying the cumulative dominance axiom.

4.2 Quasi-hyperbolic discounting

In the following section we concentrate on quasi-hyperbolic discounting functions, which induce a narrower class of weakly present-biased preferences. We say that the set of observations is rationalisable by a *quasi-hyperbolic discounted utility* function, whenever there exist a strictly increasing function $u : X \to \mathbb{R}_+$, numbers β, $\delta \in (0,1)$, and some time-delay $k \in T$ such that $v(x,t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$, where

$$\gamma(t) := \begin{cases} \beta^t\delta^t & \text{for } t<k, \\ \beta^k\delta^t & \text{otherwise.} \end{cases}$$

Note that our definition of a quasi-hyperbolic discounting function generalises the standard notion for which $k=1$. By allowing for the threshold time-delay k to vary, we are able to analyse a wider class of preferences. In particular, as k denotes a time-delay which separates the dates perceived by the agent as "present" from those regarded as "future", we allow in our test for this parameter to be determined endogenously.

Since in the case of a quasi-hyperbolic discounting function $\vartheta(t) := \gamma(t)/\gamma(t+1)$ takes the value of $(\beta\delta)^{-1}$, for $t<k-1$, and δ^{-1} otherwise, the quasi-hyperbolic specification is weakly present-biased. Hence, any set of observations rationalisable by this more specific form of time-preference obeys the cumulative dominance axiom. However, the condition

is no longer sufficient. In this section we discuss the requirement that fully characterises this class of discounted utility models. As in the previous sections, we define probability distributions μ, ν, τ, and σ as in (1) and (2) in Section 2.1.

Axiom 4 (Strong cumulative dominance)**.** *There exists some $l \in \mathcal{T}$ such that, for an arbitrary sample of $\mathcal{R}^*$, whenever*

(i) distribution ν first order stochastically dominates μ;

(ii) $\sum_{t\in\mathcal{T}} t\tau(t) \geq \sum_{t\in\mathcal{T}} t\sigma(t)$;

(iii) $\sum_{t\in\mathcal{T}} \min\{t,l\}\tau(t) \geq \sum_{t\in\mathcal{T}} \min\{t,l\}\sigma(t)$,

then $\mu = \nu$ and conditions (ii) and (iii) hold with equality.

The above requirement is stronger than cumulative dominance. Clearly, in order to verify whether the above axiom is not violated, we need to consider a wider class of samples of the directly revealed preference relation. As previously, the samples of interest need to satisfy condition (i). However, additionally, they have to obey restrictions (ii) and (iii), which impose a weaker requirement on the relation between the distribution of time-delays appearing on both sides of the directly revealed preference relation in the sample. Therefore, in order to verify the condition, it is necessary to examine a larger set of samples of $\mathcal{R}^*$ then in the case of the cumulative dominance axiom.

Proposition 5. *Set $\mathcal{O}$ is rationalisable by a quasi-hyperbolic discounted utility function if and only if it obeys the strong cumulative dominance axiom.*

The necessity of the axiom can be proven similarly as in the previous sections. Suppose that the set of observations is rationalisable by a quasi-hyperbolic discounted utility function $v(x,t) := u(x)\gamma(t)$, where γ is specified as at the beginning of this section for some β, δ in $(0,1)$, and a time-delay k. This implies that the set of observations is also rationalisable by function $w(x,t) := \phi(x) + \min\{t,k\}\hat{\beta} + t\hat{\delta}$, where $\phi := \log(u)$, $\hat{\beta} := \log(\beta)$, and $\hat{\delta} := \log(\delta)$. By construction, for any sample of $\mathcal{R}^*$, we have

$$\sum_{x\in\mathcal{X}} \phi(x)\mu(x) + \hat{\beta}\sum_{t\in\mathcal{T}} \min\{t,k\}\tau(t) + \hat{\delta}\sum_{t\in\mathcal{T}} t\tau(t) \geq \sum_{x\in\mathcal{X}} \phi(x)\nu(x) + \hat{\beta}\sum_{t\in\mathcal{T}} \min\{t,k\}\sigma(t) + \hat{\delta}\sum_{t\in\mathcal{T}} t\sigma(t),$$

where distributions μ, ν, τ, and σ are specified as in (1) and (2) in Section 2.1. Whenever the sample satisfies conditions (i)-(iii) of the strong cumulative dominance axiom, then monotonicity of function ϕ implies $\sum_{x\in\mathcal{X}} \phi(x)\mu(x) \leq \sum_{x\in\mathcal{X}} \phi(x)\nu(x)$, while

$\hat{\beta}\sum_{t\in\mathcal{T}}\min\{t,k\}\tau(t) \leq \hat{\beta}\sum_{t\in\mathcal{T}}\min\{t,k\}\sigma(t)$ and $\hat{\delta}\sum_{t\in\mathcal{T}}t\tau(t) \leq \hat{\delta}\sum_{t\in\mathcal{T}}t\sigma(t)$, since $\hat{\beta}$ and $\hat{\delta}$ are strictly negative. However, these conditions can be consistent with the initial inequality only if they all hold with equality. Moreover, the argument remains unchanged once we substitute k with $l := \min\{t \in \mathcal{T} : t \geq k\}$.

The sufficiency part of the proof is presented in Section B of the appendix. Nevertheless, we need to point out that the argument is constructed around one important observation which we summarise in the following proposition.

Proposition 6. *Set $\mathcal{O}$ obeys the strong cumulative dominance axiom for some $l \in \mathcal{T}$ if and only if there exist numbers $\{\phi_x\}_{x\in\mathcal{X}}$ and $\hat{\beta}$, $\hat{\delta}$ such that (i) $(x,t)\,\mathcal{R}^*\,(y,s)$ implies $\phi_x + \min\{t,l\}\hat{\beta} + t\hat{\delta} \geq \phi_y + \min\{s,l\}\hat{\beta} + s\hat{\delta}$, (ii) $x > y$ implies $\phi(x) > \phi(y)$, and (iii) $\hat{\beta}$ and $\hat{\delta}$ are strictly negative.*

Proposition 5 requires some comment. First of all, observe that the value of the time-delay l for which set $\mathcal{O}$ obeys the axiom, determines the empirical "kink" of the quasi-hyperbolic discounting function. In particular, this means that we do not assume prior to the test the threshold date which separates the perceived "present" from the "future", but determine it endogenously. In fact, it is possible that one set of observations admits various forms of quasi-hyperbolic discounting, not only with respect to the values of the discount factors β and δ, but also with respect to the pivotal time-delay l.

Second of all, Proposition 6 proposes an alternative method of verifying whether the set of observations obeys the axiom. As in Propositions 2 and 4, it hinges on the existence of a solution to a system of linear inequalities, conditional on l. Since the set of the observable time-delays is finite, the test can be performed in a finite number of steps.

4.3 Exponential discounting

Finally, we draw our attention to the model of exponential discounting. We say that set $\mathcal{O}$ is rationalisable by an *exponential discounted utility* function whenever there is a strictly increasing instantaneous utility function $u : X \to \mathbb{R}_+$ and some $\delta \in (0,1)$ such that $v(x,t) := \delta^t u(x)$ rationalises the set of observations. As previously, for an arbitrary sample of the directly revealed preference relation, we define probability distributions μ, ν, τ, and σ as in (1) and (2) in Section 2.1.

Proposition 7. *Set $\mathcal{O}$ is rationalisable by an exponential discounted utility function if and only if for an arbitrary subset of $\mathcal{R}^*$, whenever distribution ν first order stochastically dominates μ and $\sum_{t\in\mathcal{T}}t\tau(t) \geq \sum_{t\in\mathcal{T}}t\sigma(t)$, then $\mu = \nu$ and $\sum_{t\in\mathcal{T}}t\tau(t) = \sum_{t\in\mathcal{T}}t\sigma(t)$.*

The above propositions states a necessary and sufficient condition under which a set of observations can be rationalised by an exponential discounted utility function. The requirement significantly resembles the one stated in the definition of the strong cumulative dominance axiom. However, it differs in two aspects. First of all, the conditions stated in Proposition 7 need to be verified only for *subsets* of the directly revealed preference relation, and *not* for arbitrary samples. This allows to simplify the test substantially.

Second of all, the only distinguishable implications of the quasi-hyperbolic and exponential models are implied by condition (iii) of the strong cumulative dominance axiom. Clearly, Proposition 7 imposes a more stringent condition on the set of observations, as it admits a larger class of samples that might violate it. However, whenever the strong cumulative dominance axiom is satisfied for l equal to the least or the greatest observable time-delay, then the two requirements are equivalent.

Corollary 1. *Set $\mathcal{O}$ is rationalisable by an exponential discounted utility function if and only if it obeys the strong cumulative dominance axiom for $l = \min \mathcal{T}$ or $l = \max \mathcal{T}$.*

Proof. Let $\mathcal{O}$ be rationalisable by an exponential discounted utility function. By Proposition 7, for an arbitrary sample of $\mathcal{R}^*$, distribution ν first order stochastically dominates μ and $\sum_{t\in\mathcal{T}} t\tau(t) \geq \sum_{t\in\mathcal{T}} t\sigma(t)$ if and only if $\mu = \nu$ and $\sum_{t\in\mathcal{T}} t\tau(t) = \sum_{t\in\mathcal{T}} t\sigma(t)$. Note that, whenever $l = \min \mathcal{T}$, condition (iii) of Axiom 4 is trivially satisfied with equality. On the other hand, if $l = \max \mathcal{T}$, then conditions (ii) and (iii) of the axiom coincide. □

4.4 Discount factors indeterminacy

In the following section we discuss some indeterminacy issues that arise while rationalising the set of observations by a quasi-hyperbolic or exponential discounted utility functions.

Proposition 8. *Set of observations $\mathcal{O}$ is rationalisable by a discounted utility function $v(m,t) := u(m)\gamma(t)$ if and only if, for any $a > 0$, it is rationalisable by a discounted utility function $\hat{v}(m,t) := \hat{u}(m)\hat{\gamma}(t)$, where $\hat{u} := u^a$ and $\hat{\gamma} := \gamma^a$.*

We omit the proof. The above result simply states that whenever a set of observations is rationalisable by a discounted utility function, then it is also rationalisable by its positive exponential transformation. This observation is not new, and was already noted by Fishburn and Rubinstein (1982, Theorem 2) in their representation theorem. However, the above proposition implies a much stronger conclusion. Namely, any property of function γ that is preserved under positive exponential transformations, is also satisfied by function $\hat{\gamma}$. In particular, this means that whenever γ is weakly present-biased (respectively quasi-hyperbolic or exponential) then so is $\hat{\gamma}$.[9]

Corollary 2. *Set $\mathcal{O}$ is rationalisable by a quasi-hyperbolic discounted utility function if and only if for any β (or δ) in $(0,1)$ there is some δ (respectively β) in $(0,1)$, a strictly increasing utility function $u : X \to \mathbb{R}_+$, and a time-delay k such that $v(x,t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$, where $\gamma(t) = \beta^t\delta^t$, for $t < k$, and $\gamma(t) = \beta^k\delta^t$ otherwise.*

Proof. ($\Rightarrow$). Take a strictly increasing function $u : X \to \mathbb{R}_+$, discount factors β, δ in $(0,1)$, as well as a time-delay k such that $v(x,t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$, where function γ is defined as in the thesis of the corollary. For any $\beta' \in (0,1)$ define $a := \log(\beta')/\log(\beta)$. Clearly, function $\hat{v}(x,t) := \hat{u}(x)\hat{\gamma}(t)$, where $\hat{u} = u^a$ and $\hat{\gamma} = \gamma^a$, also rationalises $\mathcal{O}$. Moreover, $\hat{\gamma}(t) = (\beta')^t(\delta^a)^t$, for $t < k$, and $\hat{\gamma}(t) = (\beta')^k(\delta^a)^t$ otherwise. We present an analogous argument for the claim within the brackets. ($\Leftarrow$) is trivial. □

Given the above result, there are no testable restrictions for the values of discount factors β or δ, as long as we consider the two parameters separately. However, there exists a restriction on pairs (β, δ) of the two discount factors. In fact, a straightforward application of the above result allows to determine the restriction on the ratio $\log(\delta)/\log(\beta)$.

Additionally, the above result implies no observable implications for the value of the discount factor δ rationalising the set of observations under the exponential discounting. Once the choice data can be rationalised for one value of the discount factor, it can be rationalised for virtually *any* other value $\delta \in (0,1)$.

Corollary 3. *Set $\mathcal{O}$ is rationalisable by an exponential discounted utility function if and only if, for any $\delta \in (0,1)$, there exists a strictly increasing function $u : X \to \mathbb{R}_+$ such that $v(x,t) := \delta^t u(x)$ rationalises $\mathcal{O}$.*

The above observation is implied by Corollaries 1 and 2. Even though it is impossible to determine the value of the discount factor corresponding to the exponential discounted utility representation, we are able to impose a testable restriction on this class of models. Therefore, the conditions stated in Proposition 7 allow us to evaluate the shape of the discounting function, but not the parameter characterising it.

[9] This also concerns the properties of the instantaneous utility function u. In particular, classes of log-concave and power utility functions are closed under positive exponential transformations. This observations plays an important role in Section 5 and the Online Appendix (see Section C of the appendix), where we discuss the two families of instantaneous utilities.

5 Implementation

We apply our results to observations from the experiment performed in Kirby and Petry (2004), which main purpose was to determine the impact of various forms of addiction on time-preference. The test was performed on 178 adult individuals. The sample included 134 subjects who were considered to be alcohol, cocaine, and/or heroin-abusers. The remaining 44 participants were a non-substance abusing control group. See Kirby and Petry (2004) for the details concerning the selection of the participants.

Throughout this section we let $X = \mathbb{R}_+$ and $T = \mathbb{N}$. During the experiment, each subject was presented with an identical list of 27 monetary-choice questions. In every question the participant was allowed to choose between a smaller, immediate reward x (i.e., delivered at time $t = 0$), and a larger reward y, delivered with some delay $s > 0$. Using our notation, the experiment $\mathcal{E}$ consisted of 27 feasible action sets $A := \big\{(x, 0), (y, s)\big\}$, where x, y in X and $s \in T$. Since the questionnaire was constructed in a way which allowed for the monetary rewards x, y and time-delay s to vary across questions, the experiment was *not* anchored, as specified in Section 3.2.[10] Finally, the questionnaire used in the above experiment is identical to the one utilised in Kirby, Petry, and Bickel (1999), Chabris, Laibson, Morris, Schuldt, and Taubinsky (2008, 2009), and Kirby (2009).[11]

In order to precisely characterise preferences of subjects participating in the experiment, we performed a number of tests on the available data. First, we focused on the weakest form of rationalisability, discussed in Section 2.2, and determined the number of subjects whose choices could be explained by a mixed-monotone utility maximisation. Then, we tested for the discounted utility specification of time-preference, with different forms of the discounting function. Hence, we directly implemented methods introduced in Sections 3 and 4 to the sets of observations generated in the experiment.

What is more, we performed two additional series of tests, in which we imposed stronger restrictions on the form of the instantaneous utility function. In the first analysis, for each specification of discounting, we verified whether individual observations could be rationalised by a discounted utility function with a strictly increasing and *log-concave* instantaneous utility u. In the second series of tests, we restricted our attention to *power* (or log-linear) utility functions $u(x) := x^\alpha$, where $\alpha > 0$. Clearly, the above restrictions made our tests significantly more stringent, and allowed for a more precise

[10] In particular, Proposition 3 does not apply to this setting. In fact, as we show in the following subsection, not every rationalisable set of observations generated in the experiment could be supported by a discounted utility function with an arbitrary discounting. See also Example 2.

[11] We transcribe the questionnaire in Table 2 in the Online Appendix (see Section C of the appendix).

characterisation of the revealed time-preferences induced by the sets of observations. The details concerning the stricter versions of the test as well as their dual characterisations can be found in the Online Appendix (see Section C of the appendix).

5.1 Power of the experiment

Suppose that an agent participating in the experiment answered all the 27 questions randomly with the uniform probability. The *power of the test* is the likelihood of rationalising the random set of observations by a respective form of time-preference. Clearly, the lower is the probability, the higher is the power of the test.[12] As mentioned in the introduction, each participant of the Kirby and Petry (2004) experiment was presented with 27 monetary-choice questions. Since in each question the agent was allowed to choose exactly one out of two options, the total number of possible sets of observations was $2^{27} = 134\ 217\ 728$. At the same time, the total number of sets of observations that could be rationalised by a mixed-monotone utility function, defined as in Section 2.2, was 13 013. Hence, the probability of a random set of observations being rationalisable was less than 0.0001. Therefore, we consider the test to be considerably powerful.[13]

Next, we determined the probability of rationalising a random set of observations by a specific form of the discounted utility model, conditional on the set being rationalisable in the mixed-monotone sense. We present the results in Table 1. Observe that, the design of the questionnaire used in Kirby and Petry (2004), i.e., the particular combination of questions, poses significant difficulties in demarcating between different specifications of time-preference. First of all, any rationalisable set of observations can be rationalised by a discounted utility function $v(x,t) := u(x)\gamma(t)$, with a strictly increasing instantaneous utility u and a discounting function γ. In addition, conditional on the mixed-monotone rationalisation, the design of the experiment induces no testable restrictions for the weakly present-biased or quasi-hyperbolic specification of discounting, as long as we impose no conditions on the form of the instantaneous utility function. Finally, 95.7% of the rationalisable sets of observations can be explained by an exponential discounted utility maximisation. Hence, the design of the questionnaire does not allow for a conclusive

[12]The above measure is inspired by Bronars (1987), whose method has become a common practice in determining the power of revealed preference tests for consumption-price data à la Afriat.

[13]Recall that the identical design of the experiment was utilised in Kirby, Petry, and Bickel (1999), Chabris, Laibson, Morris, Schuldt, and Taubinsky (2008, 2009), and Kirby (2009). Hence, the observations presented in this subsection apply to the above studies.

	Arbitrary utility function	Log-concave utility function	Power utility function
Arbitrary discounting	1	.0373	.0052
Weakly present-biased and quasi-hyperbolic discounting	1	.0347	.0037
Exponential discounting	.9567	.0254	.0022

Table 1: Probability of a random set of observations generated in the Kirby and Petry (2004) experiment being rationalisable by a discounted utility function, conditional on the set being rationalisable in the mixed-monotone sense.

separation between different specifications of time-preference.[14]

Once we impose stronger restrictions on the specification of the instantaneous utility function, our analysis becomes significantly more conclusive. First of all, the power of the test increases substantially, as the conditional probability of a random set of observations being rationalisable by a discounted utility function with a log-concave or power utility function falls below 4%. Moreover, the additional constraints make it easier to differentiate between different forms of discounting. However, the design of the experiment still does not allow to demarcate between the weakly present-biased and quasi-hyperbolic specifications of discounting.[15]

5.2 Rationalisable choices

In the second step of our analysis we applied our tests to the sets of observations generated in the experiment. For each individual, we determined the maximal number of observations that could be rationalised by a given specification of the discounted utility.

Out of 178 agents participating in the experiment, choices of 146 (82.0%) subjects could be fully rationalised by a mixed-monotone utility function. We find the result remarkable, as only 0.01% of all conceivable observation sets generated in the experiment could be rationalised in the above sense (see Section 5.1). Moreover, 171 (96.1%) sets of

[14]The above observation implies that Proposition 3 does not apply to the experiment by Kirby and Petry (2004). In particular, 4.3% of all rationalisable observation sets cannot be explained by an exponential discounted utility function with a strictly increasing instantaneous utility.

[15]In Table 3 in the Online Appendix (Section C of the appendix) we present the corresponding probabilities for a random set of observations being rationalisable if we drop at most one observation.

observations could be rationalised once we dropped at most one element from each set.

Given the insight from the previous subsection, it is not surprising that, as long as we imposed no restrictions on the form of the instantaneous utility function, the same number of observation sets could be rationalised by the discounted utility model with an arbitrary, weakly present-biased, and/or quasi-hyperbolic discounting. In addition, the distribution of the maximal number of observations rationalisable by exponential discounting was not significantly different from the distributions corresponding to the more general specifications time-preference (see Figure 1a).[16] Therefore, the mixed-monotone model of time-preference allowed for a good approximation of the choices observed in the experiment. However, given the construction of the questionnaire, we were unable to demarcate between the narrower specifications of the discounted utility.

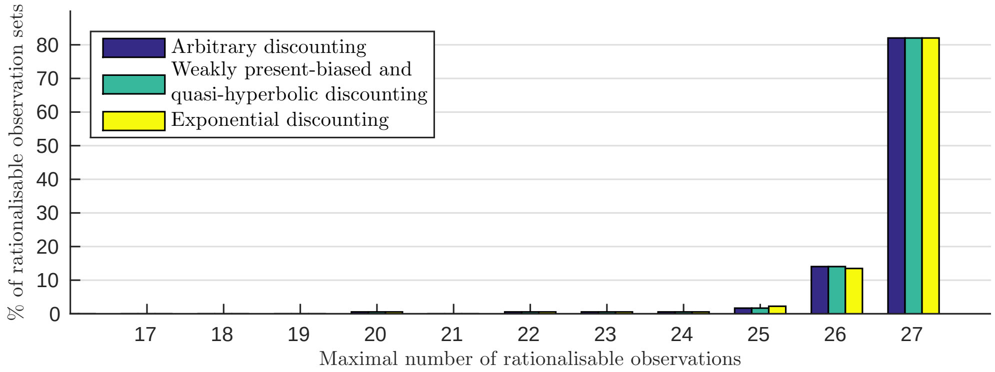
Once we restricted our attention to log-concave instantaneous utility functions, choices of 124 (69.9%) agents could be rationalised by the discounted utility model with an arbitrary, weakly present-biased, and/or quasi-hyperbolic discounting (see Figure 1b). Additionally, 162 (91.0%) sets of choices passed each of the above tests once we dropped at most one observation. Finally, the distribution of observations rationalisable by exponential discounting was significantly different from the distributions corresponding to the more general specifications of time-preference.[17] In particular, it was possible to explain 94 (52.8%) data sets with the exponential discounting model, while 148 (83.1%) sets could be rationalised in the above sense once we dropped at most one observation.

Our results changed significantly once we focused our attention on power instantaneous utility functions (see Figure 1c). In this case, choices of 70 (39.3%) subjects could be rationalised by the discounted utility model, out of which 67 (37.6%) could be explained by the weakly present-biased and/or quasi-hyperbolic, and 39 (21.9%) by the exponential specification of time-preference. Moreover, all the generated distributions were statistically distinct.[18] Once we dropped at most two observations from each of the data sets, we found that choices of more than 150 (84.2%) subjects could be rationalised by each of the above forms of the discounted utility model.

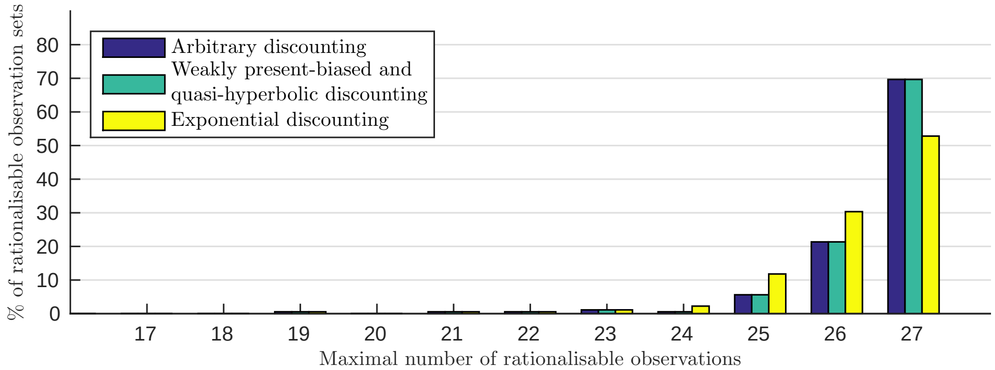
[16] At any conventional significance level, we were unable reject the null hypothesis of the Anderson-Darling test that the probability distributions were pairwise equivalent.

[17] In this case, we were able to rejected the null hypothesis of the Anderson-Darling test at any conventional significance level.

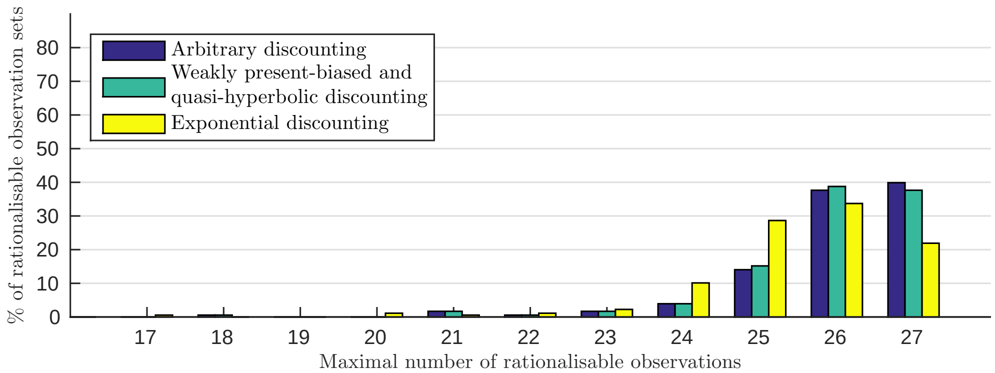
[18] We rejected the null hypothesis of the Anderson-Darling test for any conventional significance level.



(a) Rationalisation for an arbitrary instantaneous utility function.



(b) Rationalisation for a log-concave instantaneous utility function.



(c) Rationalisation for a power instantaneous utility function.

Figure 1: Empirical distributions of observation sets rationalisable by a particular form of the discounted utility model, with respect to the maximal number of rationalisable observations.

5.3 Undervalued rewards and error bounds

In this subsection we introduce an alternative measure of proximity of a set of observations from being fully rationalisable. An *error bound* is the minimal value of a coefficient ε in $[0,1]$ for which there exists an instantaneous utility function $u : X \to \mathbb{R}_+$ and a discounting function $\gamma : T \to (0,1]$, where $\gamma(0) = 1$, such that

$$(x,t)\,\mathcal{R}^*(y,s) \;\text{ implies }\; u(x)\gamma(t) \geq u\big((1-\varepsilon)y\big)\gamma(s).$$

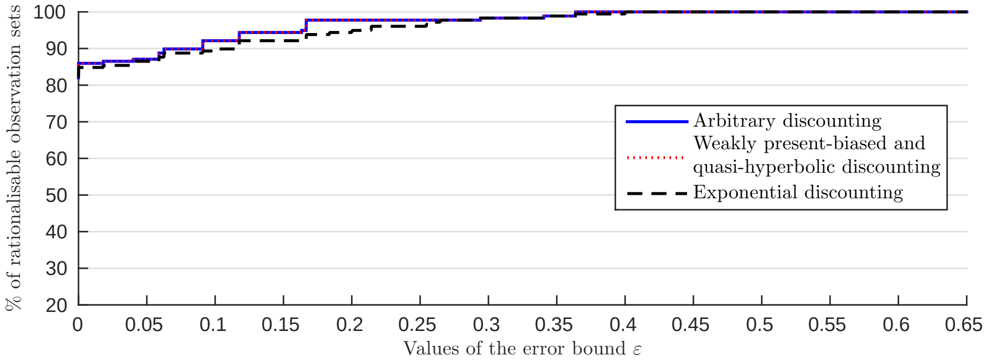
Note that, whenever the set of observations is rationalisable, the value of the error bound is equal to zero. Hence, we consider the above index to be an informative measure of proximity of choices from being rationalisable. In addition, the method can be applied to each of the discussed specifications of time-preference.

The above measure is inspired by the Critical Cost Efficiency Index which is commonly used in the Afriat-type revealed preference tests for consumption-price data. However, it is properly adjusted to our framework. In particular, note that coefficient ε affects only the value of the inferior rewards. Clearly, this specification of the test is purely arbitrary, however, it has three advantages. First of all, as the payments are bounded below by 0, the above measure belongs to the interval $[0,1]$. Second of all, the measure is monotone. That is, given an arbitrary set of observations, whenever the above condition is satisfied for some ε, then it is also satisfied for any $\varepsilon' \geq \varepsilon$.[19] Finally, since the monetary payments belong to $\mathbb{R}_+$, the measure takes continuous values.

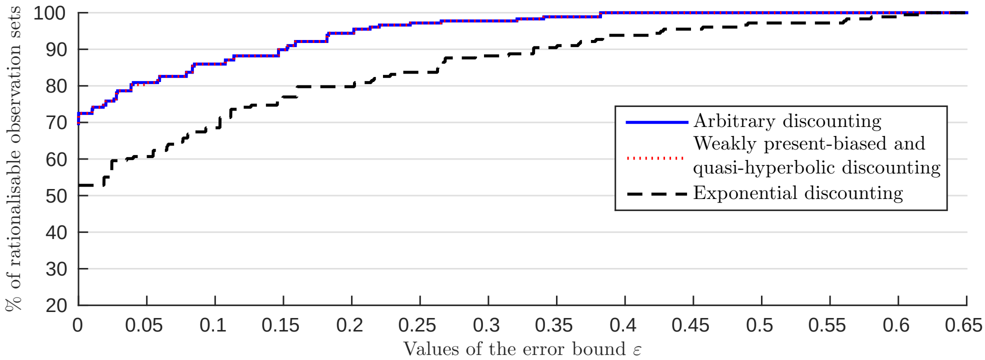
We present our results in Figure 2. For each specification of the instantaneous utility function, the distributions of error bounds for the arbitrary, weakly present-biased, and/or quasi-hyperbolic discounted utility models were statistically equivalent.[20] Second of all, in case of models with an arbitrary instantaneous utility functions, the distribution of ε for exponential discounting was statistically equivalent to distributions corresponding to the more general forms of time-preference (see Figure 2a). Nevertheless, this observation was no longer valid once we constrained out attention to either log-concave or power instantaneous utility functions. In both cases, the distribution of error bounds for

[19]Suppose that the above condition is satisfied for some ε and a mixed-monotone function v. Clearly, for any $\varepsilon' \geq \varepsilon$ and (y,s) in $X \times T$, we have $v\big((1-\varepsilon)y,s\big) \geq v\big((1-\varepsilon')y,s\big)$. In particular, this implies that whenever $(x,t)\,\mathcal{R}^*(y,s)$ then $v(x,t) \geq v\big((1-\varepsilon')y,s\big)$. Hence, the condition holds for ε'.

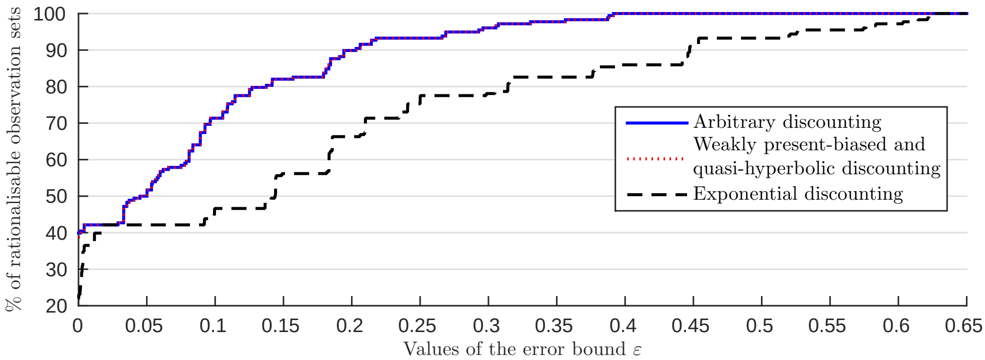
[20]In fact, in Figures 2a and 2c the distribution of error bounds for an arbitrary discounting coincides with the distribution for the weakly present-biased and/or quasi-hyperbolic representations. Moreover, for any conventional significance level, we could not reject the null hypothesis of the Andersen-Darling test, regardless of restrictions imposed on the form of the instantaneous utility function.



(a) Rationalisation for an arbitrary instantaneous utility function.



(b) Rationalisation for a log-concave instantaneous utility function.



(c) Rationalisation for a power instantaneous utility function.

Figure 2: Empirical cumulative distributions of observation sets for various specifications of the discounted utility model, with respect to the error bound ε.

exponential discounting was statistically distinct from distributions corresponding to the more general specifications of time-preference (see Figures 2b and 2c).[21]

In general, the discounted utility models seemed to approximate the observed choices remarkably well. In the case of an arbitrary specification of the instantaneous utility function, at $\varepsilon = 0.1$ it was possible to rationalise around 90% of all observation sets. Clearly, the stronger restrictions we imposed on the form of the instantaneous utility function, the greater was the minimal error at which the majority of sets could be rationalised. Nevertheless, the fraction of observation sets rationalisable under each specification of time-preference significantly exceeded the one generated randomly.[22]

5.4 Substance abuse and self-control

The main question posed in Kirby and Petry (2004) concerned the influence of substance abuse on impulsiveness and impatience of subjects participating in the experiment. By restricting their attention to a specific class of cardinal preferences with hyperbolic discounting, the authors were able to approximate values of individual discount rates.[23] In their main findings Kirby and Petry claimed that, even though rates of discounting varied with the preferred substance of abuse, substance-abusers were on average characterised by a higher value of the discount rate than the control group. Therefore, their study suggested that the level of impatience was negatively influenced by a history of addiction. The only notable exception concerned alcohol abuse, which relatively to the control group had no significant impact on the decisions made by agents.

We revisited the claim of Kirby and Petry, and using our non-parametric method we evaluated the relation between substance abuse and inter-temporal choice. Since our interest concentrated on the ordinal rather than cardinal notion of time-preference, the approach we proposed did not allow for an approximation of individual discount factors. However, using our tools we were able to verify which group of participants was more likely to be characterised by a certain type of utility representation. This allowed us to

[21]For any conventional significance level, we rejected the null hypothesis of the Anderson-Darling test.

[22]In the Online Appendix (see Section C of the appendix), for each specification of the discounted utility model, we present cumulative distributions of the minimal error bounds for a sample of 100 000 randomly generated sets of observations.

[23]In their paper, Kirby and Petry assumed that the instantaneous utility function of each agent was linear, i.e., $u(x) := x$, while the discounting function was hyperbolic, i.e., $\gamma(t) := 1/(1 + kt)$, for some $k > 0$. Under these assumptions, the authors were able to evaluate the most probable values of the coefficient k that rationalised the data set of each individual participating in the experiment. With a slight abuse of the terminology, throughout this section we refer to the coefficient k as to a *discount rate*.

draw conclusions on the impulsiveness and time-consistency of the participants.

We divided the participants into three groups: non substance-abusers, alcohol-abusers, and drug-abusers (i.e., cocaine and/or heroin).[24] For each group of subjects and each specification of the instantaneous utility function, we performed an analysis analogous to the one presented in Sections 5.2 and 5.3.

Main conclusions Our analysis implied that (i) *choices of the drug abusing participants were significantly less consistent* with utility maximisation than those of the remaining agents. In particular this concerned the exponential specification of the discounted utility model. Since this class of utility functions implies a time-consistent behaviour (see, e.g., Frederick, Loewenstein, and O'Donoghue, 2002), our results would suggest that (ii) *the drug abusing participants exhibited a lower level of time-consistency.* At the same time, there were (iii) *no significant differences in the behaviour of the control group and alcohol-abusers.* Below we present a detailed discussion on the results of our analysis.

Substance abuse and rationalisable observations We begun our analysis by determining distributions of the maximal number of observations rationalisable by a mixed-monotone utility function, for each group of participants (see Figure 3). Interestingly, the distribution corresponding to choices of drug-abusers was statistically distinct and first order stochastically dominated by distributions corresponding to the control and alcohol abusing group.[25] At the same time, there was no significant difference between the answers of the control group and alcohol-abusers.[26] Therefore, in terms of the maximal number of rationalisable observations, the non drug abusing participants did considerably better in comparison to the subjects with a history of drug addiction, while the answers of the control group and alcohol-abusers were indistinguishable.

When analysing discrepancies between the groups of substance-abusers under the more stringent forms of rationalisation, we separated those variations which were im-

[24]Note that the above groups of participants were not disjoint, as the sample contained a number of subjects who were abusing several substances simultaneously. Namely, 17 participants were abusing both alcohol and at least one type of a drug (i.e., either cocaine or heroin). However, our conclusions were not affected by this fact. In addition, since our analysis did not determine any significant differences between choices made by cocaine and heroin-abusers, we present the results for drug-abusers jointly.

[25]In either case, for any conventional significance level, we rejected the null hypothesis of the Anderson-Darling test. Moreover, the p-value from the Mann-Whitney U test was close to 0 in the comparison against the control group, and equal to 0.02 in the comparison against alcohol-abusers.

[26]We could not reject the null hypotheses of the Andersen-Darling and Mann-Whitney U tests at any conventional significance level.

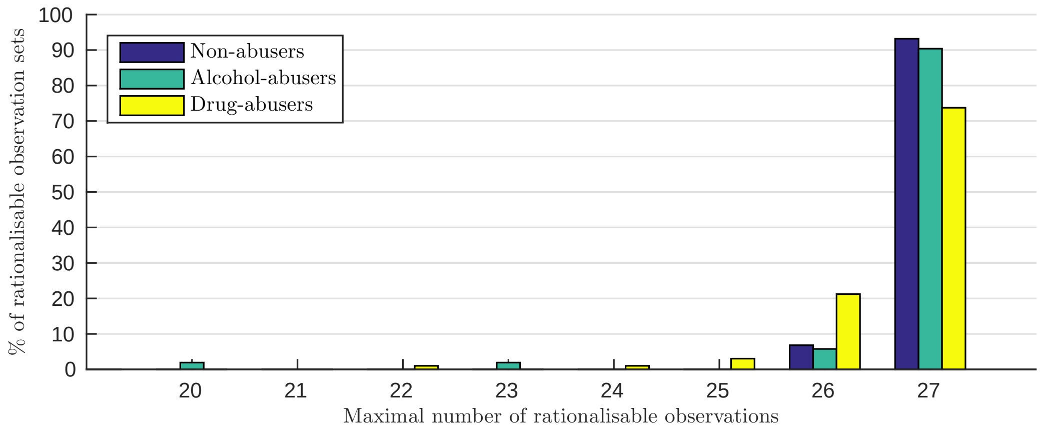


Figure 3: Empirical distributions of observation sets rationalisable by a mixed-monotone utility function for each group of substance-abusers.

plied by more general tests, from those genuine to more specific characterisations of time-preference. For example, when discussing the discounted utility model, we found it necessary to disentangle those variations that were implied by the previously determined differences in the mixed-monotone rationalisation, from those emerging only in the more stringent version of the analysis. For this reason, instead of comparing distributions of the maximal number of rationalisable choices, we focused on their relative changes.

Suppose that, for a given agent, the maximal number of observations rationalisable by a specific form of the discounted utility model is m. At the same time, let the number of observations rationalisable by a more stringent representation be n. Clearly, we have $n \leq m$. Moreover, the relative change in the number of rationalisable observations between the two tests is n/m. In particular, it is possible to construct this measure for any specification of time-preference and its superclass. Using this approach, we were able to determine the "decomposed" differences in the behaviour of the participants, which provided a more transparent view on the impact of substance abuse on time-preference.

In each possible comparison, we have determined no statistically significant differences between choices of the control group, alcohol-abusers, and drug-abusers. Therefore, as long as it concerns the maximal number of rationalisable observations, we were only able to conclude that preferences of agents with a history of drug addiction were less consistent in the mixed-monotone sense than those of the remaining participants.

Substance abuse and error bounds The analysis of error bounds allowed for a more significant differentiation between choices of the participants. Similarly as in the previous paragraph, in order to properly determine variations in the behaviour of the selected groups of substance-abusers, we needed to decompose the evaluated error bounds.

Suppose that the minimal error bound at which a set of observations is rationalisable by a certain form of the discounted utility model is ε.[27] Let ε' be the error bound at which the choices are rationalisable by a strictly narrower class of time-preference. A *decomposed error bound* for the latter representation conditional on the former is a number ζ in $[0, 1]$ such that $\big(1 - \varepsilon'\big) = (1 - \zeta)(1 - \varepsilon)$. Therefore, given that the initially perturbed set of observations can be supported by the general class of the discounted utility model, the decomposed error bound ζ determines the size of the additional perturbation that is necessary for the choices to be rationalisable by the narrower form of time-preference. Hence, the measure expresses the difficulty of rationalising the set of observations by a more stringent specification of time-preference, conditional on the data set being rationalisable by the more general characterisation. In particular, whenever the two representations are not distinguishable for a given data set, i.e., whenever $\varepsilon = \varepsilon'$, the value of the decomposed error bound ζ is zero.

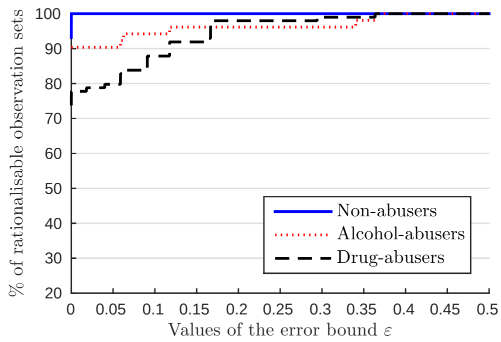
We present our results in Figure 4. Since, for each specification of the instantaneous utility function, the results for the weakly present-biased and quasi-hyperbolic specifications were indistinguishable from the general discounted utility model, we constrained our comparison to the arbitrary and exponential discounted utility rationalisations.

In almost every comparison, the distribution of errors for the drug abusing participants was statistically distinct and first order stochastically dominating the distribution corresponding to the control group.[28] The only exception concerned error bounds for the discounted utility model with a power instantaneous utility function (see Figure 4e). In this case, distributions for all groups of agents were indistinguishable. Hence, our results suggest that choices of the control group were significantly more consistent than those of the drug abusing agents. In particular, this concerned the exponential specification of the discounted utility model (for each class of instantaneous utilities), which would imply that participants with a history of drug addiction were less time-consistent.

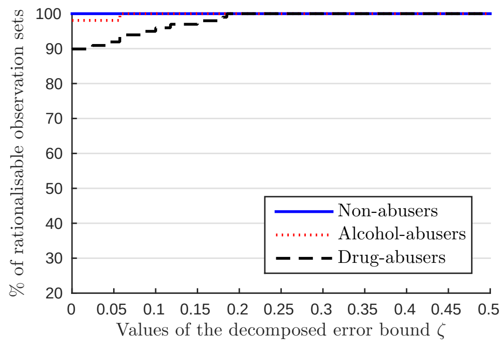
In each considered case, the answers of alcohol-abusers were statistically indifferent from those of the control group. Therefore, we obtain a similar result to Kirby and Petry (2004). However, the choices of alcohol-abusers were distinguishable and first order stochastically dominated by the decisions of drug-abusers only in the cases presented in

[27] Recall the definition of an *error bound* from Section 5.3

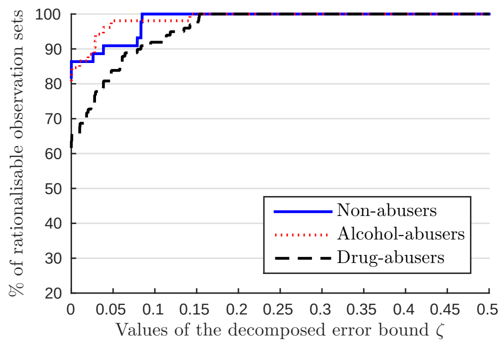
[28] Our conclusions depended on the adopted significance level. However, in general the null hypotheses of the Anderson-Darling and Mann-Whitney U tests were rejected at the p-value of 0.05. See Table 4 in the Online Appendix (Section C of the appendix) for the evaluated p-values from each comparison.



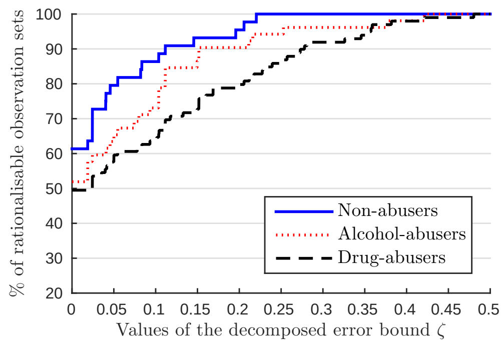
(a) Actual error bounds for the mixed-monotone and/or discounted utility model of time-preference (for an arbitrary utility).



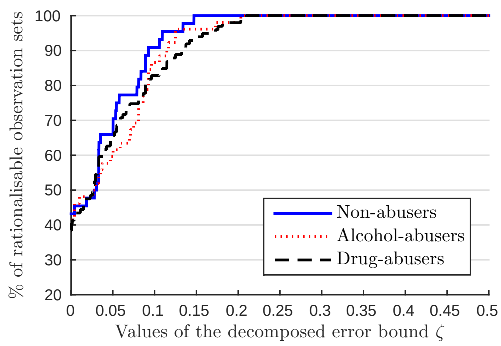
(b) Decomposed error bounds for the exponential discounted utility model (with an arbitrary utility) conditional on the discounted utility rationalisation (with an arbitrary utility).



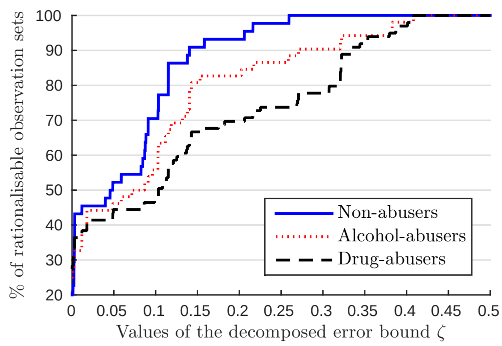
(c) Decomposed error bounds for the discounted utility model (with a log-concave utility) conditional on the discounted utility rationalisation (with an arbitrary utility).



(d) Decomposed error bounds for the exponential discounted utility model (with a log-concave utility) conditional on the discounted utility rationalisation (with a log-concave utility).



(e) Decomposed error bounds for the discounted utility model (with a power utility) conditional on the discounted utility rationalisation (with a log-concave utility).



(f) Decomposed error bounds for the exponential discounted utility model (with a power utility) conditional on the discounted utility rationalisation (with a power utility).

Figure 4: Empirical cumulative distributions of observation sets for each of the analysed groups of participants, with respect to the decomposed error bound ζ.

Figures 4a, 4b, and 4c. Hence, the comparison between the two forms of substance abuse was less conclusive than between the control group and drug-abusers.

A Auxiliary results

In this section we present the auxiliary results applied in our main arguments. We begin by introducing the so-called *Motzkin's Rational Transposition*, which is a modified version of the well-known Farkas' Lemma.

Theorem A.1 (Motzkin's Rational Transposition)**.** *Let A be a $k \times n$ matrix, let B be a $\ell \times n$ matrix, and let C be a $m \times n$ matrix, where B or C may be omitted (but not A). Exactly one of the following alternatives is true.*

(i) There exists $x \in \mathbb{R}^n$ such that $A \cdot x \gg 0$, $B \cdot x \geq 0$, and $C \cdot x = 0$.

(ii) There exist $\theta \in \mathbb{R}^k$, $\lambda \in \mathbb{R}^l$, and $\pi \in \mathbb{R}^m$ such that $\theta \cdot A + \lambda \cdot B + \pi \cdot C = 0$, where $\theta > 0$ and $\lambda \geq 0$. Moreover, if A, B, and C are rational, there exist $\theta \in \mathbb{N}^k$, $\lambda \in \mathbb{N}^l$, and $\pi \in \mathbb{N}^m$ satisfying the above condition.

The proof of the above theorem can be found in Stoer and Witzgall (1970). In the remainder of this section we apply the result to present a dual characterisation of various forms of stochastic dominance.

Let S be a finite subset of $\mathbb{R}^\ell$. Throughout the paper we find it convenient to refer to vectors $\epsilon_x \in \{0,1\}^{|S|}$, with the entry corresponding to element $x \in S$ being equal to 1 and the remaining coordinates equal to zero. For any x and $y \in S$ such that $x > y$,[29] a *monotone transformation* is a function $\rho_y^x : S \to \mathbb{R}$, defined by

$$\rho_y^x(z) := \begin{cases} 1 & \text{if } z = x, \\ -1 & \text{if } z = y, \\ 0 & \text{otherwise.} \end{cases}$$

Since set S is finite, we can represent function ρ_y^x by a vector in $\{-1,0,1\}^{|S|}$ defined as $\rho_y^x := \epsilon_x - \epsilon_y$, where ϵ_x and ϵ_y are specified as previously for x and y respectively. Denote the set of all such transformations by R.

Let Δ_S denote the set of probability distributions over S. For any μ, $\nu \in \Delta_S$, we say that μ dominates ν with respect to the *first order stochastic dominance* if for any

[29] Throughout this section we endow $\mathbb{R}^\ell$ with the coordinate-wise order $\geq$. Moreover, for any x and y in S, we say that $x > y$ whenever $x \geq y$ and $x \neq y$.

increasing function $f : S \to \mathbb{R}$, we have $\sum_{x \in S} f(x)\mu(x) \geq \sum_{x \in S} f(x)\nu(x)$.[30]

Proposition A.1. *For any μ, ν in Δ_S, probability distribution μ dominates ν with respect to the first order stochastic dominance if and only if*

$$\mu = \nu + \sum_{\rho \in R} \theta_\rho \rho.$$

for some non-negative coefficients $\{\theta_\rho\}_{\rho \in R}$.

Proof. For any increasing function $f : S \to \mathbb{R}$ and $\rho \in R$, we have $f \cdot \rho \geq 0$. In addition, equation $\mu = \nu + \sum_{\rho \in R} \theta_\rho \rho$ holds if and only if

$$\big(\mu - \nu\big) \in \Big\{ \sum\nolimits_{\rho \in R} \theta_\rho \rho : \theta_\rho \geq 0, \text{ for all } \rho \in R \Big\},$$

which is a closed convex cone. In particular, by Luenberger (1969, p. 215) the set of increasing functions constitutes the dual cone to the above set. Therefore, for any increasing function f, we have $f \cdot \big(\mu - \nu\big) \geq 0$ if and only if $\mu = \nu + \sum_{\rho \in R} \theta_\rho \rho$, for some non-negative coefficients $\{\theta_\rho\}_{\rho \in R}$. □

Given the above result, it is straightforward to show that whenever distribution μ dominates ν with respect to the first order stochastic dominance, while $\mu \neq \nu$, then for any *strictly* increasing function $f : S \to \mathbb{R}$, we have $\sum_{x \in S} f(x)\mu(x) > \sum_{x \in S} f(x)\nu(x)$. We summarise this observation in the following corollary.

Corollary A.1. *For any μ, ν in Δ_S such that $\mu \neq \nu$, if probability distribution μ first order stochastically dominates ν then, for any strictly increasing function $f : S \to \mathbb{R}$, we have $\sum_{x \in S} f(x)\mu(x) > \sum_{x \in S} f(x)\nu(x)$.*

Proof. Since μ first order stochastically dominates ν, while $\mu \neq \nu$, by Proposition A.1, there exists a subset R' of R and strictly positive numbers $\{\theta_\rho\}_{\rho \in R'}$ such that

$$\mu = \nu + \sum_{\rho \in R'} \theta_\rho \rho.$$

In addition, for any strictly increasing function $f : S \to \mathbb{R}$ and $\rho \in R$, we have $f \cdot \rho > 0$. Clearly, this implies that

$$f \cdot \nu \ < \ f \cdot \nu + \sum_{\rho \in R'} \theta_\rho f \cdot \rho \ = \ f \cdot \left(\nu + \sum_{\rho \in R'} \theta_\rho \rho \right) \ = \ f \cdot \mu,$$

which completes our argument. □

[30]We say that function $f : S \to \mathbb{R}$ is *increasing*, whenever $x \geq y$ implies $f(x) \geq f(y)$, for any x, y in S. In addition, the function is *strictly* increasing, if $x > y$ implies $f(x) > f(y)$.

Next we proceed with a dual characterisation of mean-preserving spreads and the second order stochastic dominance. For any x and y in S such that $x \neq y$, a *transformation* is a function $\pi_x^y : S \to \mathbb{R}$, defined by

$$\pi_y^x(z) := \begin{cases} 1 & \text{if } \; z = x, \\ -1 & \text{if } \; z = y, \\ 0 & \text{otherwise.} \end{cases}$$

Denote the set of all such transformations by P. Clearly, since any monotone transformation ρ is a transformation, we have $R \subseteq P$. As in the case of monotone transformations, we find it convenient to represent function π_y^x by a vector in $\{-1, 0, 1\}^{|S|}$ defined as $\pi_y^x := \epsilon_x - \epsilon_y$, where ϵ_x and ϵ_y are specified as previously.

For any two elements μ, ν in Δ_S, we say that probability distribution ν is a *mean-preserving spread* of μ, if $\sum_{x \in S} f(x)\mu(x) \geq \sum_{x \in S} f(x)\nu(x)$, for any concave function $f : \mathbb{R}^\ell \to \mathbb{R}$. Consider the following characterisation of the above relation.

Proposition A.2. *For any μ, ν in Δ_S, probability distribution ν is a mean-preserving spread of μ if and only if there exist some non-negative coefficients $\{\lambda_\pi\}_{\pi \in P}$ such that*

$$\mu = \nu + \sum_{\pi \in P} \lambda_\pi \pi,$$

while $\sum_{y \in S} \lambda_{\pi_y^x}(y - x) = 0$, for all $x \in S$.

Proof. ($\Rightarrow$). By definition, ν is a mean-preserving spread of μ if and only if there is no concave function $f : \mathbb{R}^\ell \to \mathbb{R}$ such that $\sum_{x \in S} f(x)\big(\nu(x) - \mu(x)\big) > 0$. Recall that, by the supergradient condition, for any concave function f and $x \in \mathbb{R}^\ell$, there is a vector $v_x \in \mathbb{R}^\ell$ such that $f(x) - f(y) + v_x \cdot (y - x) \geq 0$, for all $y \in \mathbb{R}^\ell$. In particular, this holds for any x and y belonging to S. Consider the following system of linear inequalities:

$$\begin{aligned} (\nu - \mu) \cdot f &> 0, \\ \pi_y^x \cdot f + (y - x) \cdot v_x &\geq 0, \text{ for all } x, y \in S, \end{aligned}$$

where $f \in \mathbb{R}^{|S|}$, while $v_x \in \mathbb{R}^{|S|}$, for all $x \in S$. Clearly, whenever ν is a mean-preserving spread of μ, the above system of inequalities has no solution. In particular, by Theorem A.1, there exist some non-negative coefficients $\{\lambda_\pi\}_{\pi \in P}$ such that

$$\mu = \nu + \sum_{\pi \in P} \lambda_\pi \pi,$$

while $\sum_{y \in S} \lambda_{\pi_y^x}(y - x) = 0$, for all $x \in S$.

(⇐). We prove this part of the proposition by contradiction. Suppose that there are no non-negative coefficients $\{\lambda_\pi\}_{\pi\in P}$ that satisfy the property specified in the thesis. By Theorem A.1 and the argument presented in the first part of the proof, this implies that there exists a solution to the system of linear inequalities specified above. Take any $f \in \mathbb{R}^{|S|}$ and $v_x \in \mathbb{R}^{|S|}$, $x \in S$, solving the system and construct function $f : \mathbb{R}^\ell \to \mathbb{R}$ by

$$f(x) := \min_{y\in S} \{f_y + v_y \cdot (x - y)\},$$

where we denote $f = (f_x)_{x\in S}$. Clearly, the function is concave, while $f(x) = f_x$. Moreover, we have $\sum_{x\in S} f(x)\big(\nu(x) - \mu(x)\big) > 0$, which contradicts that distribution ν is a mean-preserving spread of μ. □

Finally, we draw our attention to the second order stochastic dominance. For any two elements μ, ν of Δ_S, we say that probability distribution μ dominates ν with respect to the *second order stochastic dominance*, whenever for any increasing and concave function $f : \mathbb{R}^\ell \to \mathbb{R}$, we have $\sum_{x\in S} f(x)\mu(x) \geq \sum_{x\in S} f(x)\nu(x)$. In the following proposition we provide a dual characterisation of the relation.

Proposition A.3. *For any μ, ν in Δ_S, probability distribution μ dominates ν with respect to the second order stochastic dominance if and only if there exist some non-negative coefficients $\{\lambda_\pi\}_{\pi\in P}$ such that*

$$\mu = \nu + \sum_{\pi\in P} \lambda_\pi \pi,$$

while $\sum_{y\in S} \lambda_{\pi^x_y}(y - x) \leq 0$, for all $x \in S$.

Proof. (⇒). By an argument analogous to the one presented in the proof of Proposition A.2, whenever probability distribution μ dominates ν with respect to the second order stochastic dominance, the following system of linear inequalities:

$$\begin{aligned} (\nu - \mu) \cdot f &> 0, \\ \pi^x_y \cdot f + (y - x) \cdot v_x &\geq 0, \text{ for all } x, y \in S, \\ v_x &\geq 0, \text{ for all } x \in S, \end{aligned}$$

where $f \in \mathbb{R}^{|S|}$, while $v_x \in \mathbb{R}^{|S|}$, for all $x \in S$, has no solution. By Theorem A.1, this is satisfied if and only if there exist some non-negative coefficients $\{\lambda_\pi\}_{\pi\in P}$ such that

$$\mu = \nu + \sum_{\pi\in P} \lambda_\pi \pi,$$

as well as vectors $\{\theta_x\}_{x\in S}$, where $\theta_x \in \mathbb{R}^\ell_+$, such that $\sum_{y\in S} \lambda_{\pi^x_y}(y - x) = -\theta_x$, for all $x \in S$. Clearly, this implies that $\sum_{y\in S} \lambda_{\pi^x_y}(y - x) \leq 0$, for all $x \in S$.

(⇐). We prove this part of the theorem by contradiction. Whenever the premise of the implication is not satisfied, Theorem A.1 implies that there exists a solution to the system of linear inequalities specified above. Take any $f \in \mathbb{R}^{|S|}$ and $v_x \in \mathbb{R}^{|S|}$, $x \in S$, solving the system and construct function $f : \mathbb{R}^\ell \to \mathbb{R}$ by

$$f(x) := \min_{y \in S} \{f_y + v_y \cdot (x - y)\},$$

where we denote $f = (f_x)_{x \in S}$. Clearly, the function is increasing and concave, while $f(x) = f_x$. Moreover, we have $\sum_{x \in S} f(x)\big(\nu(x) - \mu(x)\big) > 0$, which contradicts that distribution μ second order stochastically dominates ν. □

There exists a close relation between the notion of a mean-preserving spread and the second order stochastic dominance. We summarise the result in the next corollary.

Corollary A.2. *For some μ, ν in Δ_S, let distribution μ second order stochastically dominate ν. Whenever ν is* not *a mean-preserving spread of μ then, for any concave and strictly increasing function $f : \mathbb{R}^\ell \to \mathbb{R}$, we have $\sum_{x \in S} f(x)\mu(x) > \sum_{x \in S} f(x)\nu(x)$.*

Proof. Whenever the premise of the corollary is satisfied, by Propositions A.2 and A.3, there exist some non-negative weights $\{\lambda_\pi\}_{\pi \in P}$ such that

$$\mu = \nu + \sum_{\pi \in P} \lambda_\pi \pi,$$

while $\sum_{y \in S} \lambda_{\pi^x_y}(y - x) \leq 0$, for all $x \in S$, and $\sum_{y \in S} \lambda_{\pi^x_y}(y - x) < 0$, for some x. First, we show that, for any concave and strictly increasing function $f : \mathbb{R}^\ell \to \mathbb{R}$, we have

$$\sum_{y \in S} \lambda_{\pi^x_y}(y - x) \leq (<)\ 0 \ \Rightarrow\ \sum_{y \in S} \lambda_{\pi^x_y} f(y) \leq (<) \sum_{y \in S} \lambda_{\pi^x_y} f(x).$$

Since coefficients $\{\lambda_\pi\}_{\pi \in P}$ are non-negative, whenever $\sum_{y \in S} \lambda_{\pi^x_y} = 0$, it must be that $\sum_{y \in S} \lambda_{\pi^x_y} f(y) = \sum_{y \in S} \lambda_{\pi^x_y} f(x) = 0$. Suppose that $\sum_{y \in S} \lambda_{\pi^x_y} > 0$. By strict monotonicity of f and Jensen's Inequality, whenever $\sum_{y \in S} \lambda_{\pi^x_y}(y - x) \leq (<)\ 0$, then

$$\sum_{y \in S} \frac{\lambda_{\pi^x_y}}{\sum_{y \in S} \lambda_{\pi^x_y}} f(y) \ \leq\ f\left(\sum_{y \in S} \frac{\lambda_{\pi^x_y}}{\sum_{y \in S} \lambda_{\pi^x_y}} y\right) \ \leq (<)\ f(x),$$

which implies that $\sum_{y \in S} \lambda_{\pi^x_y} f(y) \leq (<) \sum_{y \in S} \lambda_{\pi^x_y} f(x)$. Moreover, once we sum up all the inequalities with respect to $x \in S$, the above condition implies that

$$0 \ <\ \sum_{x \in S} \sum_{y \in S} \lambda_{\pi^x_y}\big(f(x) - f(y)\big) \ =\ \sum_{\pi \in P} \lambda_\pi \big(f \cdot \pi\big),$$

where with a slight abuse of the notation, by $f \cdot \pi_y^x$ we denote $f(x) - f(y)$, for any $\pi_y^x \in P$. In order to complete the proof, observe that

$$\begin{aligned} f \cdot \nu &< f \cdot \nu + \sum_{\pi \in P} \lambda_\pi (f \cdot \pi) \\ &= f \cdot \left(\nu + \sum_{\pi \in P} \lambda_\pi \pi \right) \\ &= f \cdot \mu, \end{aligned}$$

which concludes our argument. □

B Proofs

In this section we present arguments supporting the results presented in the main body of the paper. We proceed with the proof of Theorem 1.

B.1 Proof of Theorem 1

As the necessity part of the result was discussed in the main body of the paper, in this section we support the sufficiency of the dominance axiom for the discounted utility rationalisation. We begin with the following lemma.

Lemma B.1. *Whenever $\mathcal{O}$ obeys the dominance axiom there exist numbers $\{\phi_x\}_{x \in \mathcal{X}}$ and $\{\varphi_t\}_{t \in \mathcal{T}}$ such that (i) $(x,t)\mathcal{R}^*(y,s)$ implies $\phi_x + \varphi_t \geq \phi_y + \varphi_s$, (ii) $x > y$ implies $\phi_x > \phi_y$, and (iii) $t < s$ implies $\varphi_t > \varphi_s$.*

Proof. Enumerate the elements of $\mathcal{R}^*$ so that it is equal to $\big\{\big((x_j,t_j),(y_j,s_j)\big)\big\}_{j \in J}$. For any $x \in \mathcal{X}$, let $\epsilon_x \in \{0,1\}^{|\mathcal{X}|}$ denote a vector equal to 1 at the entry corresponding to x, and equal to 0 elsewhere. Analogously, we define vector $\varepsilon_t \in \{0,1\}^{|\mathcal{T}|}$, for each $t \in \mathcal{T}$. Finally, let $R_{\mathcal{X}}$ denote the set of monotone transformations ρ on $\mathcal{X}$, defined as in Section A, and $R_{\mathcal{T}}$ denote the set monotone transformations ϱ on $\mathcal{T}$.

In order to prove the result it is sufficient to show that, if set $\mathcal{O}$ satisfies the dominance axiom, there exist vectors $\phi \in \mathbb{R}^{|\mathcal{X}|}$ and $\varphi \in \mathbb{R}^{|\mathcal{T}|}$ such that

$$\begin{aligned} \big(\epsilon_{x_j} - \epsilon_{y_j}\big) \cdot \phi + \big(\varepsilon_{t_j} - \varepsilon_{s_j}\big) \cdot \varphi &\geq 0, \text{ for all } j \in J, \\ \rho \cdot \phi &> 0, \text{ for all } \rho \in R_{\mathcal{X}}, \\ -\varrho \cdot \varphi &> 0, \text{ for all } \varrho \in R_{\mathcal{T}}. \end{aligned}$$

We prove the claim by contradiction. Suppose that the vectors do not exist. By Theorem A.1, there are some natural numbers $\{\lambda_j\}_{j\in J}$, $\{\theta_\rho\}_{\rho\in R_{\mathcal{X}}}$, and $\{\vartheta_\varrho\}_{\varrho\in R_{\mathcal{T}}}$ such that

$$\sum_{j\in J}\lambda_j\epsilon_{x_j}+\sum_{\rho\in R_{\mathcal{X}}}\theta_\rho\rho=\sum_{j\in J}\lambda_j\epsilon_{y_j}\quad\text{and}\quad\sum_{j\in J}\lambda_j\varepsilon_{s_j}+\sum_{\varrho\in R_{\mathcal{T}}}\vartheta_\varrho\varrho=\sum_{j\in J}\lambda_j\varepsilon_{t_j},\tag{A1}$$

where either $\theta_\rho>0$, for some $\rho\in R_{\mathcal{X}}$, or $\vartheta_\varrho>0$, for some $\varrho\in R_{\mathcal{T}}$. In particular, this implies that $\lambda_j>0$, for some $j\in J$. For any $x\in\mathcal{X}$, define set $K(x):=\{j\in J: x_j=x\}$. That is, $K(x)$ is the set of all indices $j\in J$ enumerating the elements of $\mathcal{R}^*$, for which the corresponding element x_j is equal to x. Analogously, let $L(y):=\{j\in J: y_j=y\}$. Define probability distributions μ and ν, by

$$\mu(x):=\frac{\sum_{k\in K(x)}\lambda_k}{\sum_{j\in J}\lambda_j}\quad\text{and}\quad\nu(y):=\frac{\sum_{l\in L(y)}\lambda_l}{\sum_{j\in J}\lambda_j}.\text{[31]}$$

Moreover, define sequence $\{\hat\theta_\rho\}_{\rho\in R_{\mathcal{X}}}$, by $\hat\theta_\rho:=\theta_\rho/\sum_{j\in J}\lambda_j$. Similarly, for any elements t, s in $\mathcal{T}$, we can define sets $M(t):=\{j\in J: t_j=t\}$ and $N(s):=\{j\in J: s_j=s\}$, as well as probability distributions

$$\tau(t):=\frac{\sum_{m\in M(t)}\lambda_m}{\sum_{j\in J}\lambda_j}\quad\text{and}\quad\sigma(s):=\frac{\sum_{n\in N(s)}\lambda_n}{\sum_{j\in J}\lambda_j}.$$

In addition, define sequence $\{\hat\vartheta_\varrho\}_{\varrho\in R_{\mathcal{T}}}$, by $\hat\vartheta_\varrho:=\vartheta_\varrho/\sum_{j\in J}\lambda_j$. Observe that, given the definition of the above probability distributions, condition (A1) is equivalent to

$$\mu+\sum_{\rho\in R_{\mathcal{X}}}\hat\theta_\rho\rho=\nu\quad\text{and}\quad\sigma+\sum_{\varrho\in R_{\mathcal{T}}}\hat\vartheta_\varrho\varrho=\tau.$$

By Proposition A.1, this implies that ν first order stochastically dominates μ, while τ first order stochastically dominates σ. Moreover, since either $\hat\theta_\rho>0$, for some $\rho\in R_{\mathcal{X}}$, or $\hat\vartheta_\varrho>0$, for some $\varrho\in R_{\mathcal{T}}$, at least one of the above relations is strict.

In order to complete the proof, construct a sample $\{((x_i,t_i),(y_i,s_i))\}_{i\in I}$ of $\mathcal{R}^*$, where each element $((x_j,t_j),(y_j,s_j))$ of the directly revealed preference relation is taken λ_j times, $j\in J$. Clearly, the corresponding distributions μ, ν, τ, and σ are defined as above. However, this violates the dominance axiom, which leads to a contradiction. □

In the next lemma we provide a sufficient condition for the set of observation to be rationalisable by a discounted utility function.

Lemma B.2. *Set $\mathcal{O}$ is rationalisable by a discounted utility function whenever there exist numbers $\{\phi_x\}_{x\in\mathcal{X}}$ and $\{\varphi_t\}_{t\in\mathcal{T}}$ such that (i) $(x,t)\mathcal{R}^*(y,s)$ implies $\phi_x+\varphi_t\geq\phi_y+\varphi_s$, (ii) $x>y$ implies $\phi_x>\phi_y$, and (iii) $t<s$ implies $\varphi_t>\varphi_s$.*

[31] In our notation we assume that $\sum_{i\in\emptyset}a_i=0$.

Proof. Define function $\phi : \mathcal{X} \to \mathbb{R}$ by $\phi(x) := \phi_x$. Clearly, the function is strictly increasing. Moreover, by Husseinov (2010, Corollary 3), there exists a strictly increasing extension $\bar{\phi} : X \to \mathbb{R}$ of ϕ. Similarly, define a strictly decreasing function $\varphi : \mathcal{T} \to \mathbb{R}$ by $\varphi(t) := \varphi_t$, with its strictly decreasing extension $\bar{\varphi} : T \to \mathbb{R}$. Define function $u : X \to \mathbb{R}_+$ by $u := \exp\left(\bar{\phi}\right)$ and $\gamma : T \to (0, 1]$ by $\gamma := \exp\left(\bar{\varphi} - \bar{\varphi}(0)\right)$. Clearly, u is strictly increasing, while γ is strictly decreasing with $\gamma(0) = 1$. As $u(x)\gamma(t) = \exp\left(\phi_x + \varphi_t - \bar{\varphi}(0)\right)$, for any $(x, t) \in \mathcal{A}$, it is straightforward to show that $v(x, t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$. □

Lemmas B.1 and B.2 imply that whenever the set of observations obeys the dominance axiom, we can construct an instantaneous utility function $u : X \to \mathbb{R}_+$ and a discounting function $\gamma : T \to (0, 1]$ such that $v(x, t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$. Clearly, this completes the argument supporting Theorem 1.

B.2 Proof of Proposition 3

Before we present the proof, we find it convenient to state the following lemma.

Lemma B.3. *Let $\mathcal{O}$ be generated by an anchored experiment. If $\mathcal{O}$ is cyclically consistent then there exists an increasing sequence $\{z_t\}_{t\in\mathcal{T}}$ in $\mathbb{R}_+$ such that $(x, t)\, \mathcal{R}^*\, (x^*, t^*)$ implies $x \geq z_t$ and $(x^*, t^*)\, \mathcal{R}^*\, (x, t)$ implies $z_t \geq x$. Moreover, $z_{t^*} = x^*$.*

Proof. For any $t \in \mathcal{T}$, define $\underline{z}_t := \max\left\{y \in \mathcal{X} : (x^*, t^*)\, \mathcal{R}^*\, (y, s),\ \text{for } s \leq t\right\}$ as well as $\overline{z}_t := \min\left\{y \in \mathcal{X} : (y, s)\, \mathcal{R}^*\, (x^*, t^*),\ \text{for } s \geq t\right\}$. Note that, by the definition of set $\mathcal{T}$, for each t, either $\overline{z}_t$ or $\underline{z}_t$ is well-defined.

Claim 1: Whenever $\overline{z}_t$ and $\underline{z}_t$ are well-defined, then $\overline{z}_t \geq \underline{z}_t$. Assume the opposite. Then, there exist some y, $y' \in \mathcal{X}$, with $y' < y$ and s, $s' \in \mathcal{T}$, with $s' \geq s$, such that $(y', s')\, \mathcal{R}^*\, (x^*, t^*)\, \mathcal{R}^*\, (y, s)$. However, we have $(y, s) \succ (y', s')$, which violates the cyclical consistency axiom. This yields a contradiction.

Claim 2: Whenever $\overline{z}_{t^+}$ and $\underline{z}_t$ exist, then $\underline{z}_t < \overline{z}_{t^+}$, where $t^+ := \min\{s \in \mathcal{T} : s > t\}$. Assume the opposite. Then there exist some y, y' in $\mathcal{X}$ and s, s' in $\mathcal{T}$ such that $y' \leq y$ and $s' \geq t^+ > t \geq s$, while $(y', s')\, \mathcal{R}^*\, (x^*, t^*)\, \mathcal{R}^*\, (y, s)$. However, since $(y, s) \succ (y', s')$, this contradicts that the set of observations $\mathcal{O}$ is cyclically consistent.

Claim 3: For any t, t' in $\mathcal{T}$ such that $\overline{z}_t$ and $\overline{z}_{t'}$ are well-defined, whenever $t' \geq t$ then $\overline{z}_{t'} \geq \overline{z}_t$. This follows directly from the definition of $\overline{z}_t$.

Claim 4: For any t, $t' \in \mathcal{T}$ such that $t \leq t'$, whenever $\overline{z}_{t'}$ exists, so does $\overline{z}_t$. This follows directly from the definition of $\overline{z}_t$.

Given the above claims, we are able to recursively construct the sequence specified in the thesis of the lemma. Take any $t \in \mathcal{T}$ and let t^+ be its immediate successor, defined as previously. Given z_t, we determine z_{t^+} as follows. If (i) both $\underline{z}_{t^+}$ and $\overline{z}_{t^+}$ exist and $\underline{z}_{t^+} < \overline{z}_{t^+}$, let z_{t^+} be equal to any number n such that $\min\{z_t, \underline{z}_{t^+}\} < n < \overline{z}_{t^+}$. If (ii) $\underline{z}_{t^+} = \overline{z}_{t^+}$, let $z_{t^+} := \underline{z}_{t^+} = \overline{z}_{t^+}$. If (iii) only $\underline{z}_t$ exists, let z_{t^+} be equal to any number n such that $n > \min\{z_t, \underline{z}_t\}$. Finally, if (iv) only $\overline{z}_{t^+}$ exists, let z_{t^+} be equal to any number n such that $z_t < n < \overline{z}_{t^+}$. We claim that the sequence is well-defined.

Let $\underline{t} := \min \mathcal{T}$ be lowest observable time-delay. Since $z_{\underline{t}}$ is the first element of the sequence, by Claim 1, it is always well-defined. Take any $t \in \mathcal{T}$ and suppose that sequence $\{z_s\}_{s \leq t}$ is constructed as above. We claim that z_{t^+} is also well-defined. If not, then by Claim 2, it must be that $\overline{z}_{t^+} \leq z_t$. In particular, by Claim 4, we know that $\overline{z}_t$ also exists. Consider three cases. If (a) $\underline{z}_t < z_t < \overline{z}_t$ then by Claim 3, we have $z_t < \overline{z}_{t^+}$. If (b) $z_t = \underline{z}_t = \overline{z}_t$, by Claim 2 it must be that $z_t < \overline{z}_{t^+}$. Finally, if (c) $z_t < \overline{z}_t$, then by Claim 3, we have $z_t < \overline{z}_{t^+}$. Clearly, this implies that z_{t^+} is always well-defined.

In order to complete the proof, we need to show that $(x^*, t^*)\, \mathcal{R}^*\, (x, t)$ implies $z_t \geq x$, while $(x, t)\, \mathcal{R}^*\, (x^*, t^*)$ implies $x \geq z_t$. Clearly, whenever $(x^*, t^*)\, \mathcal{R}^*\, (x, t)$, then $\underline{z}_t$ is well-defined and it must be that $x \leq \underline{z}_t \leq z_t$. Similarly, whenever $(x, t)\, \mathcal{R}^*\, (x^*, t^*)$, then $\overline{z}_t$ exists and $x \geq \overline{z}_t \geq z_t$. Hence, the sequence satisfies the desired property.

Finally, we claim that it is always possible for $z_{t^*} := x^*$. Note that, for any $t < t^*$, whenever $\underline{z}_t$ exists, then $\underline{z}_t < x^*$. Otherwise, there would exist some $x \in \mathcal{X}$ such that $x \geq x^*$ and $(x^*, t^*)\, \mathcal{R}^*\, (x, t)$, which would violate the cyclical consistency axiom. Similarly, for any $t > t^*$, we have $\overline{z}_t > x^*$ (whenever it exists). Otherwise, there would exist some $x \leq x^*$ such that $(x, t)\, \mathcal{R}^*\, (x^*, t^*)$, which would violate cyclical consistency. □

Given the above lemma, we may proceed with the proof of Proposition 3. Since implication (iii) ⇒ (ii) ⇒ (i) is straightforward, we only prove (i) ⇒ (iii). Take an arbitrary discounting function $\gamma : T \to (0, 1]$ and let u^* be any strictly positive number. By Lemma B.3, there exists a strictly increasing sequence $\{z_t\}_{t \in \mathcal{T}}$ in $\mathbb{R}_+$ such that $(x^*, t^*)\, \mathcal{R}^*\, (x, t)$ implies $x \leq z_t$, while $(x, t)\, \mathcal{R}^*\, (x^*, t^*)$ implies $x \geq z_t$. Moreover, $z_{t^*} = x^*$. Define function $u : \{z_t\}_{t \in \mathcal{T}} \to \mathbb{R}_+$ by $u(z_t) := u^* \gamma(t^*)/\gamma(t)$. As γ is strictly decreasing, the function is strictly increasing. Hence, by Husseinov (2010, Corollary 3), there exists a strictly increasing extension $\bar{u} : \mathbb{R}_+ \to \mathbb{R}_+$ of u.

We claim that function $v(x, t) := \bar{u}(x)\gamma(t)$ rationalises the set of observations $\mathcal{O}$. Suppose that $(x^*, t^*)\, \mathcal{R}^*\, (x, t)$. By Lemma B.3 and monotonicity of function $\bar{u}$, we have $\bar{u}(x) \leq \bar{u}(z_t) = u(z_t) = u^* \gamma(t^*)/\gamma(t)$, which implies that $u(x)\gamma(t) \leq u^* \gamma(t^*) = u(x^*)\gamma(t^*)$.

Analogously, we can show that $(x,t)\,\mathcal{R}^*\,(x^*,t^*)$ implies $u(x)\gamma(t) \geq u(x^*)\gamma(t^*)$.

B.3 Proof of Theorem 2

We state the proof of the theorem via two lemmas.

Lemma B.4. *If set $\mathcal{O}$ obeys the cumulative dominance axiom, there exist real numbers $\{\phi_x\}_{x\in\mathcal{X}}$, $\{\varphi_t\}_{t\in\mathcal{T}}$, and $\{v_t\}_{t\in\mathcal{T}}$ such that (i) $(x,t)\mathcal{R}^*(y,s)$ implies $\phi_x + \varphi_t \geq \phi_y + \varphi_s$, (ii) $x > y$ implies $\phi_x > \phi_y$, (iii) $v_t < 0$, for all $t \in \mathcal{T}$, and (iv) for all $t \in \mathcal{T}$, we have $\varphi_t + v_t(s-t) \leq \varphi_s$, for all $s \in \mathcal{T}$.*

Proof. Enumerate the elements of $\mathcal{R}^*$ so that the set is equal to $\big\{\big((x_j,t_j),(y_j,s_j)\big)\big\}_{j\in J}$. For any $x \in \mathcal{X}$, let $\epsilon_x \in \{0,1\}^{|\mathcal{X}|}$ denote a vector equal to 1 at the entry corresponding to x, and equal to 0 elsewhere. Analogously, we define vector $\varepsilon_t \in \{0,1\}^{|\mathcal{T}|}$, for each $t \in \mathcal{T}$. Finally, let R denote the set of monotone transformations ρ on $\mathcal{X}$, defined as in Section A, and P denote the set of transformations π on $\mathcal{T}$.

In order to prove the result it is sufficient to show that, if set $\mathcal{O}$ satisfies the cumulative dominance axiom, there exist vectors $\phi \in \mathbb{R}^{|\mathcal{X}|}$ and φ, $v \in \mathbb{R}^{|\mathcal{T}|}$ such that

$$\begin{aligned}
\big(\epsilon_{x_j} - \epsilon_{y_j}\big)\cdot\phi + \big(\varepsilon_{t_j} - \varepsilon_{s_j}\big)\cdot\varphi \;&\geq\; 0, \text{ for all } j \in J,\\
\rho\cdot\phi \;&>\; 0, \text{ for all } \rho \in R,\\
-\pi^t_s\cdot\varphi - (s-t)v_t \;&\geq\; 0, \text{ for all } t,s \in \mathcal{T},\\
-v_t \;&>\; 0, \text{ for all } t \in \mathcal{T},
\end{aligned}$$

where we denote $v = (v_t)_{t\in\mathcal{T}}$. We prove the claim by contradiction. Suppose that the vectors do not exist. By Theorem A.1, there are some natural numbers $\{\lambda_j\}_{j\in J}$, $\{\theta_\rho\}_{\rho\in R}$, $\{\vartheta_\pi\}_{\pi\in P}$, and $\{\kappa_t\}_{t\in\mathcal{T}}$ such that

$$\sum_{j\in J}\lambda_j\epsilon_{x_j} + \sum_{\rho\in R}\theta_\rho\rho = \sum_{j\in J}\lambda_j\epsilon_{y_j} \quad\text{and}\quad \sum_{j\in J}\lambda_j\varepsilon_{s_j} + \sum_{\pi\in P}\vartheta_\pi\pi = \sum_{j\in J}\lambda_j\varepsilon_{t_j}, \tag{A2}$$

while $\sum_{s\in\mathcal{T}}\vartheta_{\pi^t_s}(s-t) = -\kappa_t$, for all $t \in \mathcal{T}$. Moreover, either $\theta_\rho > 0$, for some $\rho \in R$, or $\kappa_t > 0$, for some $t \in \mathcal{T}$. Clearly, this implies that $\lambda_j > 0$, for some $j \in J$.

For any $x \in \mathcal{X}$, define set $K(x) := \{j \in J : x_j = x\}$. That is, $K(x)$ is the set of all indices $j \in J$ for which the corresponding element x_j is equal to x. Analogously, let $L(y) := \{j \in J : y_j = y\}$. Define probability distributions μ and ν, by

$$\mu(x) := \frac{\sum_{k\in K(x)}\lambda_k}{\sum_{j\in J}\lambda_j} \quad\text{and}\quad \nu(y) := \frac{\sum_{l\in L(y)}\lambda_l}{\sum_{j\in J}\lambda_j}.$$

Define sequence $\{\hat{\theta}_\rho\}_{\rho\in R}$, by $\hat{\theta}_\rho := \theta_\rho / \sum_{j\in J}\lambda_j$. Similarly, for any t, s in $\mathcal{T}$, we can define sets $M(t) := \{j \in J : t_j = t\}$ and $N(s) := \{j \in J : s_j = s\}$, as well as distributions

$$\tau(t) := \frac{\sum_{m\in M(t)}\lambda_m}{\sum_{j\in J}\lambda_j} \quad \text{and} \quad \sigma(s) := \frac{\sum_{n\in N(s)}\lambda_n}{\sum_{j\in J}\lambda_j}.$$

Moreover, define sequence $\{\hat{\vartheta}_\pi\}_{\pi\in P}$, by $\hat{\vartheta}_\pi := \vartheta_\pi / \sum_{j\in J}\lambda_j$. Observe that, given the definition of the above probability distributions, condition (A2) is equivalent to

$$\mu + \sum_{\rho\in R}\hat{\theta}_\rho\rho = \nu \quad \text{and} \quad \sigma + \sum_{\pi\in P}\hat{\vartheta}_\pi\pi = \tau,$$

where $\sum_{s\in\mathcal{T}}\hat{\vartheta}_{\pi_s^t}(s-t) \leq 0$, for all $t \in \mathcal{T}$. In addition, either $\hat{\theta}_\rho > 0$, for some $\rho \in R$, or $\sum_{s\in\mathcal{T}}\hat{\vartheta}_{\pi_s^t}(s-t) < 0$, for some $t \in \mathcal{T}$. Hence, by Propositions A.1 and A.3, this implies that ν first order stochastically dominates μ, while τ second order stochastically dominates σ. Moreover, either the former relation is strict or probability distribution σ is *not* a mean-preserving spread of τ (see Corollaries A.1 and A.2 in Section A).

In order to complete the proof, construct a sample $\{((x_i, t_i), (y_i, s_i))\}_{i\in I}$ of $\mathcal{R}^*$, where each element $((x_j, t_j), (y_j, s_j))$ of the directly revealed preference relation is taken λ_j times, $j \in J$. Clearly, the corresponding distributions μ, ν, τ, and σ are defined as above. However, this violates the cumulative dominance axiom. □

In the second lemma we present a sufficient condition for a set of observations to be rationalisable by a weakly present-biased discounted utility function.

Lemma B.5. *Set $\mathcal{O}$ is rationalisable by a weakly present-biased discounted utility function, whenever there are numbers $\{\phi_x\}_{x\in\mathcal{X}}$, $\{\varphi_t\}_{t\in\mathcal{T}}$, and $\{v_t\}_{t\in\mathcal{T}}$ such that (i) $(x,t)\,\mathcal{R}^*(y,s)$ implies $\phi_x + \varphi_t \geq \phi_y + \varphi_s$, (ii) $x > y$ implies $\phi_x > \phi_y$, (iii) $v_t < 0$, for all $t \in \mathcal{T}$, and (iv) for all $t \in \mathcal{T}$, we have $\varphi_t + v_t(s-t) \leq \varphi_s$, for all $s \in \mathcal{T}$.*

Proof. Define function $\phi : \mathcal{X} \to \mathbb{R}$ by $\phi(x) := \phi_x$. Clearly, the function is strictly increasing. Moreover, by Husseinov (2010, Corollary 3), there exists a strictly increasing extension $\bar{\phi} : X \to \mathbb{R}$ of ϕ. Next, let function $\bar{\varphi} : T \to \mathbb{R}$ be defined by

$$\bar{\varphi}(t) := \max_{s\in\mathcal{T}}\{\varphi_s + v_s(t-s)\},$$

which is strictly increasing, while $\bar{\varphi}(t) = \varphi_t$, for all $t \in \mathcal{T}$. Moreover, the function has a convex extension to $\mathbb{R}$. Define function $u : X \to \mathbb{R}$ by $u := \exp(\bar{\phi})$ and $\gamma : T \to (0,1]$ by $\gamma := \exp(\bar{\varphi} - \bar{\varphi}(0))$. Note that u is strictly increasing, while γ is strictly decreasing with $\gamma(0) = 1$. Finally, as $u(x)\gamma(t) = \exp(\phi_x + \varphi_t - \bar{\varphi}(0))$, for any $(x,t) \in \mathcal{A}$, it is straightforward to show that $v(x,t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$. □

Lemmas B.4 and B.5 imply that whenever the set of observations satisfies the cumulative dominance axiom, it is possible to construct a utility function $u : X \to \mathbb{R}_+$ and a weakly present-biased discounting function $\gamma : T \to (0, 1]$ such that $v(x, t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$. This concludes the proof of Theorem 2.

B.4 Proof of Proposition 5

In the following section we present two lemmas that support the sufficiency of the strong cumulative dominance axiom for rationalisability by a quasi-hyperbolic discounted utility function. At the same time, the two results complete the proof of Proposition 5.

Lemma B.6. *Whenever set of observations $\mathcal{O}$ obeys the strong cumulative dominance axiom for some $l \in \mathcal{T}$, there exist numbers $\{\phi_x\}_{x\in\mathcal{X}}$ and β, δ such that (i) $(x, t)\,\mathcal{R}^*\,(y, s)$ implies $\phi_x + \min\{t, l\}\beta + t\delta \geq \phi_y + \min\{s, l\}\beta + s\delta$, (ii) $x > y$ implies $\phi(x) > \phi(y)$, and (iii) β and δ are strictly negative.*

Proof. Enumerate the elements of $\mathcal{R}^*$ so that the set is equal to $\{((x_j, t_j), (y_j, s_j))\}_{j\in J}$. For any $x \in \mathcal{X}$, let $\epsilon_x \in \{0, 1\}^{|\mathcal{X}|}$ denote a vector equal to 1 at the entry corresponding to x, and equal to 0 elsewhere. Moreover, let R denote the set of monotone transformations ρ on $\mathcal{X}$, defined as in Section A.

In order to prove the result it is sufficient to show that, if set $\mathcal{O}$ satisfies the dominance axiom, there exists a vector $\phi \in \mathbb{R}^{|\mathcal{X}|}$ and numbers β, δ such that

$$\begin{aligned}
\big(\epsilon_{x_j} - \epsilon_{y_j}\big)\cdot\phi + \big(\min\{t_j, l\} - \min\{s_j, l\}\big)\beta + (t_j - s_j)\delta &\geq 0, \text{ for all } j \in J,\\
\rho\cdot\phi &> 0, \text{ for all } \rho \in R,\\
-\beta &> 0,\\
-\delta &> 0.
\end{aligned}$$

We prove the claim by contradiction. Suppose that the vectors do not exist. By Theorem A.1, there are some natural numbers $\{\lambda_j\}_{j\in J}$, $\{\theta_\rho\}_{\rho\in R}$, ϑ, and κ such that

$$\sum_{j\in J}\lambda_j\epsilon_{x_j} + \sum_{\rho\in R}\theta_\rho\rho = \sum_{j\in J}\lambda_j\epsilon_{y_j}, \tag{A3}$$

while $\sum_{j\in J}\lambda_j\big(\min\{t_j, l\} - \min\{s_j, l\}\big) = \vartheta$ and $\sum_{j\in J}\lambda_j(t_j - s_j) = \kappa$. Moreover, we either have $\theta_\rho > 0$, for some $\rho \in R$, $\vartheta > 0$, or $\kappa > 0$. In particular, this implies that $\lambda_j > 0$, for some $j \in J$. For any $x \in \mathcal{X}$, define set $K(x) := \{j \in J : x_j = x\}$. Analogously, let $L(y) := \{j \in J : y_j = y\}$. Define probability distributions μ and ν, by

$$\mu(x) := \frac{\sum_{k\in K(x)}\lambda_k}{\sum_{j\in J}\lambda_j} \quad\text{and}\quad \nu(y) := \frac{\sum_{l\in L(y)}\lambda_l}{\sum_{j\in J}\lambda_j}.$$

Define sequence $\{\hat{\theta}_\rho\}_{\rho\in R}$, by $\hat{\theta}_\rho := \theta_\rho / \sum_{j\in J}\lambda_j$. For any elements t, s in $\mathcal{T}$, we can define sets $M(t) := \{j \in J : t_j = t\}$ and $N(s) := \{j \in J : s_j = s\}$, as well as distributions

$$\tau(t) := \frac{\sum_{m\in M(t)}\lambda_m}{\sum_{j\in J}\lambda_j} \quad \text{and} \quad \sigma(s) := \frac{\sum_{n\in N(s)}\lambda_n}{\sum_{j\in J}\lambda_j}.$$

Given the above definitions, condition (A3) is equivalent to

$$\mu + \sum_{\rho\in R}\hat{\theta}_\rho \rho = \nu,$$

while $\sum_{t\in\mathcal{T}}\min\{t,l\}\tau(t) \geq \sum_{t\in\mathcal{T}}\min\{t,l\}\sigma(t)$ and $\sum_{t\in\mathcal{T}} t\tau(t) \geq \sum_{t\in\mathcal{T}} t\sigma(t)$. In particular, by Proposition A.1, distribution ν first order stochastically dominates μ. Moreover, as either $\theta_\rho > 0$, for some $\rho \in R$, $\vartheta > 0$, or $\kappa > 0$, at least one of the relations is strict.

Construct a sample $\{((x_i,t_i),(y_i,s_i))\}_{i\in I}$ of $\mathcal{R}^*$, where each element $((x_j,t_j),(y_j,s_j))$ of the directly revealed preference relation is taken λ_j times, $j \in J$. Clearly, the corresponding distributions μ, ν, τ, and σ are defined as above. However, this violates the strong cumulative dominance axiom, which leads to a contradiction. □

Lemma B.7. *Set of observations $\mathcal{O}$ is rationalisable by a quasi-hyperbolic discounted utility function if there is some $l \in \mathcal{T}$ and numbers $\{\phi_x\}_{x\in\mathcal{X}}$, β, δ such that (i) $(x,t)\,\mathcal{R}^*\,(y,s)$ implies $\phi_x + \min\{t,l\}\beta + t\delta \geq \phi_y + \min\{s,l\}\beta + s\delta$, (ii) $x > y$ implies $\phi(x) > \phi(y)$, and (iii) β and δ are strictly negative.*

Proof. Define function $\phi : \mathcal{X} \to \mathbb{R}$ by $\phi(x) := \phi_x$. By Husseinov (2010, Corollary 3), there exists a strictly increasing extension $\bar{\phi} : X \to \mathbb{R}$ of ϕ. Let $u : X \to \mathbb{R}_+$ be given by $u := \exp(\bar{\phi})$, while $\hat{\delta} := \exp(\delta)$ and $\hat{\beta} := \exp(\beta)$. Define $\gamma : T \to (0,1]$ by $\gamma(t) := \hat{\beta}^t\hat{\delta}^t$, whenever $t < l$, and $\gamma(t) := \hat{\beta}^l\hat{\delta}^t$ otherwise. Since for any $(x,t) \in \mathcal{A}$, we have $v(x,t) := u(x)\gamma(t) = \exp(\phi_x + t\delta + \min\{t,l\}\beta)$, it is straightforward to show that function v rationalises the set of observations $\mathcal{O}$. □

Note that Lemmas B.6 and B.7 imply that if set $\mathcal{O}$ obeys the strong cumulative dominance axiom, then it is rationalisable by a quasi-hyperbolic discounting function. This completes the sufficiency part of that argument supporting Proposition 5. At the same time, jointly with the discussion in Section 4.2, the above lemmas support Proposition 6.

B.5 Proof of Proposition 7

We show the necessity of the conditions stated in the proposition analogously to the quasi-hyperbolic case. In order to show sufficiency, enumerate the elements of $\mathcal{R}^*$ so that the set is equal to $\left\{\left((x_j,t_j),(y_j,s_j)\right)\right\}_{j\in J}$. For any $x\in\mathcal{X}$, let $\epsilon_x\in\{0,1\}^{|\mathcal{X}|}$ denote a vector equal to 1 at the entry corresponding to x, and 0 elsewhere. Moreover, let R denote the set of monotone transformations ρ on $\mathcal{X}$, specified as in Section A.

We show that, whenever set $\mathcal{O}$ satisfies the conditions stated in the thesis of the proposition, there exists a vector $\phi\in\mathbb{R}^{|\mathcal{X}|}$ and a number δ such that

$$\begin{aligned}\left(\epsilon_{x_j}-\epsilon_{y_j}\right)\cdot\phi+(t_j-s_j)\delta &\geq 0, \text{ for all } j\in J,\\ \rho\cdot\phi &> 0, \text{ for all } \rho\in R,\\ -\delta &> 0.\end{aligned}$$

We prove the claim by contradiction. Suppose that the solution does not exist. By Theorem A.1, there are some real positive numbers $\{\lambda_j\}_{j\in J}$, $\{\theta_\rho\}_{\rho\in R}$, and ϑ such that

$$\sum_{j\in J}\lambda_j\epsilon_{x_j}+\sum_{\rho\in R}\theta_\rho\rho=\sum_{j\in J}\lambda_j\epsilon_{y_j} \quad\text{and}\quad \sum_{j\in J}\lambda_j(t_j-s_j)=\vartheta. \tag{A4}$$

Moreover, we either have $\theta_\rho>0$, for some $\rho\in R$, or $\vartheta>0$. In particular, this implies that $\lambda_j>0$, for some $j\in J$. We claim that there are positive numbers $\{\kappa_k\}_{k\in K}$, as well as sets $J_k\subseteq J$ and $R_k\subseteq R$, $k\in K$, such that, for all $k\in K$,

$$\sum_{j\in J_k}\left(\epsilon_{t_j}-\epsilon_{s_j}\right)+\sum_{\rho\in R_k}\rho=0,$$

while $\sum_{k\in K}\kappa_k\sum_{j\in J_k}(t_j-s_j)=\sum_{j\in J}\lambda_j(t_j-s_j)=\vartheta$.

First, we show how to recursively construct numbers $\{\kappa_k\}_{k\in K}$, as well as sets J_k and R_k, $k\in K$. Take $k=1$ and let $\hat{\lambda}_j:=\lambda_j$, for all $j\in J$, and $\hat{\theta}_\rho:=\theta_\rho$, for all $\rho\in R$. Define $\underline{\hat{\lambda}}:=\min_{j\in J}\left\{\hat{\lambda}_j:\hat{\lambda}_j>0\right\}$. By the previous argument, such number exists. Take any $i\in J$ such that $\hat{\lambda}_i=\underline{\hat{\lambda}}$. Since for each $t\in\mathcal{T}$ vector ϵ_t has a single entry equal to 1 and all the remaining equal to 0, while $\rho_s^t=\epsilon_t-\epsilon_s$, there are sets $J_1\ni i$ and R_1 such that

$$\sum_{j\in J_1}\left(\epsilon_{t_j}-\epsilon_{s_j}\right)+\sum_{\rho\in R_1}\rho=0,$$

while $\hat{\lambda}_j\geq\underline{\hat{\lambda}}$, for all $j\in J_1$, and $\hat{\theta}_\rho\geq\underline{\hat{\lambda}}$, for all $\rho\in R_1$. Otherwise, condition (A4) would not be satisfied. Set $\kappa_1:=\underline{\hat{\lambda}}$. Redefine numbers $\lambda_j:=\hat{\lambda}_j-\kappa_1$, for all $j\in J_1$, and $\lambda_j:=\hat{\lambda}_j$ for all $j\notin J_1$. For each $k>1$, repeat the above argument. Clearly, the algorithm is finite. Moreover, by construction, we have $\sum_{k\in K}\kappa_k\sum_{j\in J_k}(t_j-s_j)=\sum_{j\in J}\lambda_j(t_j-s_j)=\vartheta$.

Observe that, by definition, for each $k\in K$, subset $\left\{(x_i,t_i),(y_i,s_i)\right\}_{i\in J_k}$ of $\mathcal{R}^*$ is such that distribution of x_i first order stochastically dominates the distribution of y_i. Since set $\mathcal{O}$ obeys the condition presented in the proposition, it must be that $\sum_{j\in J_k}(t_j-s_j)\leq 0$. This implies that $\vartheta=0$ and $\theta_\rho=0$, for all $\rho\in R$, which leads to a contradiction.

C Online Appendix

In this appendix we present some additional results that were not included in the paper. First, we discuss an axiomatic characterisation of the discounted utility model with additional restrictions imposed on the form of the instantaneous utility function. In the second part, we include results complementary to Section 5. Throughout this appendix we use the notation introduced in the main body of the paper.

C.1 Log-concavity and rationalisation

We proceed with presenting the necessary and sufficient conditions for rationalisation with a *log-concave* instantaneous utility function. Our approach is analogous to the one presented in the paper. However, for a strictly technical reason, we assume that the set of observable prizes $\mathcal{X}$ is a subset of $\mathbb{Q}^{\ell}_{+}$ rather that $\mathbb{R}^{\ell}_{+}$. This restriction is not without loss of generality. Nevertheless, given our application, we find it acceptable.

Before we proceed with the first result, recall that for any sample $\big\{\big((x_i,t_i),(y_i,s_i)\big)\big\}_{i\in I}$ of $\mathcal{R}^*$, the distributions of rewards in the superior and inferior options are denoted respectively by μ and ν, while the distributions of time-delays are given by τ and σ, as in (1) and (2) in Section 2.1 of the paper.

Proposition C.1. *The following statements are equivalent.*

(i) There is a log-concave, strictly increasing instantaneous utility function $u : X \to \mathbb{R}$ and a strictly decreasing discounting function $\gamma : T \to (0,1]$, where $\gamma(0) = 1$, such that $v(x,t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$.

(ii) For any sample $\big\{\big((x_i,t_i),(y_i,s_i)\big)\big\}_{i\in I}$ of $\mathcal{R}^*$, whenever distribution ν second order stochastically dominates μ, while τ first order stochastically dominates σ, then μ must be a mean-preserving spread of ν and $\tau = \sigma$.

(iii) There are real numbers $\{\phi_x\}_{x\in\mathcal{X}}$, $\{\varphi_t\}_{t\in\mathcal{T}}$, and vectors $\{\upsilon_x\}_{x\in\mathcal{X}}$ in $\mathbb{R}^{\ell}$ such that (a) $(x,t)\,\mathcal{R}^*(y,s)$ implies $\phi_x + \varphi_t \geq \phi_y + \varphi_s$, (b) $\upsilon_x \gg 0$, for all $x \in \mathcal{X}$, (c) for all $x \in \mathcal{X}$, we have $\phi_x + \upsilon_x \cdot (y - x) \geq \phi_y$, for all $y \in \mathcal{X}$, and (d) $t < s$ implies $\varphi_t > \varphi_s$.

Proof. To show that (i) $\Rightarrow$ (ii), assume that there is a log-concave, strictly increasing instantaneous utility function $u : X \to \mathbb{R}$ and a strictly decreasing discounting function $\gamma : T \to (0,1]$, where $\gamma(0) = 1$, such that $v(x,t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$. Clearly, function $w(x,t) := \phi(x) + \varphi(t)$, where $\phi := \log(u)$ and $\varphi := \log(\gamma)$, also rationalises the set. Moreover, ϕ is concave and strictly increasing, while φ is strictly decreasing.

Take any sample $\{((x_i,t_i),(y_i,s_i))\}_{i\in I}$ of $\mathcal{R}^*$. By definition, for any $i \in I$, it must be that $\phi(x_i)+\varphi(t_i) \geq \phi(y_i)+\varphi(s_i)$. Moreover, once we sum up the above inequalities with respect to the elements of the sample, we obtain

$$\sum_{x\in\mathcal{X}} \phi(x)\mu(x) + \sum_{t\in\mathcal{T}} \varphi(t)\tau(t) \geq \sum_{x\in\mathcal{X}} \phi(x)\nu(x) + \sum_{t\in\mathcal{T}} \varphi(t)\sigma(t),$$

where distributions μ, ν, τ, and σ are defined as in (1) and (2) in Section 2.1. Suppose that the sample is specified as in statement (ii). Since ϕ is concave and strictly increasing, we have $\sum_{x\in\mathcal{X}} \phi(x)\mu(x) \leq \sum_{x\in\mathcal{X}} \phi(x)\nu(x)$. Moreover, as the distribution of time-delays t first order stochastically dominates the distribution of s, the monotonicity of φ implies that $\sum_{t\in\mathcal{T}} \varphi(t)\tau(t) \leq \sum_{t\in\mathcal{T}} \varphi(t)\sigma(t)$. However, the two inequalities can be consistent with the initial condition only if they are satisfied with equality. By Corollaries A.1 and A.2 (see Section A of the main appendix to the paper), this requires for μ to be a mean-preserving spread of ν and for τ to be equivalent to σ.

Next we show that (ii) $\Rightarrow$ (iii). Enumerate the elements of $\mathcal{R}^*$ so that the set is equal to $\{((x_j,t_j),(y_j,s_j))\}_{j\in J}$. For any $x \in \mathcal{X}$, let $\epsilon_x \in \{0,1\}^{|\mathcal{X}|}$ be a vector equal to 1 at the entry corresponding to x, and 0 elsewhere. Analogously, we define a vector $\varepsilon_t \in \{0,1\}^{|\mathcal{T}|}$, for each $t \in \mathcal{T}$. Finally, let P denote the set of transformations π on $\mathcal{X}$, as in Section A of the main appendix, and R be the set of monotone transformations ϱ on $\mathcal{T}$.

In order to prove the result it is sufficient to show that, if set $\mathcal{O}$ satisfies the condition in statement (ii), there exist vectors $\phi \in \mathbb{R}^{|\mathcal{X}|}$, $\varphi \in \mathbb{R}^{|\mathcal{T}|}$, and $\upsilon_x \in \mathbb{R}^{\ell}$, $x \in \mathcal{X}$, such that

$$\begin{aligned}
(\epsilon_{x_j}-\epsilon_{y_j})\cdot\phi + (\varepsilon_{t_j}-\varepsilon_{s_j})\cdot\varphi &\geq 0, \text{ for all } j \in J,\\
-\varrho\cdot\varphi &> 0, \text{ for all } \rho \in R,\\
\pi^x_y\cdot\phi + (y-x)\cdot\upsilon_x &\geq 0, \text{ for all } x,\, y \in \mathcal{X},\\
\upsilon_x &\gg 0, \text{ for all } x \in \mathcal{X}.
\end{aligned}$$

We prove the claim by contradiction. Suppose that the vectors do not exist. Since $\mathcal{X}$ is a subset of $\mathbb{Q}^{\ell}_{+}$, by Theorem A.1, there are some natural numbers $\{\lambda_j\}_{j\in J}$, $\{\theta_\pi\}_{\pi\in P}$, $\{\vartheta_\varrho\}_{\varrho\in R}$, and vectors $\kappa_x \in \mathbb{N}^{\ell}$, $x \in \mathcal{X}$, such that

$$\sum_{j\in J}\lambda_j\epsilon_{x_j} + \sum_{\pi\in P}\theta_\pi\pi = \sum_{j\in J}\lambda_j\epsilon_{y_j} \quad\text{and}\quad \sum_{j\in J}\lambda_j\varepsilon_{s_j} + \sum_{\varrho\in R}\vartheta_\varrho\varrho = \sum_{j\in J}\lambda_j\varepsilon_{t_j}, \tag{A5}$$

while $\sum_{y\in\mathcal{X}}\theta_{\pi^x_y}(y-x) = -\kappa_x$, for all $x \in \mathcal{X}$. Moreover, either $\vartheta_\varrho > 0$, for some $\varrho \in R$, or $\kappa_x > 0$, for some $x \in \mathcal{X}$. Clearly, this implies that $\lambda_j > 0$, for some $j \in J$.

For any $x \in \mathcal{X}$, define set $K(x) := \{j \in J : x_j = x\}$. That is, $K(x)$ is the set of all indices $j \in J$ for which the corresponding element x_j is equal to x. Analogously, let

$L(y) := \{j \in J : y_j = y\}$. Define probability distributions μ and ν, by

$$\mu(x) := \frac{\sum_{k\in K(x)} \lambda_k}{\sum_{j\in J} \lambda_j} \quad \text{and} \quad \nu(y) := \frac{\sum_{l\in L(y)} \lambda_l}{\sum_{j\in J} \lambda_j}.$$

Define sequences $\{\hat{\theta}_\pi\}_{\pi\in P}$ and $\{\hat{\kappa}_x\}_{x\in\mathcal{X}}$, by $\hat{\theta}_\pi := \theta_\pi / \sum_{j\in J} \lambda_j$ and $\hat{\kappa}_x := \big(1/\sum_{j\in J}\lambda_j\big)\kappa_x$.[32] For any t, s in $\mathcal{T}$, we define sets $M(t) := \{j \in J : t_j = t\}$ and $N(s) := \{j \in J : s_j = s\}$, as well as probability distributions

$$\tau(t) := \frac{\sum_{m\in M(t)} \lambda_m}{\sum_{j\in J} \lambda_j} \quad \text{and} \quad \sigma(s) := \frac{\sum_{n\in N(s)} \lambda_n}{\sum_{j\in J} \lambda_j}.$$

Moreover, define sequence $\{\hat{\vartheta}_\varrho\}_{\varrho\in R}$, by $\hat{\vartheta}_\varrho := \vartheta_\varrho / \sum_{j\in J} \lambda_j$. Observe that, given the definition of the above probability distributions, condition (A5) is equivalent to

$$\mu + \sum_{\pi\in P} \hat{\theta}_\pi \pi = \nu \quad \text{and} \quad \sigma + \sum_{\varrho\in R} \hat{\vartheta}_\varrho \varrho = \tau,$$

where $\sum_{y\in\mathcal{X}} \hat{\theta}_{\pi^x_y}(y - x) = -\hat{\kappa}_x$, for all $x \in \mathcal{X}$. In addition, we either have $\hat{\vartheta}_\varrho > 0$, for some $\varrho \in R$, or $\hat{\kappa}_x > 0$, for some $x \in \mathcal{X}$. Clearly, the second condition would imply that $\sum_{y\in\mathcal{T}} \hat{\theta}_{\pi^x_y}(y - x) < 0$, for some $x \in \mathcal{X}$. Therefore, by Propositions A.1 and A.3 from Section A of the appendix, this means that ν second order stochastically dominates μ, while τ first order stochastically dominates σ. Moreover, either μ is *not* a mean-preserving spread of ν or τ strictly dominates σ (see Corollaries A.1 and A.2 in Section A).

In order to complete this part of the proof, construct a sample $\big\{\big((x_i, t_i), (y_i, s_i)\big)\big\}_{i\in I}$ of $\mathcal{R}^*$, where each element $\big((x_j, t_j), (y_j, s_j)\big)$ of the directly revealed preference relation is taken λ_j times, $j \in J$. Clearly, the corresponding distributions μ, ν, τ, and σ are defined as above. However, this violates the cumulative dominance axiom.

Finally, we show that (iii) $\Rightarrow$ (i). Define function $\bar{\phi} : X \to \mathbb{R}$ by

$$\bar{\phi}(x) := \min_{y\in\mathcal{X}} \big\{\phi_y + v_y \cdot (x - y)\big\}.$$

Clearly, it is concave and strictly increasing, while $\phi(x) = \phi_x$, for all $x \in \mathcal{X}$. Next, let $\varphi : \mathcal{T} \to \mathbb{R}$ be defined by $\varphi(t) = \varphi_t$. By assumption, the function is strictly decreasing. Moreover, by Husseinov (2010, Corollary 3), there exists a strictly decreasing extension $\bar{\varphi} : T \to \mathbb{R}$ of φ. Finally, define $u : X \to \mathbb{R}_+$ by $u := \exp\big(\bar{\phi}\big)$ and $\gamma : T \to (0, 1]$ by $\gamma := \exp\big(\bar{\varphi} - \bar{\varphi}(0)\big)$. Clearly, u is strictly increasing and log-concave, while γ is strictly decreasing with $\gamma(0) = 1$. As $u(x)\gamma(t) = \exp\big(\phi_x + \varphi_t - \bar{\varphi}(0)\big)$, for any $(x, t) \in \mathcal{A}$, it is straightforward to show that $v(x, t) := u(x)\gamma(t)$ rationalises $\mathcal{O}$. $\square$

[32] Clearly, sequence $\{\hat{\kappa}_x\}_{x\in\mathcal{X}}$ consists of vectors in $\mathbb{R}^\ell_+$.

Given the line of the above argument, as well as the results in the main body of the paper, it is possible to determine the necessary and sufficient conditions for the log-concave rationalisation with the more stringent restrictions imposed on the specification of the discounting function. We leave this part of the argument to the reader.

C.2 Power utility and rationalisation

In this subsection we concentrate on the necessary and sufficient conditions for a set of observations to be rationalisable with a *power utility* function. Throughout this section we assume that $X = \mathbb{R}_+$. The restriction is driven by our application presented in Section 5.

As previously, recall that for any sample $\{((x_i, t_i), (y_i, s_i))\}_{i \in I}$ of $\mathcal{R}^*$, the distributions of rewards in the superior and inferior options are denoted respectively by μ and ν, while the distributions of time-delays are given by τ and σ, as in (1) and (2) in Section 2.1.

Proposition C.2. *The following statements are equivalent.*

(i) There is some $\alpha > 0$ and a strictly decreasing discounting function $\gamma : T \to (0, 1]$, where $\gamma(0) = 1$, such that $v(x, t) := x^\alpha \gamma(t)$ rationalises $\mathcal{O}$.

(ii) For a subset $\{((x_i, t_i), (y_i, s_i))\}_{i \in I}$ of $\mathcal{R}^*$, if $\sum_{x \in \mathcal{X}} \log(x)\nu(x) \geq \sum_{x \in \mathcal{X}} \log(x)\mu(x)$, while probability distribution τ first order stochastically dominates σ, then it must be that $\sum_{x \in \mathcal{X}} \log(x)\nu(x) = \sum_{x \in \mathcal{X}} \log(x)\mu(x)$ and $\tau = \sigma$.

(iii) There are real numbers α and $\{\varphi_t\}_{t \in \mathcal{T}}$ such that (a) whenever $(x, t)\mathcal{R}^*(y, s)$ then $\alpha \log(x) + \varphi_t \geq \alpha \log(y) + \varphi_s$, (b) $\alpha > 0$, and (c) $t < s$ implies $\varphi_t > \varphi_s$.

Proof. To show that (i) $\Rightarrow$ (ii), assume that there is some $\alpha > 0$ and a strictly decreasing discounting function $\gamma : T \to (0, 1]$ such that $v(x, t) := x^\alpha \gamma(t)$ rationalises $\mathcal{O}$. Clearly, function $w(x, t) := \alpha \log(x) + \varphi(t)$, where $\varphi := \log(\gamma)$, also rationalises the set of observations. Moreover, function φ is strictly decreasing.

Take any subset $\{((x_i, t_i), (y_i, s_i))\}_{i \in I}$ of $\mathcal{R}^*$. By definition, for any $i \in I$, it must be that $\alpha \log(x_i) + \varphi(t_i) \geq \alpha \log(y_i) + \varphi(s_i)$. Moreover, once we sum up the above inequalities with respect to the elements in the set, we obtain

$$\alpha \sum_{x \in \mathcal{X}} \log(x)\mu(x) + \sum_{t \in \mathcal{T}} \varphi(t)\tau(t) \geq \alpha \sum_{x \in \mathcal{X}} \log(x)\nu(x) + \sum_{t \in \mathcal{T}} \varphi(t)\sigma(t),$$

where distributions μ, ν, τ, and σ are defined as in (1) and (2) in Section 2.1. Suppose that the set is specified as in statement (ii). Since α is strictly positive, we have $\sum_{x \in \mathcal{X}} \log(x)\mu(x) \leq \sum_{x \in \mathcal{X}} \log(x)\nu(x)$. Moreover, as the distribution of time-delays t first order stochastically dominates the distribution of s, the monotonicity of φ implies that $\sum_{t\in\mathcal{T}}\varphi(t)\tau(t) \leq \sum_{t\in\mathcal{T}}\varphi(t)\sigma(t)$. However, the two inequalities can be consistent with the initial condition only if they are satisfied with equality. By Corollary A.1 (see Section A of the appendix), this requires that $\sum_{x\in\mathcal{X}}\log(x)\mu(x) = \sum_{x\in\mathcal{X}}\log(x)\nu(x)$ and $\tau=\sigma$.

Next we show that (ii) $\Rightarrow$ (iii). Enumerate the elements of $\mathcal{R}^*$ so that the set is equal to $\{((x_j,t_j),(y_j,s_j))\}_{j\in J}$. For each $t\in\mathcal{T}$, define vector $\varepsilon_t\in\{0,1\}^{|\mathcal{T}|}$ equal to 1 at the coordinate corresponding to t, and 0 elsewhere. Finally, let R denote the set of transformations ϱ on $\mathcal{T}$, defined as in Section A of the main appendix to the paper.

In order to prove the result, it is sufficient to show that, if set $\mathcal{O}$ satisfies the condition in statement (ii), there is some real number α and a vector $\varphi\in\mathbb{R}^{|\mathcal{T}|}$ such that

$$\begin{aligned}\big(\log(x_j)-\log(y_j)\big)\alpha + \big(\varepsilon_{t_j}-\varepsilon_{s_j}\big)\cdot\varphi &\geq 0, \text{ for all } j\in J,\\ -\varrho\cdot\varphi &> 0, \text{ for all } \rho\in R,\\ \alpha &> 0.\end{aligned}$$

We prove the claim by contradiction. Suppose that the vectors do not exist. By Theorem A.1, there are some real numbers $\{\lambda_j\}_{j\in J}$, θ, and $\{\vartheta_\varrho\}_{\varrho\in R}$ such that

$$\sum_{j\in J}\lambda_j\big(\log(x_j)-\log(y_j)\big) = -\theta \quad\text{and}\quad \sum_{j\in J}\lambda_j\varepsilon_{s_j} + \sum_{\varrho\in R}\vartheta_\varrho\varrho = \sum_{j\in J}\lambda_j\varepsilon_{t_j}, \tag{A6}$$

Moreover, either $\vartheta_\varrho>0$, for some $\varrho\in R$, or $\theta>0$. Clearly, this implies that $\lambda_j>0$, for some $j\in J$. We claim that there exist positive numbers $\{\kappa_k\}_{k\in K}$, as well as sets $J_k\subseteq J$ and $R_k\subseteq R$, $k\in K$, such that, for all $k\in K$, we have

$$\sum_{j\in J_k}\big(\varepsilon_{t_j}-\varepsilon_{s_j}\big) + \sum_{\varrho\in R_k}\varrho = 0,$$

while $\sum_{k\in K}\kappa_k\sum_{j\in J_k}\big(\log(x_j)-\log(y_j)\big) = \sum_{j\in J}\lambda_j\big(\log(x_j)-\log(y_j)\big) = -\theta$.

First, we show how to recursively construct numbers $\{\kappa_k\}_{k\in K}$, as well as sets J_k and R_k, $k\in K$. Take $k=1$ and let $\hat{\lambda}_j := \lambda_j$, for all $j\in J$, and $\hat{\vartheta}_\rho := \vartheta_\rho$, for all $\rho\in R$. Define $\underline{\hat{\lambda}} := \min_{j\in J}\{\hat{\lambda}_j : \hat{\lambda}_j>0\}$. By the previous argument, such number exists. Take any $i\in J$ such that $\hat{\lambda}_i = \underline{\hat{\lambda}}$. Since for each $t\in\mathcal{T}$ vector ε_t has a single entry equal to 1 and all the remaining equal to 0, while $\varrho^t_s = \varepsilon_t - \varepsilon_s$, there are sets $J_1\ni i$ and R_1 such that

$$\sum_{j\in J_1}\big(\varepsilon_{t_j}-\varepsilon_{s_j}\big) + \sum_{\varrho\in R_1}\varrho = 0,$$

while $\hat{\lambda}_j \geq \underline{\hat{\lambda}}$, for all $j\in J_1$, and $\hat{\vartheta}_\rho \geq \underline{\hat{\lambda}}$, for all $\rho\in R_1$. Otherwise, condition (A6) would not be satisfied. Set $\kappa_1 := \underline{\hat{\lambda}}$. Redefine numbers $\lambda_j := \hat{\lambda}_j - \kappa_1$, for all $j\in J_1$, and $\lambda_j := \hat{\lambda}_j$

for all $j \notin J_1$. For each $k > 1$, repeat the above argument. Clearly, the algorithm is finite. Moreover, by construction, we have $\sum_{k\in K} \kappa_k \sum_{j\in J_k} \big(\log(x_j) - \log(y_j)\big) = -\theta$.

Observe that, by definition, for each $k \in K$, subset $\big\{(x_i, t_i), (y_i, s_i)\big\}_{i\in J_k}$ of $\mathcal{R}^*$ is such that distribution of t_i first order stochastically dominates the distribution of s_i. Since set $\mathcal{O}$ obeys the condition presented in statement (ii) of the proposition, it must be that $\sum_{j\in J_k} \big(\log(x_j) - \log(y_j)\big) \geq 0$, which implies that $\theta = 0$, as well as $\vartheta_\varrho = 0$, for all $\varrho \in R$, which would lead to a contradiction.

Finally, we show that (iii) $\Rightarrow$ (i). Let $\varphi : \mathcal{T} \to \mathbb{R}$ be defined by $\varphi(t) = \varphi_t$. By assumption, the function is strictly decreasing. Moreover, by Husseinov (2010, Corollary 3), there exists a strictly decreasing extension $\bar{\varphi} : T \to \mathbb{R}_-$ of φ. Define $\gamma : T \to (0, 1]$ by $\gamma := \exp\big(\bar{\varphi} - \bar{\varphi}(0)\big)$. As $x^\alpha\gamma(t) = \exp\big(\alpha \log(x) + \varphi_t - \bar{\varphi}(0)\big)$, for any $(x, t) \in \mathcal{A}$, it is straightforward to show that $v(x, t) := x^\alpha\gamma(t)$ rationalises $\mathcal{O}$. □

As in the case of log-concavity, given our previous results, it is possible to determine analogous necessary and sufficient conditions for the discounted utility model with instantaneous power utility function and the alternative specifications of discounting.

C.3 Appendix to Section 5

First, in Table 2 we present the questionnaire employed in the experiment by Kirby and Petry (2004). Recall that it is identical to the one utilised in Kirby, Petry, and Bickel (1999), Chabris, Laibson, Morris, Schuldt, and Taubinsky (2008, 2009), and Kirby (2009).

Next, we discuss the probability with which a random set of observations generated in the Kirby and Petry (2004) experiment can be rationalised once we allowed to drop at most one observation from each set. The total number of observation sets that could be rationalised in this sense by a mixed-monotone utility function was 177 502. Therefore, given the number of all conceivable observation sets equal to $2^{27} = 134\ 217\ 728$, the probability was 0.0014. In the table below, we present the corresponding probabilities for each specification of the discounted utility model.

Given the questionnaire employed in the experiment by Kirby and Petry (2004), we generated a sample of 100 000 random sets of observations. In Figure 5, for each specification of the discounted utility model, we present cumulative probability distributions of rationalisable sets, with respect to the minimal error bound ε (see Section 5.3 of the main body of the paper). Moreover, please note that the distribution of error bounds for the mixed-monotone time-preference coincided in this case with the arbitrary discounted utility model for an arbitrary instantaneous

Order	Reward values		
	Smaller immediate reward	Larger delayed reward	Delay (in days)
1	$54	$55	117
2	$55	$75	61
3	$19	$25	53
4	$31	$85	7
5	$14	$25	19
6	$47	$50	160
7	$15	$35	13
8	$25	$60	14
9	$78	$80	162
10	$40	$55	62
11	$11	$30	7
12	$67	$75	119
13	$34	$35	186
14	$27	$50	21
15	$69	$85	91
16	$49	$60	89
17	$80	$85	157
18	$24	$35	29
19	$33	$80	14
20	$28	$30	179
21	$34	$50	30
22	$25	$30	80
23	$41	$75	20
24	$54	$60	111
25	$54	$80	30
26	$22	$25	136
27	$20	$55	7

Table 2: The questionnaire employed in the experiment by Kirby and Petry (2004).

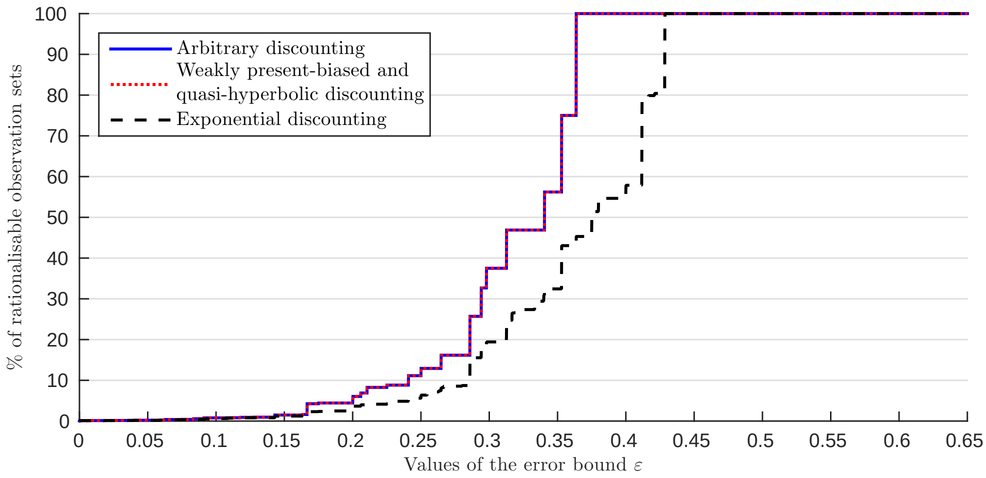
	Arbitrary utility function	Log-concave utility function	Power utility function
Arbitrary discounting	1	.0439	.0060
Weakly present-biased and quasi-hyperbolic discounting	1	.0408	.0043
Exponential discounting	0.9590	.0303	.0025

Table 3: Probability of rationalising a random set of observations generated in the Kirby and Petry (2004) experiment by a discounted utility model once we allow to drop at most one observation, conditional on the set being rationalisable in the above sense by a mixed-monotone utility function (i.e., with at most one observations dropped).

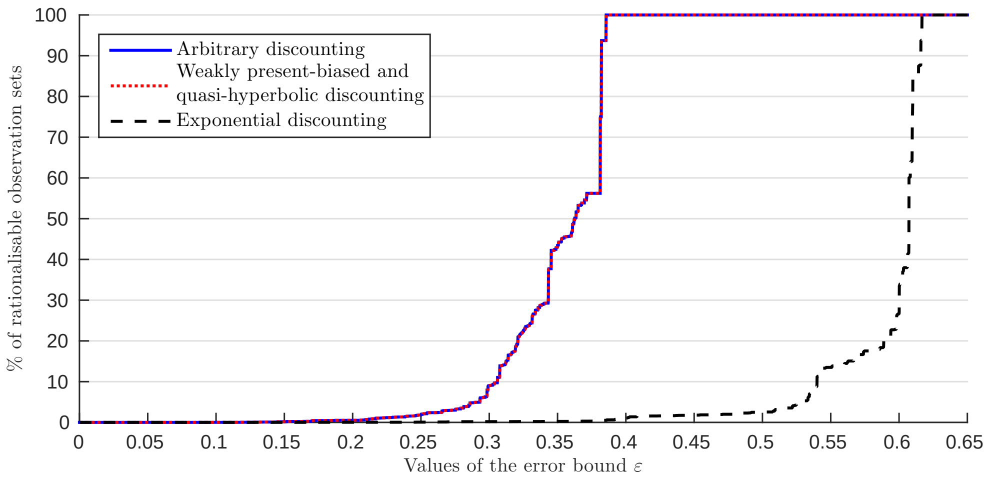
Finally, we display the p-values from Anderson-Darling and Mann-Whitney U tests corresponding to the analysis presented in Figure 4 of the paper (see Section 5.4).

	Arbitrary discounting			Exponential discounting		
Compared groups	Arbitrary utility	Log-concave utility	Power utility	Arbitrary utility	Log-concave utility	Power utility
Control group vs. alcohol-abusers	.6834 .5496	.8268 .9742	.6058 .4487	.4250 .4487	.2677 .2072	.1843 .3282
Control group vs. drug-abusers	.0002 .0045	.0031 .0187	.5806 .5417	.0041 .0299	.0149 .0256	.0038 .0614
Alcohol-abusers vs. drug-abusers	.0025 .0231	.0015 .0098	.7649 .9037	.0194 .0667	.3652 .3043	.3081 .3559

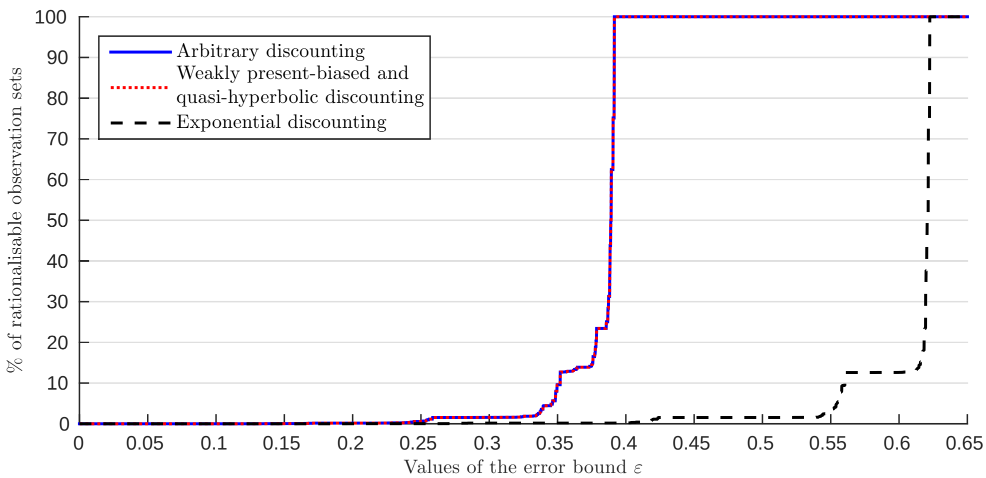
Table 4: P-values from Anderson-Darling (the upper value) and Mann-Whitney U tests (the lower value) comparing distributions presented in Figure 4, for each pair of groups considered in Section 5.4 and each specification of the discounted utility model.



(a) Rationalisation for an arbitrary instantaneous utility function.



(b) Rationalisation for a log-concave instantaneous utility function.



(c) Rationalisation for a power instantaneous utility function.

Figure 5: Cumulative distributions of a sample of 100 000 random observation sets for various specifications of the discounted utility model, with respect to the error bound ε.

References

ANDERSEN, S., G. W. HARRISON, M. I. LAU, AND E. E. RUTSTRÖM (2008): "Eliciting risk and time preferences," *Econometrica*, 76(3), 583–618.

——— (2013): "Discounting behaviour and the magnitude effect: Evidence from a field experiment in Denmark," *Economica*, 80(320), 670–697.

——— (2014): "Discounting behavior: A reconsideration," *European Economic Review*, 71(C), 15–33.

BENHABIB, J., A. BISIN, AND A. SCHOTTER (2010): "Present-bias, quasi-hyperbolic discounting, and fixed costs," *Games and Economic Behavior*, 69(2), 205–223.

BLEICHRODT, H., K. I. ROHDE, AND P. P. WAKKER (2009): "Non-hyperbolic time inconsistency," *Games and Economic Behavior*, 66(1), 27–38.

BORDER, K. C. (1992): "Revealed preference, stochastic dominance, and the expected utility hypothesis," *Journal of Economic Theory*, 56, 20–42.

BRONARS, S. G. (1987): "The power of nonparametric tests of preference maximization: The nonparametric approach to demand analysis," *Econometrica*, 55(3), 693–98.

CHABRIS, C. F., D. LAIBSON, C. L. MORRIS, J. P. SCHULDT, AND D. TAUBINSKY (2008): "Individual laboratory-measured discount rates predict field behavior," *Journal of Risk and Uncertainty*, 37(2), 237–269.

——— (2009): "The allocation of time in decision-making," *Journal of the European Economic Association*, 7(2-3), 628–637.

COLLER, M., G. W. HARRISON, AND E. E. RUTSTRÖM (2012): "Latent process heterogeneity in discounting behavior," *Oxford Economic Papers*, 64(2), 375–391.

COLLER, M., AND M. WILLIAMS (1999): "Eliciting individual discount rates," *Experimental Economics*, 2(2), 107–127.

DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2012): "Interpreting time horizon effects in inter-temporal choice," Discussion Paper No. 6385, IZA.

ECHENIQUE, F., T. IMAI, AND K. SAITO (2014): "Testable implications of quasi-hyperbolic and exponential discounting," Caltech SS Working Paper 1388.

Fishburn, P. C., and A. Rubinstein (1982): "Time preference," *International Economic Review*, 23(3), 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), 351–401.

Harrison, G. W., M. I. Lau, and M. B. Williams (2002): "Estimating individual discount rates in Denmark: A field experiment," *American Economic Review*, 92(5), 1606–1617.

Husseinov, F. (2010): "Monotonic extension," Departmental Working Papers 1004, Bilkent University, Department of Economics.

Kirby, K. N. (2009): "One-year temporal stability of delay-discount rates," *Psychonomic Bulletin & Review*, 16, 457–462.

Kirby, K. N., and N. M. Petry (2004): "Heroin and cocaine abusers have higher discount rates for delayed rewards than alcoholics or non-drug-using controls," *Addiction*, 99, 461–471.

Kirby, K. N., N. M. Petry, and W. K. Bickel (1999): "Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls," *Journal of Experimental Psychology: General*, 128, 78–87.

Kraft, C. H., J. W. Pratt, and A. Seidenberg (1959): "Intuitive probability on finite sets," *Annals of Mathematical Statistics*, 30, 408–419.

Luenberger, D. (1969): *Vector Space Optimization*. Wiley, New York.

Nishimura, H., E. A. Ok, and J. K.-H. Quah (2013): "A unified approach to revealed preference theory: The case of rational choice," Economics Series Working Papers 686, University of Oxford, Department of Economics.

Noor, J. (2011): "Intertemporal choice and the magnitude effect," *Games and Economic Behavior*, 72(1), 255–270.

Ok, E. A., and Y. Masatlioglu (2007): "A theory of (relative) discounting," *Journal of Economic Theory*, 137(1), 214–245.

PENDER, J. L. (1996): "Discount rates and credit markets: Theory and evidence from rural India," *Journal of Development Economics*, 50(2), 257–296.

SCOTT, D. (1964): "Measurement structures and linear inequalities," *Journal of Mathematical Psychology*, 1, 233–247.

STOER, J., AND C. WITZGALL (1970): *Convexity and optimization in finite dimensions I*. Springer Verlag, Berlin.