University of Leicester

DEPARTMENT OF ECONOMICS

# A GENERAL THEORY OF TIME DISCOUNTING: THE REFERENCE-TIME THEORY OF INTERTEMPORAL CHOICE

**Ali al-Nowaihi, University of Leicester, UK**

**Sanjit Dhami, University of Leicester, UK**

**Working Paper No. 08/34**
**September 2008**

# A general theory of time discounting: The reference-time theory of intertemporal choice

**Ali al-Nowaihi*** **Sanjit Dhami†**

9 September 2008

**Abstract**

We develop a general theory of intertemporal choice: the reference-time theory, RT. RT is a synthesis of ideas from the hyperbolic model and subadditivity of time discounting. These models are extended to allow for a reference point for time as well as wealth. RT is able to account for all the 6 main anomalies of time discounting: gain-loss asymmetry, magnitude effect, common difference effect, delay-speedup asymmetry, apparent intransitivity of time preferences, and non-additivity of time discounting. We provide a class of utility functions compatible with RT. We show how RT can be extended to incorporate uncertainty and attribute models of intertemporal choice.

*Keywords: Anomalies of the discounted utility model, Hyperbolic discounting, Prospect theory, $\gamma$-delay, $\alpha$-subadditivity, SIE value functions.*

*JEL Classification Codes: C60(General: Mathematical methods and programming); D91(Intertemporal consumer choice).*

*Department of Economics, University of Leicester, University Road, Leicester. LE1 7RH, UK. Phone: +44-116-2522898. Fax: +44-116-2522908. E-mail: aa10@le.ac.uk.

†Department of Economics, University of Leicester, University Road, Leicester. LE1 7RH, UK. Phone: +44-116-2522086. Fax: +44-116-2522908. E-mail: Sanjit.Dhami@le.ac.uk.

# 1. Introduction

The standard model of intertemporal choice, exponentially discounted utility (EDU), is contradicted by a large body of empirical evidence. See, for example, Loewenstein and Prelec (1992), Frederick et al. (2002) and Manzini and Mariotti (2008). Moreover, these anomalies are not simply mistakes; see Frederick et al. (2002), section 4.3. To develop models that provide a better explanation of economic behavior over time, it is imperative to take account of these anomalies. Subsection 1.1, below, outlines the anomalies. Subsection 1.2 overviews the current explanations. Subsection 1.3 outlines our approach.

## 1.1. Anomalies of intertemporal choice

Loewenstein and Prelec (1992), henceforth 'LP', described the following four anomalies, all with good empirical support:

1. *Gain-loss asymmetry (or sign effect).* Subjects in a study by Loewenstein (1988b) were, on average, indifferent between receiving \$10 immediately and receiving \$21 one year later (an implied discount rate of 74% per annum[1]). They were also indifferent between loosing \$10 immediately and losing \$15 dollars one year later (an implied discount rate of 40.5% per annum). This is inconsistent with EDU because the implied discount rates are different and they are both too high (even allowing for capital market imperfections and liquidity constraints).

2. *Magnitude effect.* Thaler (1981) reported that subjects were, on average, indifferent between receiving \$15 immediately and \$60 one year later (an implied discount rate of 139% per annum). They were also indifferent between receiving \$3000 immediately and receiving \$4000 one year later (an implied discount rate of 29% per annum). This contradicts EDU in that the implied discount rate is magnitude dependent and is too high.

3. *Common difference effect*[2]. Thaler (1981): A person might prefer one apple today

[1]The estimated discount factor $D^e$ is the ratio of current to future reward times the ratio of marginal utilities. Assuming that the marginal utilities are approximately the same, the ratio of rewards is simply used to approximate $D^e$. Thus, in this case, $D^e = \frac{10}{21} = 0.476\,19$. Assuming continuous compounding, $D^e \approx D = e^{-\theta}$, where $\theta$ is the discount rate. Taking logs on both sides, $\theta = -\ln D^e$, which in this case is $-\ln(0.476\,19) = 0.741\,94$, as claimed. The same method is used to report the other discount rates in experiments, below.

[2]The *common difference effect* is also known as the *delay effect* or *preference reversal*. However, 'preference reversal' is also used to describe (for example) the following situation: A consumer strictly prefers bundle $x$ to bundle $y$. But when a new bundle, $z$, is introduced, the consumer now strictly prefers $y$ to $x$. The latter phenomenon is a framing effect and leads to violation of independence. The 'common difference effect' need not be due to framing nor need it lead to a violation of independence. On this point, see Manzini and Mariotti (2008, subsection 3.1).

to two apples tomorrow (an implied real discount rate of 25300% per annum), but at the same time prefer two apples in 51 days to one apple in 50 days (there is no evidence that this is due to an expected change in the real discount rate).[3] This is a violation of a fundamental principle of EDU, namely, that discounting only depends on the time interval between two rewards.

4. *Delay-speedup asymmetry.* Loewenstein (1988a) reported that, in general, the amount required to compensate for delay in receiving a real reward by a given interval, from $s$ to $s+t$, was two to four times greater than what the subjects were willing to sacrifice to bring consumption forward from $s+t$ to $s$. (Neo)classically, these quantities should be the same.

   Recent scholarship has added two further anomalies:

5. *Non-additivity of time discounting.* Discounting from time $t$ back to time $s$ then further back to time $r$ is not the same as discounting from time $t$ back to time $r$ in one step (Read, 2001 and Scholten and Read, 2006a).

6. *Intransitivity of time preferences.* The following cycles have been observed: $x$ received at time $r$ is preferred to $y$ received at time $s$, which is preferred to $z$ received at time $t$, which is preferred to $x$ received at time $r$ (Roelofsma and Read, 2000).

### 1.2. Explaining the anomalies

We very briefly summarize some attempts at explaining anomalies 1-6. These attempts will be discussed in greater depth later in the paper.

The main explanations of the magnitude effect and the gain-loss asymmetry are due to Loewenstein and Prelec (1992) (henceforth, 'LP'). LP explained the magnitude effect through a value function with increasing elasticity. This makes higher magnitudes more salient. They explained gain-loss asymmetry by adopting a value function with greater elasticity for losses than for gains, which makes losses more salient. We are not aware of alternative generally accepted explanations of these two anomalies.

There are several competing explanations of the common difference effect (CDE). The CDE is explained in LP by the notion of *declining impatience* (roughly, one is more impatient as the date of the reward approaches). Hence, despite identical intervals separating two time-outcome pairs, the choice among the two depends on how close to the current period they are. LP also provided an axiomatic derivation of their *generalized hyperbolic discount function*, which we shall call the *LP-discount function*. A similar explanation

[3] However, Andersen et al (2007), by jointly eliciting attitudes to risk and time, estimate much lower discount rates, of about 10.1%.

of the CDE is provided in the quasi-hyperbolic discounting model of Phelps and Pollak (1968) and Laibson (1997), 'PPL' for short.[4]

The experimental work of Read (2001) confirmed the CDE but rejected declining impatience in favour of constant impatience and, hence, rejected the LP (and PPL) explanation of the common difference effect in terms of declining impatience. Read (2001) and Scholten and Read (2006a) found experimental evidence for subadditivity and introduced the concept of an *interval discount function*[5]. Furthermore, based on empirical evidence, Scholten and Read (2006a) developed a specific interval discount function, which we shall call the *RS-discount function*. The RS-discount function (depending on parameter values) can explain the common difference effect as due to either declining impatience, subadditivity or a combination of both. We shall refer to this work as collectively RRS.[6]

Manzini and Mariotti (2006), in their 'theory of vague time preferences', develop an attribute model that can explain the common difference effect; we review this in more detail in subsection 5.3, below.

Under uncertainty, the common difference effect never arises when we use expected utility with exponential discounting. Hence, it is quite possible that the experimental finding of the common difference effect is a rejection of expected utility rather than exponential discounting. Halevy (2007) shows that when non-expected utility is combined with exponential discounting, the theory is consistent with the presence of a common difference effect, provided uncertainty is present but sufficiently small (see subsection 5.2, below).

The main explanation of non-additivity of time discounting is through the work of RRS and, in particular, using the interval discount function of Scholten and Read (2006a). Intransitivity of time preferences can be incorporated either through a variety of attribute models (Rubinstein, 2003; Manzini and Mariotti, 2006). Yet another explanation relies on models that do not assume transitivity of preferences in the first place, such as Ok and Masatlioglu (2007); see subsection 5.4, below. However, it cannot account for either gain-loss asymmetry or delay-speedup asymmetry.

A problem, however, is that there is no single model that can explain all the anomalies.

### 1.3. Towards a general theory of intertemporal choice: The reference-time theory (RT)

Within the class of time preference models which are separable in time and outcomes (also known as delay-discounting models), what should a general theory of time preference

[4]Hyperbolic discounting alone, however, cannot explain anomalies 5 and 6.

[5]An interval discount function, $D(r,t)$, $0 \leq r \leq t$, discounts a quantity from time $t$ back to time $r$. If $D$ is additive (and only if $D$ is additive), then $D(r,t) = D(r,s)\,D(s,t)$, for all $s$, $r \leq s \leq t$.

[6]The justification for the acronym RRS is that in related work, Roelofsman and Read (2000) provide experimental evidence for intransitivity.

aspire to? We suggest two desirable elements. First, it should be able to explain anomalies 1-6. Second, it should provide a framework that can incorporate recent developments in time discounting such as uncertainty (Halevy, 2007) and attribute models (Manzini and Mariotti, 2006 and Scholten and Read, 2006b).

The aim of this paper is to develop a theory of intertemporal choice that incorporates the two desirable elements mentioned above. We call this theory the *reference-time* theory of intertemporal choice, 'RT' for short. It is a synthesis of three important works, namely, LP, PPL and RRS. In a nutshell, RT is basically LP extended to allow for non-additive time discounting by incorporating a reference point for time. Using RT, we explain the anomalies 1-6 as follows.

1. We follow LP in explaining the gain-loss asymmetry (anomaly 1) by assuming that the elasticity of the value function for losses exceeds the elasticity for gains. This allows us to use the same discount function for gains and losses, in agreement with strict separability of time and outcomes.

2. Also like LP, we explain the magnitude effect (anomaly 2) by assuming that the elasticity of the value function is increasing. However, al-Nowaihi and Dhami (2008) show that several popular classes of utility functions violate this condition. They propose the class of *simple increasing elasticity* value functions (SIE) that is tractible and consistent with LP's explanation of the magnitude effect. Furthermore, it is compatible with any theory where preferences are separable in time and outcomes. The SIE value function also satisfies LP's requirement that the elasticity of the value function for losses exceeds the elasticity for gains in order to explain the sign effect. We follow al-Nowaihi and Dhami (2008) in using the SIE value function and integrating it within the RT theory.

3. LP provided an axiomatic derivation of their generalized hyperbolic discount function (which we call the LP-discount function). For this, they added the extra assumption of *linear delay* to that of the common difference effect. While there is considerable empirical evidence for the common difference effect, the assumption of linear delay is added purely for convenience. We extend the LP derivation as follows. At the most general level, which requires neither linear delay nor the common difference effect, we have our Representation Theorem 2 (Proposition 12, below). We introduce a weaker notion of subadditivity, which we call $\alpha$-subadditivity (Definition 12, below). According to our Characterization Theorem 4 (Proposition 21, below), preferences exhibit the common difference effect if, and only if, $\alpha$-subadditivity holds. We also introduce a generalization of the concept of linear delay of LP. We call this $\gamma$-delay. Our Proposition 23, below, then shows that $\gamma$-delay implies the common difference

effect. Imposing additivity, as well as $\gamma$-delay, gives our Proposition 24, below. The special case of the latter with $\gamma = 1$ gives the LP-discount function. Our more general approach also allows us to derive the RS-discount function (Proposition 25, below). In particular, as with RRS, we can explain the common difference effect (anomaly 3) as due to either declining impatience, subadditivity or a combination of both. However, our approach is more general, as RT can also explain the common difference effect as due to the presence of a small amount of irremovable uncertainty, as in Halevy (2007), or as a consequence of multiple decision criteria, as in Manzini and Mariotti (2006) (see section 5, below). Thus, RT can accommodate all the known explanations of the common difference effect. In the spirit of RRS, we leave it to empirical evidence to select the correct explanation.

4. We show the delay-speedup asymmetry (anomaly 4) follows from our other assumptions (see Proposition 29, below).

5. We follow LP in adopting prospect theory (Kahneman and Tversky, 1979 and Tversky and Kahneman, 1992) as our underlying decision theory. But, in addition to a reference point for wealth, we introduce a reference point for time. If preferences are additive, as in LP, then the choice of the reference point for time is immaterial (Proposition 7, below). However, if preferences are non-additive, then the choice of the reference point for time matters. Thus, we can accommodate non-additive preferences (anomaly 5) by having a reference point for time. A consequence is that all our discount functions are interval discount functions, as in RRS.

6. Given a reference point for wealth, $w_0$, and a reference point for time, $r$, our preferences are complete and transitive (subsection 2.2, below). Thus they may be called *conditionally complete* and *conditionally transitive* (conditional on $w_0$ and $r$). We explain observed intransitivity as due to a change in the reference point for time (see subsections 2.3 and 2.7, below). This is in contrast to Ok and Masatlioglu (subsection 5.4, below), where preferences are complete but intransitive (in our terminology we may describe such preferences as *unconditionally* complete but not even conditionally transitive). Thus the relative-discounting theory Ok and Masatlioglu and the reference-time theory of this paper are not compatible and neither is a special case of the other.

To summarize, the theory presented in this paper (reference-time theory or RT) can explain anomalies 1-6 (section, 4, below) and can be extended to incorporate uncertainty, as in Halevy (2007), and the attribute models of Manzini and Mariotti (2006) and Scholten and Read (2006b) (section 5, below).

All proofs are contained in the appendix.

## 2. A reference-time theory of intertemporal choice (RT)

This section is structured as follows. We first outline prospect theory (subsection 2.1), which is an essential building block of RT. We define preferences for RT as early as possible (subsection 2.2). This is followed by further essential material on discount functions, additivity, impatience, intransitivity and the common difference effect (subsections 2.4, 2.5, 2.6, 2.7 and 2.8). The main technical machinery: representation, extension and characterization theorems are developed in subsections 2.9, 2.10 and 2.11. Finally, in subsection 2.12 we are in a position to extend LP to allow for non-additive time discounting by incorporating a reference point for time.

### 2.1. Prospect theory

We follow LP in using prospect theory (Kahneman and Tversky, 1979, and Tversky and Kahneman, 1992) rather than standard utility theory. Prospect theory distinguishes between two phases of decision making: *editing* and *evaluation*.

In the editing phase, a decision maker simplifies a real world problem to make it amenable to formal analysis and reduce the associated cognitive load. As part of the editing phase a *reference point* is chosen to which outcomes are to be compared.

In the evaluation phase, a value (a real number) is attached to each feasible action by the decision maker. The action with highest value is chosen. The function, $v$, that assigns values to actions in prospect theory is called the *value function*, it is the analogue of the indirect utility function of standard utility theory. In standard utility theory carriers of utility are the *outcomes* of actions. But in prospect theory carriers of utility are *deviations* in outcomes from the reference point. In general, the action chosen in the evaluation phase will depend on the reference point chosen in the editing phase.

The value function, $v$, in prospect theory has four main properties: *reference dependence*, *monotonicity*, *declining sensitivity*, *loss aversion*. Furthermore, in prospect theory, there is *non-linear transformation of probabilities*. There is good empirical support for these feature; see, for instance, Kahneman and Tversky (2000).

We take $v$ to be the value function introduced by Kahneman and Tversky (1979). Thus $v$ satisfies:

$$v : (-\infty, \infty) \to (-\infty, \infty) \text{ is continuous, strictly increasing (monotonicity).} \tag{2.1}$$

$$v(0) = 0 \text{ (reference dependence) and is twice differentiable except at 0.} \tag{2.2}$$

$$\text{For } x > 0\text{: } -v(-x) > v(x) \text{ (loss aversion).} \tag{2.3}$$

Following LP, we define the elasticity of the value function as follows.

**Definition 1** : *The elasticity of* $v$ *is*

$$\epsilon_v(x) = \frac{x}{v}\frac{dv}{dx},\ x \neq 0. \tag{2.4}$$

### 2.2. Preferences

We consider a decision maker who, at time $t_0$, takes an action that results in the level of wealth $w_i$ at time $t_i$, $i = 1, 2, ..., n$, where

$$t_0 \leq r \leq t_1 < ... < t_n. \tag{2.5}$$

Time $r$ is the *reference time*: the time back to which all values are to be discounted. We can choose any moment of time as time zero and measure all other times relative to it. We choose to set $t_0 = 0$, i.e., the time a decision is made is always time $t = 0$. If it is desired to set $t_0 \neq 0$, then simply replace all times, $t$, below, with $t - t_0$. The decision maker's intertemporal utility function is given by:

$$V_r((x_1, t_1), (x_2, t_2), ..., (x_n, t_n)) = \Sigma_{i=1}^{n} v(x_i) D(r, t_i), \tag{2.6}$$

where $v$ is the value function introduced by Kahneman and Tversky (1979). $x_i = w_i - w_0$ is the difference between the wealth level, $w_i$, at time, $t_i$, and the reference level for wealth, $w_0$. $D(r, t_i)$ is the discount function, it discounts $v(x_i)$ from time, $t_i$, back to the reference time, $r$.

More formally, we assume that, for each $(w_0, r) \in (-\infty, \infty) \times [0, \infty)$, the decision maker has a complete transitive preference relation, $\preceq_{w_0,r}$ on $(-\infty, \infty) \times [r, \infty)$. We think of $w_0$ as the reference point for wealth and $r$ the reference point for time. If $(w, t) \in (-\infty, \infty) \times [r, \infty)$, with $w \geq w_0$, we say that $(w, t)$ is an outcome in the domain of gains. If $(w, t) \in (-\infty, \infty) \times [r, \infty)$, with $w < w_0$, we say that $(w, t)$ is an outcome in the domain of losses. We assume that $\preceq_{w_0,r}$ is represented by a utility function $v(w - w_0) D(r, t)$. Thus $(w_1, t_1) \preceq_{w_0,r} (w_2, t_2)$ if, and only if, $v(w_1 - w_0) D(r, t_1) \leq v(w_2 - w_0) D(r, t_2)$.

Using (2.6), we extend $\preceq_{w_0,r}$ to a complete transitive preference relation on sequences from $(-\infty, \infty) \times [r, \infty)$, as follows[7]:

$$\begin{aligned} &((x_1, s_1), (x_2, s_2), ..., (x_m, s_m)) \preceq_{w_0,r} ((y_1, t_1), (y_2, t_2), ..., (y_n, t_n)) \\ &\Leftrightarrow V_r((x_1, s_1), (x_2, s_2), ..., (x_m, s_m)) \leq V_r((y_1, t_1), (y_2, t_2), ..., (y_n, t_n)) \end{aligned} \tag{2.7}$$

We depart from LP in the following ways. LP have a reference point for wealth but not for time. We have a reference point for wealth, $w_0$, *and* a reference point for time, $r$. LP implicitly assume that the discount function is additive (Definition 5, below). We allow

[7]This also holds for infinite sequences, provided the sum in (2.6) is convergent.

the discount function to be non-additive, to accommodate the empirical evidence of RRS. If the discount function is additive (as in LP), then the choice of the reference point for time is irrelevant, since $(x,s) \preceq_{w_0,r} (y,t)$ if, and only if, $(x,s) \preceq_{w_0,0} (y,t)$ (Proposition 7, below). However, if the discount function is non-additive, then the choice of the reference point for time matters. We use this to explain (apparent) intransitivity as a framing effect due to a change in the reference point for time.

### 2.3. Determination of the reference point for time

Let $S$ be a non-empty set of sequences from $(-\infty,\infty)\times[0,\infty)$. Suppose a decision maker is interested in comparing the members of $S$. For example, for the purpose of choosing the optimal member (if $S$ is compact). For this he needs a reference point for time. Let $T$ be the set of times involved, i.e.,

$$T=\{t\in[0,\infty): t=t_i \text{ for some sequence } \{(x_1,t_1),(x_2,t_2),...,(x_i,t_i),...\} \text{ in } S\}. \quad (2.8)$$

Since $T$ is bounded below (by 0) and non-empty, it follows that $T$ has a greatest lower bound, $r$. We make the following tentative assumption:

**A0** *Reference time.* Given $S$, $T$, $r$, as described just above, we assume that the decision maker takes $r$ as the reference point for time.

For example, if a decision maker wants to compare $x$ received at time $s$ with $y$ received at time $t$, $s\leq t$, then $S$ consists of just two sequences, each with just one element: $S=\{(x,s),(y,t)\}$ and $T=\{s,t\}$. Thus A0 implies that $r=s$. If $v(x)<v(y)\,D(s,t)$ then the decision maker chooses $(y,t)$ over $(x,s)$.

A0 does not have the status of the LP assumptions A1-A4, introduced in subsection 2.12, below. While there is considerable, though debated, empirical evidence for A1-A4, A0 should be regarded as a tentative assumption, whose implications are to be explored. We will only use A0 in subsections 2.7 and 2.12. In subsection 2.7, we use A0 to explain (apparent) intransitivity as due to a shift in the reference point for time. In subsection 2.12, we use A0 to prove that assumption A4 (Delay-speedup asymmetry) follows from the other assumptions (Proposition 29).

### 2.4. Discount functions

The five discount functions that will be important for this paper are:

$$\textbf{Exponential}:\ D(r,t)=e^{-\beta(t-r)},\ \beta>0. \quad (2.9)$$

$$\textbf{PPL}:\ D(r,t)=\left\{\begin{array}{ccc} 1 & \text{when} & r=t=0 \\ e^{-(\delta+\beta t)} & \text{when} & r=0,t>0 \\ e^{-\beta(t-r)} & \text{when} & 0<r\leq t \end{array}\right\};\ \beta>0,\ \delta>0. \quad (2.10)$$

$$\textbf{LP}\text{: } D(r,t) = \left(\frac{1+\alpha t}{1+\alpha r}\right)^{-\frac{\beta}{\alpha}},\ t \geq 0,\ r \geq 0,\ \alpha > 0,\ \beta > 0. \tag{2.11}$$

$$\textbf{RS}\text{: } D(r,t) = \left[1+\alpha\left(t^{\tau}-r^{\tau}\right)^{\rho}\right]^{-\frac{\beta}{\alpha}},\ 0 \leq r \leq t,\ \alpha > 0,\ \beta > 0,\ \rho > 0,\ \tau > 0. \tag{2.12}$$

$$\textbf{generalized RS}\text{: } D(r,t) = e^{-Q[w(t)-w(r)]},\ 0 \leq r \leq t, \text{ where}$$
$$Q : [0,\infty) \overset{\text{onto}}{\to} [0,\infty) \text{ and } w : [0,\infty) \overset{\text{onto}}{\to} [0,\infty) \text{ are strictly increasing.} \tag{2.13}$$

The exponential discount function (2.9) was introduced by Samuelson (1937). Aside from its tractability, the main attraction of EDU is that it leads to time-consistent choices. If the plan $(x_1,t_1),(x_2,t_2),...,(x_n,t_n)$ is optimal at time 0, then at time $t_k$ the plan $(x_{k+1},t_{k+1}),(x_{k+2},t_{k+2}),...,(x_n,t_n)$ is also optimal. But this may no longer be true for more general specifications of the discount function.

The $\beta\delta$ (or quasi-hyperbolic) discount function (2.10) was proposed by Phelps and Pollak (1968) and Laibson (1997). The generalized hyperbolic discount function (2.11) was proposed by Loewenstein and Prelec (1992). For the special case, $\alpha = \beta$, (2.11) it reduces to the hyperbolic discount function. These three discount functions are *additive* (Definition 3, below). They can account for the common difference effect through declining impatience (Definition 6, below) but they cannot account for either non-additivity or intransitivity.

The interval discount function (2.12) was introduced by Scholten and Read (2006a). It can account for both non-additivity and intransitivity. It can account for the common difference effect though declining impatience, subadditivity or a combination of both (subsections 2.7 and 2.8, below).

In subsection 5.1, below, we shall show that the attribute model of Scholten and Read (2006b) is equivalent to a discounted utility model with the discount function (2.13), which is a generalization of their RS-discount function (2.12).

Note that (2.11) approaches (2.9) as $\alpha \to 0$. In general, neither of (2.11) or (2.12) is a special case of the other. However, for $r = 0$ (and only for $r = 0$), (2.12) reduces to (2.11) when $\rho = \tau = 1$. Scholten and Read (2006a) report incorrectly that the LP-discount function is a special case of the RS-discount function. One needs to restrict $r = 0$ (in addition to $\rho = \tau = 1$) in order to generate the LP from the RS-discount function. While $\rho, \tau$ are parameters, $r$ is a variable. Hence, neither discount function is a special case of the other.

**Definition 2** *(Discount functions): Let*

$$\Delta = \{(r,t) \in \mathbb{R} \times \mathbb{R} : 0 \leq r \leq t\}. \tag{2.14}$$

*A discount function is a mapping, $D : \Delta \to (0,1]$, satisfying:*
*(i) For each $r \in [0,\infty)$, $D(r,t)$ is a strictly decreasing function of $t \in [r,\infty)$ into $(0,1]$*

*with* $D(r,r) = 1$.
*(ii) For each* $t \in [0,\infty)$, $D(r,t)$ *is a strictly increasing function of* $r \in [0,t]$ *into* $(0,1]$.
*Furthermore, if D satisfies (i) with 'into' replaced with 'onto', then we call D a continuous discount function.*

Our terminology suggests that a continuous discount function is continuous. That this is partly true, is established by the following Proposition.

**Proposition 1** *: A continuous discount function,* $D(r,t)$*, is continuous in* $t$.

**Proposition 2** *: Each of (2.9), (2.11) and (2.12) is a continuous discount function in the sense of Definition 2. However, (2.10) is a discount function but not a continuous discount function.*

The reason that (2.10) fails to be a *continuous* discount function is that $\lim_{t\to 0+} D(0,t) = e^{-\delta} < 1 = D(0,0)$.

From (2.11) and (2.12) we see that the restrictions $r \geq 0$ and $t \geq 0$ are needed. From (2.12) we see that the further restriction $r \leq t$ is needed.[8] From (2.9) we see that the 'into' in Definition 2(ii) cannot be strengthened to 'onto'.

**Proposition 3** *(Time sensitivity*[9]*): Let D be a continuous discount function. Suppose* $r \geq 0$. *If* $0 < x \leq y$, *or if* $y \leq x < 0$, *then* $v(x) = v(y)\,D(r,t)$ *for some* $t \in [r,\infty)$.

**Proposition 4** *(Existence of present values): Let D be a discount function. Let* $r \leq t$. *Let* $y \geq 0$ $(y \leq 0)$. *Then, for some* $x$, $0 \leq x \leq y$ $(y \leq x \leq 0)$, $v(x) = v(y)\,D(r,t)$.

### 2.5. Additivity

We now define additivity and related concepts.

**Definition 3** *(Additivity): A discount function,* $D(r,t)$*, is*

$$\textit{additive if } D(r,s)\,D(s,t) = D(r,t)\,,\ \textit{for } r \leq s \leq t, \tag{2.15}$$

$$\textit{subadditive if } D(r,s)\,D(s,t) < D(r,t)\,,\ \textit{for } r < s < t, \tag{2.16}$$

$$\textit{superadditive if } D(r,s)\,D(s,t) > D(r,t)\,,\ \textit{for } r < s < t. \tag{2.17}$$

[8] One alternative is to *define* $D(t,s)$ to be $1/D(s,t)$. But we do not know if people, when compounding forward, use the inverse of discount function (as they should, from a normative point of view). Fortunately, we have no need to resolve these issues in this paper.

[9] We have chosen the phrase 'time sensitivity' to conform with the terminology of OM, Axiom A1, p219, and Claim 3, p235.

Additivity (2.15) implies that discounting a quantity from time $t$ back to time $s$ and then further back to time $r$ is the same as discounting that quantity from time $t$ back to time $r$ in one step.

To aid further development, we define a generating function, whose interpretation will become apparent from Proposition 5 that follows the definition.

**Definition 4** *(The generating function): Let $\varphi : [0, \infty) \to (0, 1]$ be a strictly decreasing function with $\varphi(0) = 1$. Then we call $\varphi$ a generating function. If, in addition, $\varphi$ is onto, we call $\varphi$ a continuous generating function.*

A 'continuous generating function' is continuous. The proof is the same as that of Proposition 1 and, therefore, will be omitted.

**Proposition 5** *(Additive extension theorem):*
*(a) Let $D(r,t) = [\varphi(r)]^{-1} \varphi(t)$ for some strictly decreasing real valued function, $\varphi : [0, \infty) \to (0, 1]$. Then the following hold:*
*(i) $D$ is an additive discount function.*
*(ii) If $\varphi(0) = 1$ (so that $\varphi$ is a generating function), then $D(0,t) = \varphi(t)$.*
*(iii) If $\varphi$ is onto (so that $\varphi$ is a continuous generating function), then $D$ is an additive continuous discount function and $D(0,t) = \varphi(t)$.*
*(b) Let $D$ be an additive discount function. Then the following hold:*
*(i) $D(r,t) = [\varphi(r)]^{-1} \varphi(t)$ for some strictly decreasing real valued function, $\varphi : [0, \infty) \to (0, 1]$ with $\varphi(0) = 1$ (hence, $\varphi$ is a generating function).*
*(ii) If $D$ is a continuous discount function, then $\varphi$ is onto (hence, $\varphi$ is a continuous generating function).*
*(iii) $D(0,t) = \varphi(t)$.*

Proposition 5 justifies the following definition.

**Definition 5** *(Additive extensions): Let $\varphi$ be a generating function. Let $D(r,t) = [\varphi(r)]^{-1} \varphi(t)$. Then:*
*(i) We call $\varphi$ the generating function of the additive discount function, $D$.*
*(ii) We call $D$ the additive extension of $\varphi$.*
*(iii) We also refer to $D(r,t)$ as the additive extension of $D(0,t)$.*

**Proposition 6** *: The discount functions (2.9), (2.10) and (2.11) are additive. In each case, $D(r,t)$ is the additive extension of $D(0,t)$ and $\varphi(t) = D(0,t)$ is the generating function for $D(r,t)$. However, (2.12) is not additive. (2.13) is additive if, and only if, $Q[w(t) - w(r)] = Q[w(t)] - Q[w(r)]$, in which case $D(r,t) = e^{Q[w(r)] - Q[w(t)]}$ is the additive extension of $D(0,t) = e^{-Q[w(t)]}$ and $e^{-Q[w(t)]}$ is the generating function.*

**Proposition 7** *(Invariance to the choice of reference time) If $D(r,t)$ is additive, then*

$$\begin{aligned}&((x_1,s_1),(x_2,s_2),...,(x_m,s_m)) \preceq_{w_0,r} ((y_1,t_1),(y_2,t_2),...,(y_n,t_n))\\ \Leftrightarrow\ &((x_1,s_1),(x_2,s_2),...,(x_m,s_m)) \preceq_{w_0,0} ((y_1,t_1),(y_2,t_2),...,(y_n,t_n)).\end{aligned}$$

Thus, if the discount function is additive (as is the case with LP) then the choice of the reference time, $r$, back to which all utilities are discounted, is irrelevant. Discounting back to time $r$ is equivalent to discounting back to time 0.

### 2.6. Impatience

The following concepts are also useful.

**Definition 6** *(Impatience): A discount function, $D(r,s)$, exhibits*[10]

$$\textit{declining impatience if } D(r,s) < D(r+t,s+t), \textit{ for } t>0 \textit{ and } r<s, \tag{2.18}$$

$$\textit{constant impatience if } D(r,s) = D(r+t,s+t), \textit{ for } t\geq 0 \textit{ and } r\leq s, \tag{2.19}$$

$$\textit{increasing impatience if } D(r,s) > D(r+t,s+t), \textit{ for } t>0 \textit{ and } r<s. \tag{2.20}$$

**Proposition 8** *: Let $D(r,t)$ be the RS-discount function (2.12), then:*
*(a) If $0<\rho\leq 1$, then $D$ is subadditive.*
*(b) If $\rho>1$, then $D$ is neither subadditive, additive nor superadditive.*
*(c) (i) If $0<\tau<1$, then $D$ exhibits declining impatience.*
*(ii) If $\tau=1$, then $D$ exhibits constant impatience.*
*(iii) If $\tau>1$, then $D$ exhibits increasing impatience.*

Scholten and Read (2006a), bottom of p1425, state: $\alpha>0$ implies subadditivity (incorrect), $\rho>1$ implies superadditivity (incorrect) and $0<\tau<1$ implies declining impatience (correct but incomplete). Proposition 8 clarifies these points.

In the light of Proposition 8, we can now see the interpretation of the parameters $\tau$ and $\rho$ in the RS-discount function (2.12). $\tau$ controls impatience, independently of the values of the other parameters $\alpha$, $\beta$ and $\rho$: $0<\tau<1$, gives declining impatience, $\tau=1$ gives constant impatience and $\tau>1$gives increasing impatience. If $0<\rho\leq 1$, then we get subadditivity, irrespective of the values of the other parameters $\alpha$, $\beta$ and $\tau$. However, if $\rho>1$, then (2.12) can be neither subadditive, additive nor superadditive (depending on the particular values of $r$, $s$ and $t$, we may have $D(r,s)<D(r+t,s+t)$, $D(r,s)=D(r+t,s+t)$ or $D(r,s)>D(r+t,s+t)$).

[10]Some authors use 'present bias' for what we call 'declining impatience'. But other authors use 'present bias' to mean that the discount function, $D(s,t)$ is declining in $t$. So we prefer 'declining impatience' to avoid confusion. It is common to use 'stationarity' for what we call 'constant impatience'. We prefer the latter, to be in conformity with 'declining impatience' and 'increasing impatience'.

### 2.7. Intransitive preferences: Real or apparent?

Consider the following hypothetical situation. A decision maker prefers a payoff of 1 now to a payoff of 2 next period, i.e., (2, next period) $\prec$ (1, now). The decision maker also prefers a payoff of 2 next period to a payoff of 3 two periods from now, i.e., (3, 2 two periods from now) $\prec$ (2, next period). Finally, the same decision maker prefers a payoff of 3 two periods from now to a payoff of 1 now, i.e., (1, now) $\prec$ (3, 2 two periods from now). Schematically:

$$(1, \text{now}) \prec (3, 2 \text{ two periods from now}) \prec (2, \text{next period}) \prec (1, \text{now}) . \tag{2.21}$$

Ok and Masatlioglu (2007, p215) use a similar example to motivate their intransitive theory of relative discounting.

Alternatively, we may view (2.21) as due to a framing effect resulting in a shift in the reference point for time. Assume that the choice of reference time in each pairwise comparison is the sooner of the two dates, in conformity with Assumption A0, subsection 2.3. Then (2.21) can be formalized as follows.

$$V_0(1,0) < V_0(3,2) , \; V_1(3,2) < V_1(2,1) , \; V_0(2,1) < V_0(1,0) . \tag{2.22}$$

Thus, the decision maker prefers a payoff of 1 now to a payoff of 2 next period, both discounted back to the present. The decision maker also prefers a payoff of 2 next period to a payoff of 3 the following period, both discounted back to next period. Finally, the decision maker prefers a payoff of 3 in two periods from now to a payoff of 1 now, both discounted back to the present. If this view is accepted, then the apparent intransitivity in (2.21) arises from conflating $V_0(3,2)$ with $V_1(3,2)$ and $V_1(2,1)$ with $V_0(2,1)$. The following example shows that (2.22) is consistent with a reference-time theory of intertemporal choice.

**Example 1** *: Take the reference point for wealth be the current level of wealth, so each payoff is regarded as a gain to current wealth. Take the value function to be*[11]

$$v(x) = x^{\frac{1}{2}} (1+x)^{\frac{1}{2}} , \; x \geq 0. \tag{2.23}$$

*Thus (working to five significant figures),*

$$v(1) = 1.414\,2, \; v(2) = 2.449\,5 \text{ and } v(3) = 3.464\,1. \tag{2.24}$$

*As our discount function we take the Read-Scholten discount function (2.12) with* $\alpha = \beta = 1$ *and* $\rho = \tau = \frac{1}{2}$*:*

$$D(r,t) = \left(1 + \left(t^{\frac{1}{2}} - r^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)^{-1} . \tag{2.25}$$

[11] The reasons for this choice will become clear in sections 3.1 and 3.2, below. Many other choices are possible.

*Thus,*

$$D(0,1) = \frac{1}{2},\ D(1,2) = 0.608\,42 \text{ and } D(0,2) = 0.456\,79. \tag{2.26}$$

*From (2.24) and (2.26) we get*

$$V_0(1,0) = v(1)\,D(0,0) = 1.414\,2, \tag{2.27}$$

$$V_0(3,2) = v(3)\,D(0,2) = 1.582\,4, \tag{2.28}$$

$$V_1(3,2) = v(3)\,D(1,2) = 2.107\,6, \tag{2.29}$$

$$V_1(2,1) = v(2)\,D(1,1) = 2.449\,5, \tag{2.30}$$

$$V_0(2,1) = v(2)\,D(0,1) = 1.224\,8. \tag{2.31}$$

*From (2.27) to (2.31), we get*

$$V_0(1,0) < V_0(3,2),\ V_1(3,2) < V_1(2,1),\ V_0(2,1) < V_0(1,0), \tag{2.32}$$

*confirming (2.22).*

A consequence of Proposition 7 is that no additive discount function (e.g., exponential (2.9), PPL (2.10) or LP (2.11)) can explain (apparently) intransitive choices as exhibited in (2.21). The reason is that, under the conditions of that proposition, all utilities can be discounted back to time zero and, hence, can be compared and ordered.

### 2.8. The common difference effect: Declining impatience or subadditivity?

Let us reconsider the common difference effect, using Thaler's apples example (anomaly 3 in the list of subsection 1.1). A decision maker prefers one apple today to two apples tomorrow, so that

$$v(1) > v(2)\,D(0,1). \tag{2.33}$$

However, the decision maker, today, prefers to receive two apples in 51 days' time to receiving one apple in 50 days' time, so that

$$v(1)\,D(0,50) < v(2)\,D(0,51). \tag{2.34}$$

From (2.33) and (2.34) we get

$$D(0,50)D(0,1) < D(0,51). \tag{2.35}$$

If we assume additivity (as did LP), so that $D(0,51) = D(0,50)D(50,51)$, we get, from (2.35), $D(0,1) < D(50,51)$. So the decision maker exhibits declining impatience. However, and as Read (2001) pointed out, subadditivity could be an alternative explanation. To see this, assume constant impatience, so that $D(0,1) = D(50,51)$. Then (2.35) gives $D(0,50)D(50,51) < D(0,51)$. Thus, the common difference effect is consistent with constant impatience if the discount function is sufficiently subadditive.

**Example 2** *: (Thaler's apples example)*
*Take the reference point to be 'no apples' and take the value function to be*

$$v(x) = x^{\frac{1}{2}}(1+x)^{\frac{1}{2}}, \; x \geq 0. \tag{2.36}$$

*Thus,*

$$v(1) = 1.414\,2 \text{ and } v(2) = 2.449\,5. \tag{2.37}$$

*We compare the resolution of the 'common difference effect' anomaly under the LP-discount function (2.11) and the RS-discount function (2.12). To simplify as much as possible, choose the parameters:* $\alpha = \beta = \tau = \rho = 1$. *We shall use these parameters in other examples too. We tabulate the relevant magnitudes below:*

| | *LP:* $D(s,t) = (1+s)(1+t)^{-1}$ | *RS:* $D(s,t) = (1+t-s)^{-1}$ |
|---|---|---|
| $D(0,1)$ | $1/2$ | $1/2$ |
| $D(0,50)$ | $1/51$ | $1/51$ |
| $D(0,51)$ | $1/52$ | $1/52$ |
| $D(50,51)$ | $51/52$ | $1/2$ |
| $D(0,50)\,D(50,51)$ | $1/52$ | $1/102$ |

*Recall that the decision maker prefers one apple today to two apples tomorrow if, and only if,*

$$v(1) > v(2)\,D(0,1). \tag{2.38}$$

*On the other hand, the decision maker, today, prefers to receive two apples in 51 days' time to receiving one apple in 50 days' time if, and only if,*

$$v(1)\,D(0,50) < v(2)\,D(0,51). \tag{2.39}$$

*Substituting from the above table and (2.37) into (2.38) and (2.39) gives, respectively,*[12]

$$v(1) = 1.4 > 1.2 = v(2)\,D(0,1), \tag{2.40}$$

$$v(1)\,D(0,50) = 0.028 < 2.4 = v(2)\,D(0,51), \tag{2.41}$$

*for both the LP-discount function and the RS-discount function. This illustrates that both approaches can explain the common difference effect. However, they explain in very different ways. Comparing rows one and four of the table, we see that the LP-discount function exhibits declining impatience,* $D(0,1) = \frac{1}{2} < \frac{51}{52} = D(50,51)$, *while the RS-discount function exhibits constant impatience,* $D(0,1) = \frac{1}{2} = D(50,51)$. *On the other hand, comparing rows three and five of the table, we see that the LP-discount function is*

[12] In the presentation of the results, we write down the first two significant figures, which is entirely adequate.

*additive,* $D(0,51) = \frac{1}{52} = D(0,50)\,D(50,51)$*, while the RS-discount function subadditive,* $D(0,51) = \frac{1}{52} > \frac{1}{102} = D(0,50)\,D(50,51)$*. Thus, the LP-discount function explains the common difference effect as exclusively due to declining impatience, while the RS-discount function explains this effect as due (in this example, exclusively) to subadditivity. More generally, and provided* $0 < \rho \leq 1$*, the RS-discount function can combine subadditivity with declining impatience (*$0 < \tau < 1$*), constant impatience (*$\tau = 1$*) or increasing impatience (*$\tau > 1$*).*

Of course, and as Read (2001) pointed out, the common difference effect could be due to both declining impatience and subadditivity. Read (2001), conducted a series of experiments that tested for the common difference effect and could also discriminate between subadditivity and declining impatience. He found support for the common difference effect and for subadditivity but rejected declining impatience in favour of constant impatience. Read (2001) also discusses the psychological foundation for subadditivity.

### 2.9. Representation theorems

Suppose that $x$ received at time 0 is equivalent to $y$ received at time $t$ (when both are discounted back to time 0), so that $v(x) = v(y)\,D(0,t)$. Suppose that the receipt of $x$ is delayed to time $s$. We ask, at what time, $T$, will $y$ received at time $T$ be equivalent to $x$ received at time $s$ (when both are discounted back to time 0)? Or, for what time, $T$, will the following hold: $v(x)\,D(0,s) = v(y)\,D(0,T)$? For the exponential discount function (2.9) the answer is clear: $T = s + t$. More generally, does such a $T$ exist? Is it unique? Does it depend on $x, y$ as well as $s, t$? What are its properties? These questions are answered by Propositions 9 and 10, below. But first, a definition.

**Definition 7** *(Delay functions): Let* $D$ *be a discount function. Suppose the function,* $\Psi$*, has the property* $D(0,s)\,D(0,t) = D(0,\Psi(s,t))$*,* $s \geq 0$*,* $t \geq 0$*. Then we call* $\Psi$ *a delay function corresponding to the discount function,* $D$*. We also say that the discount function,* $D$*, exhibits* $\Psi$*-delay.*

**Proposition 9** *(Properties of a delay function): Let* $D$ *be a discount function and* $\Psi$ *a corresponding delay function. Then* $\Psi$ *has the following properties:*
*(a)* $\Psi$ *is unique,*
*(b)* $\Psi(s,t)$ *is strictly increasing in each of* $s$ *and* $t$*,*
*(c)* $\Psi(s,t) = \Psi(t,s)$*,*
*(d)* $\Psi(0,t) = \Psi(t,0) = t$*,*
*(e)* $v(x) = v(y)\,D(0,t)$ *if, and only if,* $v(x)\,D(0,s) = v(y)\,D(0,\Psi(s,t))$*.*

Suppose that $x$ received at time 0 is equivalent to $y$ received at time $t$ (when both are discounted back to time 0), so that $v(x) = v(y) D(0,t)$. Suppose that the receipt of $x$ is delayed to time $s$. Then, according to Proposition 9(e), the delay function, $\Psi(s,t)$, if it exists, gives the time to which the receipt of $y$ has to be deferred, so as to retain equivalence to $x$ (when both are discounted back to time 0). Therefore, we called $\Psi$ a *delay function.*

**Proposition 10** *(Existence of a delay function): A continuous discount function has a unique delay function.*

We now introduce our fourth defined function (the others were: the discount function, the generating function and the delay function).

**Definition 8** *(The seed function): Let $\psi : [0,\infty) \to [0,\infty)$ be strictly increasing, with $\psi(0) = 0$. We call $\psi$ a seed function. If, in addition, $\psi$ is onto, we call $\psi$ a continuous seed function.*

A 'continuous seed function' is continuous. The proof is similar to that of Proposition 1 and, therefore, will be omitted.

The following definition gives a useful representation for discount functions.

**Definition 9** *(Representations): Let $\alpha > 0$ and $\beta > 0$. We call $D(0,t) = [1 + \alpha\psi(t)]^{-\frac{\beta}{\alpha}}$ an $(\alpha,\beta)$-representation of the discount function $D(r,t)$.*

Proposition 11, below, establishes the existence of $(\alpha,\beta)$-representations for continuous discount functions and shows their connection to delay functions.

**Proposition 11** *(Representation Theorem 1): Let $D$ be a continuous discount function. Let $\Psi$ be the corresponding delay function. Then, for each $\alpha > 0$ and each $\beta > 0$, $D$ has the unique $(\alpha,\beta)$-representation $D(0,t) = [1 + \alpha\psi(t)]^{-\frac{\beta}{\alpha}}$. Moreover, $\psi$ has the properties:*
*(a) $\psi : [0,\infty) \overset{onto}{\to} [0,\infty)$ is strictly increasing (hence $\psi(0) = 0$ and $\psi$ is a continuous seed function).*
*(b) $\psi^{-1}$ exists and $\psi^{-1} : [0,\infty) \overset{onto}{\to} [0,\infty)$ is strictly increasing with $\psi^{-1}(0) = 0$.*
*(c) $\Psi(s,t) = \psi^{-1}[\psi(s) + \psi(t) + \alpha\psi(s)\psi(t)]$.*

From Proposition 11, we see that if $\Psi$ is to be the delay function of some continuous discount function, then it must take the form given in part (c) of that proposition. In the light of this, when considering possible delay functions, we can restrict ourselves, without loss of generality, to the class of functions of the form $\Psi(s,t) = \psi^{-1}[\psi(s) + \psi(t) + \alpha\psi(s)\psi(t)]$, where $\psi$ is as in part (a), i.e., a continuous seed function.

The following proposition is a generalization of LP's derivation of their generalized hyperbolic discount function.

**Proposition 12** *(Representation Theorem 2): Let* $\psi : [0,\infty) \overset{onto}{\to} [0,\infty)$ *be strictly increasing,* $\alpha > 0$ *and* $\Psi(s,t) = \psi^{-1}[\psi(s) + \psi(t) + \alpha\psi(s)\psi(t)]$. *Let* $D$ *be a continuous discount function with delay function,* $\Psi$. *Then, for some* $\beta > 0$, $D(0,t) = [1+\alpha\psi(t)]^{-\frac{\beta}{\alpha}}$.

According to Propositions 9(a) and 10, a continuous discount function determines a unique delay function. Hence, we can partition the set of all continuous discount functions into equivalence classes, two continuous discount functions being in the same equivalence class if, and only if, they have the same delay function. Many different discount functions could have the same delay function. In fact, according to Representation Theorem 1 (Proposition 11), all (the different) continuous discount functions, $D$, for which $D(0,t) = [1+\alpha\psi(t)]^{-\frac{\beta}{\alpha}}$ (fixed $\alpha$ and $\psi$, different $\beta$'s) have the same delay function and, hence, lie in the same equivalence class. But does an equivalence class contain other continuous discount functions? Representation Theorem 2 (Proposition 12) gives the answer 'no': Consider an arbitrary equivalence class. Choose some member of that class. Let it have the $(\alpha,\beta)$-representation $D(0,t) = [1+\alpha\psi(t)]^{-\frac{\beta}{\alpha}}$. Then *all* members of its class can be obtained by varying $\beta$, keeping $\alpha$ and $\psi$ fixed.

### 2.10. Extension theorems

Proposition 8(a) and (b) established that the RS-discount function (2.12) is not additive and, hence, cannot be obtained as an additive extension of a strictly decreasing function $\varphi : [0,\infty) \to (0,1]$, $\varphi(0) = 1$. We, therefore, need a more general way to extend such a strictly decreasing function to a discount function. This is what we turn to in this subsection. We start by introducing our fifth, and final, defined function.

**Definition 10** *(extension functions): Let* $f : \Delta \to [0,\infty)$ *satisfy:*
*(i) For each* $r \in [0,\infty)$, $f(r,t)$ *is a strictly increasing function of* $t \in [r,\infty)$ *into* $[0,\infty)$, *with* $f(r,r) = 0$.
*(ii) For each* $t \in [0,\infty)$, $f(r,t)$ *is a strictly decreasing function of* $r \in [0,t]$ *into* $[0,t]$, *with* $f(0,t) = t$.
*Then we call* $f$ *an extension function. If, in (i), 'into' is replaced with 'onto', then we call* $f$ *a continuous extension function.*

A 'continuous extension function', $f(r,t)$, is continuous in $t$. The proof is the same as that of Proposition 1 and, therefore, will be omitted.

**Definition 11** *($f$-extensions): Let* $D$ *be a discount function. Let* $f : \Delta \to [0,\infty)$ *satisfy* $D(r,t) = D(0, f(r,t))$, *then*
*(a) we call* $f$ *an extension function corresponding to* $D$,
*(b) we refer to* $D(r,t)$ *as an* $f$*-extension of* $D(0,t)$, *or just an* $f$*-extension.*

Definition 10 defines extension functions independently of any discount function. By contrast, Definition 11 defines an extension function corresponding to a give discount function. Our terminology suggests that 'an extension function corresponding to a given discount function' is, in fact, 'an extension function'. That this is indeed the case, is established in the following proposition.

**Proposition 13** *(Extension Theorem 1): Let $D$ be a discount function. Let $f$ be a corresponding extension function in the sense of Definition 11(a). Then:*
*(a) $f$ is unique.*
*(b) $f$ is an extension function in the sense of Definition 10.*
*(c) $v(x) = v(y)\, D(r,t)$ if, and only if, $v(x) = v(y)\, D(0, f(r,t))$.*

Suppose that $x$ received at time $r$ is equivalent to $y$ received at time $t$, $0 \leq r \leq t$, time $r$ being the reference time; so that $v(x) = v(y)\, D(r,t)$. Suppose that the receipt of $x$ is brought forward to time 0. We ask, at what time, $T$, will $y$ received at time $T$ be equivalent to $x$ received at time 0, time 0 being the new reference time? Or, for what time, $T$, will the following hold: $v(x) = v(y)\, D(0,T)$? For the exponential discount function (2.9) the answer is clear: $T = t - r$. More generally, Proposition 13(c) gives the answer as $T = f(r,t)$.

**Proposition 14** *(Extension Theorem 2): Let the discount function, $D$, be continuous. Then there exists an extension function, $f$, corresponding to $D$ in the sense of Definition 11. Moreover, $f$ is unique and is a continuous extension function in the sense of Definition 10.*

**Proposition 15** *(Extension Theorem 3): Let $\varphi$ be a generating function and $f$ an extension function. Then:*
*(a) $D(r,t) = \varphi(f(r,t))$ is a discount function.*
*(b) $f$ is the extension function corresponding to $D$ and $D(r,t)$ is the $f$-extension of $D(0,t)$.*
*(c) If $\varphi$ is a continuous generating function and $f$ a continuous extension function, then $D$ is a continuous discount function.*

To summarize, given a generating function, $\varphi$, and an extension function, $f$, by Extension Theorem 3 (Proposition 15), we can construct a discount function $D$ so that $D(r,t)$ is the $f$-extension of $\varphi(t) = D(0,t)$. Extension Theorem 2 (Proposition 14) tells us that *all* continuous discount functions are obtainable in this way from continuous generating functions and continuous extension functions.

### 2.11. Characterization theorems

We can combine the representation and extension theorems of the previous two subsections to produce further useful results, which we now turn to.

**Proposition 16** *(Characterization Theorem 1): $D$ is a continuous discount function if, and only if, $D(r,t) = [1+\alpha\psi(f(r,t))]^{-\frac{\beta}{\alpha}}$, $\alpha > 0$, $\beta > 0$, where $\psi$ is a continuous seed function and $f$ is a continuous extension function. Furthermore, $f$ is uniquely determined by $D$.*

**Proposition 17** *(Characterization Theorem 2): A continuous discount function with the extension function, $f$, exhibits:*
*(a) declining impatience if, and only if, $f(r,s) > f(r+t, s+t)$, for $t > 0$ and $r < s$,*
*(b) constant impatience if, and only if, $f(r,s) = f(r+t, s+t)$, for $t \geq 0$ and $r \leq s$,*
*(c) increasing impatience if, and only if, $f(r,s) < f(r+t, s+t)$, for $t > 0$ and $r < s$,*

**Proposition 18** *(Characterization Theorem 3): A continuous discount function, $D$, is additive if, and only if, $f(r,t) = \psi^{-1}\left(\frac{\psi(t)-\psi(r)}{1+\alpha\psi(r)}\right)$, in which case $D(r,t) = \left[\frac{1+\alpha\psi(t)}{1+\alpha\psi(r)}\right]^{-\frac{\beta}{\alpha}}$.*

**Proposition 19** *: The following two tables give a seed function, $\psi$, the generating function, $\varphi$, the extension function, $f$, and the delay function, $\Psi$, of each of the discount functions $D(r,t)$ (2.9) to (2.12).*

| | $\psi(t)$ | $\varphi(t)$ | $f(r,t)$ | $\Psi(s,t)$ |
|---|---|---|---|---|
| ***exponential*** | $\frac{e^{\alpha t}-1}{\alpha}$ | $e^{-\beta t}$ | $t-r$ | $s+t$ |
| ***LP*** | $t$ | $(1+\alpha t)^{-\frac{\beta}{\alpha}}$ | $\frac{t-r}{1+\alpha r}$ | $s+t+\alpha st$ |
| ***RS*** | $t^{\tau\rho}$ | $[1+\alpha t^{\tau\rho}]^{-\frac{\beta}{\alpha}}$ | $(t^{\tau}-r^{\tau})^{\frac{1}{\tau}}$ | $[s^{\tau\rho}+t^{\tau\rho}+\alpha(st)^{\tau\rho}]^{\frac{1}{\tau\rho}}$ |
| ***generalized RS*** | $\frac{e^{\frac{\alpha}{\beta}Q[w(t)]}-1}{\alpha}$ | $e^{-Q[w(t)]}$ | $w^{-1}[w(t)-w(r)]$ | $w^{-1}Q^{-1}[Q(w(s))+Q(w(t))]$ |
| ***PPL*** $(r=t=0)$ | $0$ | $1$ | $0$ | $s$ |
| ***PPL*** $(0=r<t)$ | $\frac{e^{\frac{\alpha\delta}{\beta}+\alpha t}-1}{\alpha}$ | $e^{-(\delta+\beta t)}$ | $t$ | $\frac{\delta}{\beta}+s+t$ |
| ***PPL*** $(0<r\leq t)$ | $\frac{e^{\alpha t}-1}{\alpha}$ | $e^{-\beta t}$ | $t-r$ | $NA$ |

| | $D(r,t)$ |
|---|---|
| ***exponential*** | $e^{-\beta(t-r)}$ |
| ***LP*** | $\left(\frac{1+\alpha t}{1+\alpha r}\right)^{-\frac{\beta}{\alpha}}$ |
| ***RS*** | $[1+\alpha(t^{\tau}-r^{\tau})^{\rho}]^{-\frac{\beta}{\alpha}}$ |
| ***generalized RS*** | $e^{-Q[w(t)-w(r)]}$ |
| ***PPL*** $(r=t=0)$ | $1$ |
| ***PPL*** $(0=r<t)$ | $e^{-(\delta+\beta t)}$ |
| ***PPL*** $(0<r\leq t)$ | $e^{-\beta(t-r)}$ |

Starting with a continuous seed function, $\psi$, an $\alpha > 0$ and a $\beta > 0$, we can 'grow' from them a unique generating function, $\varphi(t) = [1 + \alpha\psi(t)]^{-\frac{\beta}{\alpha}}$ (which turns out to be continuous). Given this generating function and a continuous extension function, $f(r,t)$, we obtain a unique discount function $D(r,t) = \varphi(f(r,t)) = [1 + \alpha\psi(f(r,t))]^{-\frac{\beta}{\alpha}}$ (which also turns out to be continuous). This discount function determines a unique delay function, $\Psi(s,t) = \psi^{-1}[\psi(s) + \psi(t) + \alpha\psi(s)\psi(t)]$.

Conversely, a continuous discount function, $D$, determines a unique (continuous) generating function, $\varphi(t) = D(0,t)$ and a unique (continuous) extension function, $f$, so that $D$ is the $f$-extension of $\varphi$: $D(r,t) = \varphi(f(r,t))$.

Although a continuous discount function, $D$, determines unique generating, extension and delay functions, $\varphi$, $f$ and $\Psi$, it does *not* determine unique $\alpha$, $\beta$ or $\psi$ in the representation $D(0,t) = [1 + \alpha\psi(t)]^{-\frac{\beta}{\alpha}}$. For example, the LP-discount function $D(r,t) = (1+\alpha r)^{\frac{\beta}{\alpha}}(1+\alpha t)^{-\frac{\beta}{\alpha}}$, $\alpha > 0, \beta > 0$, has, obviously, the representation $D(0,t) = (1+\alpha t)^{-\frac{\beta}{\alpha}}$ (with $\psi(t) = t$) and, hence, the delay function $\Psi(s,t) = s + t + \alpha st$. But it also has many other representations: $D(0,t) = [1 + a\psi(t)]^{-\frac{b}{a}}$, $\psi(t) = \frac{(1+\alpha t)^{\frac{\beta a}{\alpha b}} - 1}{a}$, for all $a > 0$ and all $b > 0$. However, it can easily be check that $[1 + a\psi(t)]^{-\frac{b}{a}} = (1+\alpha t)^{-\frac{\beta}{\alpha}}$ and $\psi^{-1}[\psi(s) + \psi(t) + a\psi(s)\psi(t)] = s + t + \alpha st$. Since the delay function, $\Psi$, but not the seed function, $\psi$, is uniquely determined by $D$, it is better to say that $D$ exhibits $\Psi$ delay rather than $\psi$ delay.

### 2.12. Assumptions and consequences

LP introduce four assumptions, all with good experimental support (LP, II pp574-578). We adapt these assumption to allow for discount functions that are not, necessarily, additive. Under this condition, the reference point for time becomes important (Proposition 7). Let the discount function, $D$, be given by:

$$D(r,t) = [1 + \alpha\psi(f(r,t))]^{-\frac{\beta}{\alpha}}, \alpha > 0, \beta > 0, \tag{2.42}$$

where $\psi$ and $f$ are, respectively, the seed and extension functions. If $D$ is a continuous discount function then it can always be represented in the form (2.42), where $\psi$ is a continuous seed function and $f$ is a continuous extension function. Moreover, $f$ is determined uniquely by $D$ (Characterization Theorem 1 (Proposition 16)). Furthermore, under the assumption of continuity, $D$ will have a unique delay functions, $\Psi$ (Propositions 9 and 10 and Representation Theorem 1 (Proposition 11)), and is given by:

$$\Psi(s,t) = \psi^{-1}[\psi(s) + \psi(t) + \alpha\psi(s)\psi(t)]. \tag{2.43}$$

Assumptions A1 to A4, below, correspond to anomalies 1 to 4, above (subsection 1.1). Thus, what is regarded as anomalous behavior from the neoclassical point of view is at the core of the RT theory.

Given a discount function, $D(r,t)$, the assumption A1 to A4, below, place restrictions *only* on $D(0,t)$, i.e., only on discounting from an arbitrary time, $t \geq 0$, *back to time zero*. Hence, to derive results for $D(r,t)$, further assumptions are needed. In particular, for Proposition 24 we assume that $D(r,t)$ is an additive extension of $D(0,t)$, while for Proposition 25 we assume that $D(r,t)$ is an $f$-extension of $D(0,t)$ for $f(r,t) = (t^\tau - r^\tau)^{\frac{1}{\tau}}$.

**A1** *Gain-loss asymmetry.* If $0 < x < y$ and $v(x) = v(y) D(0,t)$, then $v(-x) > v(-y) D(0,t)$.

**A2** *Magnitude effect.* If $0 < x < y$, $v(x) = v(y) D(0,t)$ and $a > 1$, then $v(ax) < v(ay) D(0,t)$. If $y < x < 0$, $v(x) = v(y) D(0,t)$ and $a > 1$, then $v(ax) > v(ay) D(0,t)$.

**A3** *Common difference effect.* If $0 < x < y$, $v(x) = v(y) D(0,t)$ and $s > 0$, then $v(x) D(0,s) < v(y) D(0,s+t)$. If $y < x < 0$, $v(x) = v(y) D(0,t)$ and $s > 0$, then $v(x) D(0,s) > v(y) D(0,s+t)$.

**A4** *Delay-speedup asymmetry.* For $c > 0$, $s > 0$ and $t > 0$, $V_0((0,0),(c,s),(-c,s+t)) < -V_0((0,0),(-c,s),(c,s+t))$.[13]

**Definition 12** *($\alpha$-subadditivity): Let $\alpha > 0$. A function, $\psi$, is $\alpha$-subadditive if, for all $s$ and $t$ for which $\psi$ is defined and non-zero: $\psi(s+t) < \psi(s) + \psi(t) + \alpha\psi(s)\psi(t)$.*

A function, $\psi$, is subadditive (in the standard sense) if, for all $s$ and $t$ for which $\psi$ is defined: $\psi(s+t) \leq \psi(s) + \psi(t)$. A function that is $\alpha$-subadditive, for some $\alpha > 0$, need not be subadditive. However, a function is subadditive if, and only if, it is $\alpha$-subadditive for all $\alpha > 0$.[14]

[13] A4 is to be understood as follows. In the LHS of the inequality, the reference stream of the decision maker is $0, 0, c$ (for dates $0, s, s+t$) i.e. a reward is contractually promised at time $s+t > 0$. The individual is then offered a choice to receive the reward early, at time $s$ (*speedup*). Given the assumption on reference time, in A0, the income stream, relative to reference wealth, $0, c, -c$, can be explained as follows. The individual was not expecting anything at times 0 and $s$ so relative to reference wealth, he gets $0-0, c-0$ at times 0 and $s$. Having received a reward of $c$ at time $s$, his reference wealth is $c$. Hence, at time $t+s$ his wealth relative to the reference wealth is $0-c=-c$. For the RHS of the inequality, the contractually promised income stream is $0, c, 0$ (for dates $0, s, s+t$). The individual is then told that the reward will now instead be available only at time $s+t$ (*delay*). Proceeding as before, the stream of income relative to the reference point is now $0, -c, c$.

[14] Similarly, a function, $\psi$, is additive (in the standard sense) if, for all $s$ and $t$ for which $\psi$ is defined: $\psi(s+t) = \psi(s) + \psi(t)$. Consider the exponential discount function, $D(r,t) = e^{-\beta(t-r)}$, $\beta > 0$. Then $\ln D(0,t)$ is additive in this sense. And, of course, $D(r,t)$ is additive in the sense of Definition 3. Also note that $\alpha$-subaddivity, as in Definition 12, neither implies nor is implied by subadditivity of the discount function, as in Definition 3.

**Proposition 20** *(Representation Theorem 3): Let $D$ be a continuous discount function. Let $\Psi$ be the corresponding delay function. Let $\alpha > 0$ and $\beta > 0$. Let $D(0,t) = [1+\alpha\psi(t)]^{-\frac{\beta}{\alpha}}$ be an $(\alpha,\beta)$-representation of $D$. Then the following are equivalent:*
*(a) $\psi$ is $\alpha$-subadditive.*
*(b) If $s > 0$ and $t > 0$, then $\Psi(s,t) > s + t$.*
*(c) If $s > 0$ and $t > 0$, then $D(0,s)\,D(0,t) < D(0,s+t)$.*

**Proposition 21** *(Characterization Theorem 4): Let $D$ be a continuous discount function. Then preferences exhibit the common difference effect for gains if, and only if, the seed function, $\psi$, is $\alpha$-subadditive.*[15]

**Definition 13** *($\gamma$-delay): Preferences exhibit $\gamma$-delay, if the delay function (2.43) is*

$$\Psi(s,t) = \left(s^\gamma + t^\gamma + \alpha s^\gamma t^\gamma\right)^{\frac{1}{\gamma}},\ \alpha > 0,\ 0 < \gamma \leq 1. \tag{2.44}$$

*In particular, if $\gamma = 1$, we say that preferences exhibit linear delay.*

A delay function, if it exists, is unique (Proposition 9) and it always exists for a continuous discount function (Proposition 10). Hence, Definition 13 is a sound definition. However, it should be remembered that $\gamma$-delay is a property of the delay function, $\Psi$, not of the seed function, $\psi$ (see discussion at end of subsection 2.11).

**Proposition 22** *: If preferences exhibit $\gamma$-delay, then they also exhibit the common difference effect.*

**Proposition 23** *: If preferences, with a continuous discount function, $D$, exhibits $\gamma$-delay, then, necessarily, $D(0,t) = (1+\alpha t^\gamma)^{-\frac{\beta}{\alpha}}$ and $D(0,t) = (1+\alpha t^\gamma)^{-\frac{\beta}{\alpha}}$, where $\alpha$ are $\beta$ are positive.*

**Proposition 24** *(generalization of LP): If preferences, with an additive continuous discount function, $D$, exhibits $\gamma$-delay, then, necessarily, $D(r,t) = \left(\frac{1+\alpha t^\gamma}{1+\alpha r^\gamma}\right)^{-\frac{\beta}{\alpha}}$, where $\alpha$ and $\beta$ are positive.*
*In particular, $\gamma = 1$ gives the LP-discount function (2.11).*

**Proposition 25** *(RS-discount functions): Let $\tau > 0$, $\rho = \frac{\gamma}{\tau}$ and $f(r,t) = \left(t^\tau - r^\tau\right)^{\frac{1}{\tau}}$. Let preferences with continuous discount function, $D$, exhibit $\gamma$-delay. Let the extension function be $f$. Then $D$ is the RS-discount functions, $D(r,t) = \left[1+\alpha\left(t^\tau - r^\tau\right)^\rho\right]^{-\frac{\beta}{\alpha}}$,*
*where $\alpha$ and $\beta$ are positive.*

[15] We noted earlier that the common difference effect can be explained by a combination of declining impatience and subadditivity *of the discount function*. However, $\alpha$-subadditivity is not to be confused with subadditivity of the discount function.

**Proposition 26** *(LP, p583): For a continuous discount function, A1 implies the following:*
*(a) Losses are discounted less heavily than gains in the following sense:* $0 < x < y \Rightarrow \frac{v(x)}{v(y)} > \frac{v(-x)}{v(-y)}$.
*(b) The value function is more elastic for losses than for gains:* $x > 0 \Rightarrow \epsilon_v(-x) > \epsilon_v(x)$.

**Proposition 27** *(LP, p584): For a continuous discount function, A2 implies that the value function is*
*(a) subproportional:* $(0 < x < y \text{ or } y < x < 0) \Rightarrow \frac{v(x)}{v(y)} > \frac{v(ax)}{v(ay)}$, *for* $a > 1$,
*(b) more elastic for outcomes of larger absolute magnitude:* $(0 < x < y \text{ or } y < x < 0) \Rightarrow \epsilon_v(x) < \epsilon_v(y)$.

Intuitively, increasing elasticity of the value function implies greater sensitivity of $v$ to increases in $x$. This in turn increases the weight of larger outcomes ($D(r,t)\,v(x_t)$) in intertemporal plans. A similar intuition applies to the result in Proposition 26(b).

We now add the standard assumption from prospect theory that the value function is strictly concave for gains and strictly convex for losses (Kahneman and Tversky, 1979):

**A5** *Declining sensitivity.* For $x > 0$, $v''(x) < 0$ (strict concavity for gains). For $x < 0$, $v''(x) > 0$ (strict convexity for losses).

Combining A5 with Proposition 27 we get:

**Proposition 28** *: A2 and A5 imply that* $0 < \epsilon_v < 1$.

**Proposition 29** *: Assumption A4 (delay-speedup asymmetry) follows from assumptions A0 (reference time) and A1 (gain-loss asymmetry).*

## 3. Simple increasing elasticity value functions (SIE)

A natural question that arises is 'Is the RT theory developed in section 2 consistent?' A related question is 'Is there a tractable functional form for the value function which can be combined with RT theory to produce a useful model?' We address these questions in this section. In subsection 3.2, below, we answer the second question in the affirmative. We call the value function developed there a *simple increasing elasticity* (SIE) value function. Our affirmative answer to the second question also provides an affirmative answer to the first question. But first, in subsection 3.1, immediately below, we show that none of several popular families of functions is compatible with RT theory or, indeed, any theory (e.g., LP) that attempts to explain the magnitude effect on the basis of increasing elasticity of the value function.

### 3.1. Decreasing elasticity of HARA utility functions

We consider several popular classes of value functions:

1. Constant relative risk aversion functions (CRRA)[16]

$$v(x) = \frac{x^{1-\gamma}}{1-\gamma},\ 0 < \gamma < 1;\ \epsilon_v(x) = \frac{xv'(x)}{v(x)} = 1-\gamma. \tag{3.1}$$

2. Hyperbolic absolute risk aversion functions (HARA)[17]

$$\begin{aligned} v(x) &= \frac{\gamma}{1-\gamma}\left[\left(\mu + \frac{\theta x}{\gamma}\right)^{1-\gamma} - \mu^{1-\gamma}\right],\ \theta > 0,\ \mu > 0,\ 0 < \gamma < 1,\ x \geq 0, \\ \epsilon_v(x) &= \frac{xv'(x)}{v(x)} = \frac{(1-\gamma)\frac{\theta x}{\gamma}\left(\mu + \frac{\theta x}{\gamma}\right)^{-\gamma}}{\left(\mu + \frac{\theta x}{\gamma}\right)^{1-\gamma} - \mu^{1-\gamma}}. \end{aligned} \tag{3.2}$$

3. Constant absolute risk aversion functions (CARA)

$$v(x) = 1 - e^{-\theta x},\ \theta > 0,\ x \geq 0;\ \epsilon_v(x) = \frac{1}{\sum_{n=1}^{\infty}\frac{(\theta x)^{n-1}}{n!}}. \tag{3.3}$$

4. Logarithmic functions

$$v(x) = \ln(1+\theta x),\ \theta > 0,\ x \geq 0;\ \epsilon_v(x) = \frac{\theta x}{(1+\theta x)\ln(1+\theta x)}. \tag{3.4}$$

5. Quadratic functions

$$v(x) = \frac{1}{2}\mu^2 - \frac{1}{2}(\mu - \theta x)^2,\ \theta > 0,\ 0 \leq x < \frac{\mu}{\theta};\ \epsilon_v(x) = \frac{2\theta x(\mu - \theta x)^2}{\mu^2 - (\mu - \theta x)^2}. \tag{3.5}$$

**Proposition 30** : *For members of the CRRA class of value functions (3.1), $\epsilon_v(x)$ is constant. For members of the HARA (3.2), CARA (3.3), logarithmic (3.4) and quadratic (3.5) classes of functions, $\epsilon_v(x)$ is declining. Hence these families violates Proposition 26.*

[16] The general restriction is that $\gamma \neq 1$. However, we need the stronger restriction, $0 < \gamma < 1$, in order to satisfy Proposition 28.

[17] The general restrictions are $\theta > 0$, $\left(\mu + \frac{\theta x}{\gamma}\right)^{1-\gamma} > 0$, $\gamma \neq 1$. Since we allow $x \in [0, \infty)$, the restriction $\left(\mu + \frac{\theta x}{\gamma}\right)^{1-\gamma} > 0$ implies that $\mu > 0$ and $\gamma > 0$. We then also need $\gamma < 1$ in order to satisfy Proposition 28.

Note that, traditionally, the HARA class is defined by $v(x) = \frac{\gamma}{1-\gamma}\left(\mu + \frac{\theta x}{\gamma}\right)^{1-\gamma}$, and that $\epsilon_v(x) = (1-\gamma)\left(1 + \frac{\gamma\mu}{\theta x}\right)^{-1}$, which is *increasing* in $x$, as required by Proposition 26. While an additive constant, of course, makes no difference in expected utility theory; its absence here would violate the assumption $v(0) = 0$. However, including the constant $-\frac{\gamma}{1-\gamma}\mu^{1-\gamma}$, to make $v(0) = 0$, results in $\epsilon_v(x)$ *decreasing* with $x$, as will be shown by Proposition 30, and, hence, violates Proposition 26.

### 3.2. Simple increasing elasticity utility functions (SIE)

Consider the value function

$$\begin{aligned} v(x) &= \frac{\gamma x^{\sigma}}{1-\gamma}\left(\mu+\frac{\theta_+}{\gamma}x\right)^{1-\gamma},\ x\geq 0, \\ v(x) &= -\lambda\frac{\gamma(-x)^{\sigma}}{1-\gamma}\left(\mu-\frac{\theta_-}{\gamma}x\right)^{1-\gamma},\ x<0, \\ \mu &> 0,\ \theta_->\theta_+>0,\ \lambda\geq 1,\ 0<\sigma\leq\gamma<1. \end{aligned} \tag{3.6}$$

It may be interesting to note that (3.6) is a product of a CRRA function, $x^{\sigma}$, and a HARA function, $\frac{\gamma}{1-\gamma}\left(\mu+\frac{\theta_{\pm}}{\gamma}x\right)^{1-\gamma}$.

Proposition 31, below, establishes that the value function (3.6) is compatible with Propositions 26, 27 and 28.

**Proposition 31** *: From (3.6) it follows that*
*(a)* $v:(-\infty,\infty)\to(-\infty,\infty)$ *is continuous,* $v(0)=0$ *(reference dependence),* $v$ *is* $C^{\infty}$ *except at* $x=0$ *and, for* $x>0$, $-v(-x)>v(x)$ *(loss aversion).*
*(b)* $\epsilon_v(x)=\sigma+\frac{1-\gamma}{1+\frac{\gamma\mu}{\theta_+ x}}>0$, $\epsilon_v'(x)>0$, $x>0$.
*(c)* $\epsilon_v(x)=\sigma+\frac{1-\gamma}{1-\frac{\gamma\mu}{\theta_- x}}>0$, $\epsilon_v'(x)<0$, $x<0$.
*(d)* $x>0\Rightarrow\epsilon_v(x)<\epsilon_v(-x)$.

**Remark 1** *(The sign effect): The restriction* $\theta_->\theta_+>0$ *(along with the other restrictions) guarantees that* $x>0\Rightarrow\epsilon_v(x)<\epsilon_v(-x)$, *as required by Proposition 27(b).*[18]

**Corollary 1** *: From (b) and (c) of Proposition 31, we get that* $\epsilon_v(x)\to\sigma$ *as* $x\downarrow 0$ *and as* $x\uparrow 0$. *Hence we can define* $\epsilon_v(x)$ *as a continuous function for all* $x\in(-\infty,\infty)$ *as follows.* $\epsilon_v(0)=\sigma$, $\epsilon_v(x)=\sigma+\frac{1-\gamma}{1+\frac{\gamma\mu}{\theta_\pm|x|}}>0$ *for* $x\neq 0$. *Note that* $\epsilon_v(x)$ *is increasing in* $|x|$ *and* $\epsilon_v(x)\to\sigma+1-\gamma\leq 1$, *as* $|x|\to\infty$.

**Remark 2** *(SIE value function): In the light of Corollary 1, we may call the value function (3.6) a simple increasing elasticity (SIE) value function.*

## 4. Explaining the anomalies

Here, we put together the results of sections 2 and 3.

**Proposition 32** *: Each of the four discount functions (2.10), (2.11), (2.12) with* $\rho\leq 1$ *and* $\tau\rho\leq 1$, *and (2.13), when combined with the SIE value function (3.6), satisfies assumptions A0 to A5, i.e., all the assumptions.*

[18] One may wonder whether it is possible to allow the other parameters to take different values in the domains of gains and losses. The answer is no, as can be easily shown.

### 4.1. A summing up

To sum up so far, the PPL-discount function (2.10), the LP-discount function (2.11) and the RS-discount function (2.12) can all explain the common difference effect. But they explain it in different ways. The PPL (2.10) and LP (2.11) discount functions explain the common difference effect with declining impatience. For PPL, there is a sudden drop in impatience from time $t = 0$ to times $t > 0$, with impatience being constant for all times $t > 0$. For LP, on the other hand, the decline in impatience is continuous (recall Example 2). By contrast, the RS-discount function (2.12), on account of its subadditivity (for $0 < \rho \leq 1$) can explain the common difference whether we have declining impatience ($0 < \tau < 1$), constant impatience ($\tau = 1$) or increasing impatience ($\tau > 1$), provided $\tau\rho \leq 1$ (recall Proposition 8 and Example 2).

On the other hand, none of the discount functions (2.9), (2.10) or (2.11) can explain (apparent) intransitive preferences such as that exhibited by (2.21), recall Proposition 7 and subsection 2.7.

Thus, it emerges that of discount functions (2.9), (2.10), (2.11) and (2.12), the RS-discount function (2.12) is the most satisfactory because, when combined with the SIE value function, reference time/wealth, it can explain all the anomalies: gain-loss asymmetry, the magnitude effect, the common difference effect, delay-speedup asymmetry as well as subadditivity and (apparent) intransitivity.

## 5. Alternatives and extensions

In this section we compare the reference-time theory (RT) of section 2 with four recent developments.

First, we consider the tradeoff model of intertemporal choice of Scholten and Read (2006b), SR for short. We will argue that SR's tradeoff criterion can be represented by a discount function. Hence, it can be incorporated within RT. The gain is that their psychological arguments for their tradeoff model give support for RT theory and, in particular, their own RS-discount function.

The second development we consider is Halevy (2007), H for short, who shows that the common difference effect is compatible with exponential discounting, provided subjects are non-expected utility maximizers and exhibit the *certainty effect*. The certainty effect was first proposed as an explanation of the Allais paradox: subjects are much more sensitive to a change from certainty to uncertainty than they are to changes in the middle range of probabilities.

The third is the theory of vague time preferences of Manzini and Mariotti (2006), MM for short. Again, they can explain the common difference effect without departing

from exponential discounting. However, we believe that the importance of H and MM far transcends their ability to explain the common difference effect. On the other hand, and because in their present formulations they do not include any reference dependence, they are unable to explain *gain-loss asymmetry*, *delay-speedup asymmetry*, *subadditivity* and (apparent) *intransitivity*. By contrast, RT theory can explain all the anomalies. Nevertheless, we believe that it is desirable, and easy, to extend RT theory to incorporate uncertainty, as in H, and multiple criteria, as in MM. We show this, below, in the context of simple examples.

The fourth recent development we discuss here is the theory of intransitive preferences and relative discounting of Ok and Masatlioglu (2007), OM for short. This is the most radical of all the theories considered so far. From the outset it neither assumes transitivity nor additivity and, hence, is compatible with these two phenomena. In its present formulation, it cannot account for either *gain-loss asymmetry* or *delay-speedup asymmetry*. Furthermore, the lack of transitivity will make it hard to work with this theory, as the authors themselves explain. On the other hand, these problems can all be resolved in the special case of a transitive preference relation. But then their model becomes additive. In this case, OM would reduce to a standard discounting model.

Finally, all five theories (SR, H, MM, OM and RT) can explain the magnitude effect, when combined with the SIE value function (3.6).

### 5.1. The tradeoff model of intertemporal choice

Read and Scholten's critique of discounting models, including their own, led them to develop their *tradeoff model* of intertemporal choice (Read and Scholten, 2006). It is worth quoting their abstract in full:

> "Research on intertemporal judgement and choices between a smaller-sooner and a larger-later outcome has revealed many anomalies to the discounted-utility model. Attempts to account for these anomalies within the discounting paradigm have resulted in convoluted and psychologically opaque models. We therefore develop a new model of intertemporal choice, the tradeoff model, in which choice results from a tradeoff between the perceived time difference (interval) and the perceived outcome difference (compensation). This model is both more parsimonious and more intuitive than any rival discounting model of comparable scope. Moreover, it accurately describes archival data as well as data from new experiments."

We argue that the tradeoff model of Scholten and Read (2006b) can be incorporated within RT-theory. If this is accepted, then their tradeoff model lends further support to

the RT-theory and, in particular, their own discount function (2.12) and its generalization (2.13), above.

We proceed by first recasting their model in a more general form (and indicate how their model is to be obtained as a special case). However, there should be no presumption that they would agree with our reformulation. They develop their model through three successive versions. We concentrate on their fourth and final version, page 15.

Let $r \geq 0$ be the reference point for time.[19] The tradeoff model establishes preference relationships, $\prec_r$ and $\sim_r$ between outcome pairs $(x, s)$ and $(y, t)$. Thus $(x, s) \prec_r (y, t)$ if, and only if, $y$ received at time $t$ is strictly preferred to $x$ received at time $s$. Similarly, $(x, s) \sim_r (y, t)$ if, and only if, $y$ received at time $t$ is equivalent to $x$ received at time $s$. These relationship are established using three functions, a *value function*, $u$, a *tradeoff function* $Q$ and a *delay-perception function*, $w$. We make the following assumptions: $Q : [0, \infty) \overset{\text{onto}}{\to} [0, \infty)$ is strictly increasing, $w : [0, \infty) \overset{\text{onto}}{\to} [0, \infty)$ is strictly increasing (the same as in (2.13), above).[20]

First, let $x > 0$, $y > 0$ and $s \geq r \geq 0$, $t \geq r$. Then:

$$(x, s) \sim_r (y, t) \Leftrightarrow Q[w(t) - w(r)] - Q[w(s) - w(r)] = u(y) - u(x), \tag{5.1}$$

$$(x, s) \prec_r (y, t) \Leftrightarrow Q[w(t) - w(r)] - Q[w(s) - w(r)] < u(y) - u(x). \tag{5.2}$$

Second, let $x < 0$, $y < 0$ and (as before) $s \geq r \geq 0$, $t \geq r$. Then:

$$(x, s) \sim_r (y, t) \Leftrightarrow Q(w(t) - w(r)) - Q(w(s) - w(r)) = u(x) - u(y), \tag{5.3}$$

$$(x, s) \prec_r (y, t) \Leftrightarrow Q(w(t) - w(r)) - Q(w(s) - w(r)) > u(x) - u(y). \tag{5.4}$$

For completeness, we also need (again, $s \geq r \geq 0$, $t \geq r$):

$$(0, s) \sim_r (0, t), \tag{5.5}$$

$$x < 0 \Rightarrow (x, s) \prec_r (0, t), \tag{5.6}$$

$$y > 0 \Rightarrow (0, s) \prec_r (y, t), \tag{5.7}$$

$$x < 0, y > 0 \Rightarrow (x, s) \prec_r (y, t). \tag{5.8}$$

To get the tradeoff model of Read and Scholten, set $r = s$ in the above equations.[21]

[19] To ease the burden of notation, we shall suppress reference to the reference point for wealth, $w_0$. Thus, in what follows, we write $\prec_r$ and $\sim_r$ when we should have written $\prec_{r,w_0}$ and $\sim_{r,w_0}$, respectively.

[20] They explicitly state two assumptions: $Q' > 0$, $Q'' < 0$. However, in the next paragraph, they say that $Q'' > 0$ for sufficiently small intervals. So, we make no assumptions on $Q''$. They explicitly state no further assumptions on $Q$ and $w$. However, we believe our other assumptions on $Q$ and $w$ are in line with what they intend (see their equations (2) and (5) for the earlier, and simpler, versions of their model).

[21] They explicitly state only (5.1) and (5.3) (with $r = s$). However, we believe that our other equations are in line with their framework.

To define a discount function, $D$, that expresses these preferences, let

$$v(x) = e^{u(x)}, \text{ for } x > 0, \tag{5.9}$$

$$v(x) = -e^{-u(x)}, \text{ for } x < 0. \tag{5.10}$$

Then all the above relations, (5.1) to (5.8), can be summarized by the following. For all $x, y$ and all $r$, $s$, $t$ such that $s \geq r \geq 0$, $t \geq r$:

$$(x,s) \sim_r (y,t) \Leftrightarrow v(x)\, e^{-Q[w(s)-w(r)]} = v(y)\, e^{-Q[w(t)-w(r)]}, \tag{5.11}$$

$$(x,s) \prec_r (y,t) \Leftrightarrow v(x)\, e^{-Q(w(s)-w(r))} < v(y)\, e^{-Q[w(t)-w(r)]}. \tag{5.12}$$

(5.11) and (5.12) suggest we take our discount function to be the generalized RS function (2.13), which is a generalization of the discount function (2.12) of Scholten and Read (2006a).

Thus, RT-theory can incorporate the tradeoff model.

### 5.2. The certainty effect

A test of a theory (T) is always a test of T plus auxiliary assumptions (O). Thus, a refutation of T&O may be a refutation of O rather than T. However, since O is often left implicit, a refutation of T&O may be misconstrued as a refutation of T rather than O. A case in point may be T = 'exponential discounting' and O = 'uncertainty is not relevant'. In testing the common difference effect, not only is it better if subjects are paid 'real money', the delays should be realistic too, i.e., quite long. Despite the experimenters' best efforts to eliminate uncertainty, there will always be a residual risk that the subjects will not receive their promised payoffs. If subjects were expected utility (EU) maximizers, then risk would not matter (Example 3, below). However, if subjects overweight low probabilities and underweight high probabilities (as in many non-EU theories), then risk matters (Example 4, below). Moreover, the *lower* the residual risk the *greater* will be its effect! (Example 5, below.)[22] Thus, Halevy (2007) argues that the common difference effect may, in fact, be a refutation of EU rather than exponential discounting.

The above points are illustrated by the following three examples. They all involve a choice between receiving \$1000 now or \$1100 next year and, simultaneously, a choice between receiving these two sums 10 and 11 years from now, respectively. We use the SIE value function (3.6), so that

$$v(1000) = 1000.5 \text{ and } v(1100) = 1100.5. \tag{5.13}$$

[22]So, experimenters, by doing their best to reduce residual risk with the aim of getting a sharper refutation of exponential discounting are, actually, achieving the opposite.

Let the discount function be $D(s,t)$ and the probability weighting function be $w(p)$, where $p$ is the probability that the payoff will actually be paid one year from now. We assume independence across years so that the probability of receiving the payoff $t$ years from now is $p^t$. Let $(x,t)$ be the event \$$x$ is received in year $t$ and let $(x,s) \prec (y,t)$ mean $(y,t)$ is strictly preferred to $(x,s)$. We take the current level of wealth, $w_0$, and present time, $r=0$, to be the reference points for wealth and time, respectively (and, to simplify notation, we have dropped the subscripts, $w_0, r$, from $\prec_{w_0,r}$). We thus have:

$$\text{No common difference effect: } (1100,1) \prec (1000,0) \Rightarrow (1100,11) \prec (1000,10) \tag{5.14}$$

$$(1100,1) \prec (1000,0) \Leftrightarrow v(1100)\, D(0,1)\, w(p) < v(1000), \tag{5.15}$$

$$(1100,11) \prec (1000,10) \Leftrightarrow v(1100)\, D(0,11)\, w\left(p^{11}\right) < v(1000)\, D(0,10)\, w\left(p^{10}\right). \tag{5.16}$$

**Example 3** *(exponential discounting with expected utility): Assume exponential discounting, so* $D(s,t) = e^{-\beta(t-s)}$, $\beta > 0$, *and expected utility, so* $w(p) = p$. *Substitute in (5.15), (5.16) to get:*

$$(1100,1) \prec (1000,0) \Leftrightarrow v(1100)\, e^{-1} p < v(1000), \tag{5.17}$$

$$(1100,11) \prec (1000,10) \Leftrightarrow v(1100)\, e^{-11\beta} p^{11} < v(1000)\, e^{-10\beta} p^{10}, \tag{5.18}$$

*(5.18) is equivalent to:*

$$(1100,11) \prec (1000,10) \Leftrightarrow v(1100)\, e^{-1} p < v(1000). \tag{5.19}$$

*From (5.17) and (5.19) we see that* $(1100,1) \prec (1000,0) \Leftrightarrow (1100,11) \prec (1000,10)$. *Thus, exponential discounting together with expected utility*[23] *imply no common difference effect. Hence, the observation of a common difference effect is a rejection of the joint hypothesis of exponential discounting and expected utility. Thus, it would imply the rejection of one or the other (or both) but not, necessarily, exponential discounting.*

**Example 4** *(Exponential discounting with non-expected utility): We take cumulative prospect theory (Tversky and Kahneman, 1992) as our model of non-expected utility. Take* $D(r,t) = e^{-\beta(t-r)}$, $\beta = 0.04$, *and* $w(p) = e^{-(-\ln p)^{\alpha}}$ *(Prelec, 1998) with* $\alpha = 0.5$. *Let* $p = 0.98$. *Substitute in (5.15), (5.16), using (5.13), to get:*

$$(1100,1) \prec (1000,0) \Leftrightarrow 1100.5 e^{-0.04} e^{-(-\ln 0.98)^{0.5}} < 1000.5, \tag{5.20}$$

$$(1100,11) \prec (1000,10) \Leftrightarrow 1100.5 e^{-0.44} e^{-\left(-\ln(0.98)^{11}\right)^{0.5}} < 1000.5 e^{-0.4} e^{-\left(-\ln(0.98)^{10}\right)^{0.5}}. \tag{5.21}$$

[23] Or no risk so that $p = 1$ and, hence, $w(p) = w(1) = 1$.

*A calculation shows that the inequality of the right hand side of (5.20) holds while the corresponding inequality in (5.21) does not hold. Hence,* $(1100, 1) \prec (1000, 0)$ *but* $(1100, 11) \succ (1000, 10)$. *Thus, exponential discounting may be consistent with an observation of a common difference effect, if subjects do not behave according to expected utility.*

**Example 5** *(It's the certainty effect that's doing the work): The same as for Example (4), except that now* $p = 0.4$, *rather than* $p = 0.98$. *Substitute in (5.15), (5.16), using (5.13), to get:*

$$(1100, 1) \prec (1000, 0) \Leftrightarrow 1100.5e^{-0.04}e^{-(-\ln 0.4)^{0.5}} < 1000.5, \tag{5.22}$$

$$(1100, 11) \prec (1000, 10) \Leftrightarrow 1100.5e^{-0.44}e^{-\left(-\ln(0.4)^{11}\right)^{0.5}} < 1000.5e^{-0.4}e^{-\left(-\ln(0.4)^{10}\right)^{0.5}}. \tag{5.23}$$

*A calculation shows that the inequality of the right hand side of (5.22) holds and also the corresponding inequality in (5.23). Hence,* $(1100, 1) \prec (1000, 0)$ *and* $(1100, 11) \prec (1000, 10)$. *Thus, the common difference effect is due to the certainty effect in particular, rather than uncertainty as such.*

Example 5 suggests that if the common difference effect is due to the certainty effect alone, rather than a combination of the certainty effect and non-exponential discounting, then the phenomenon should disappear for probabilities around 0.4.

### 5.3. Vague time preferences

Manzini and Mariotti (2006) develop a theory of vague time preferences and discuss the psychological foundations for such an approach. The intuition behind this theory is that the choice between, say, receiving \$1000 now and \$1100 next year is clearer than the choice between these two sums received 10 and 11 years from now, respectively. MM propose three criteria to choose between $(x, t)$ and $(y, s)$. The primary criterion is to choose whichever has the highest present utility value. If the two present values are not 'significantly' different, then the subject chooses the one with the highest monetary value (secondary criteria). If they have the same monetary values, so that the secondary criterion fails, then the subject behaves according to the third criterion: 'choose the outcome that is delivered sooner'. If all three criteria fail, then the subject is indifferent. Thus, MM achieve a complete, though intransitive, ordering. In particular, indifference here is not an equivalence relationship. Suppose that two present values are significantly different if their difference is greater than $\sigma$, where $\sigma$ is positive real number. Then we can state these criteria formally as follows. $(x, t) \prec_{w_0,r} (y, s)$ if, and only if, one of the following holds[24]:

[24]More generally, $\sigma$ is a 'vagueness function', in which case '$|v(y) D(r, s) - v(x) D(r, t)| \leq \sigma$' is replaced by '$v(y) D(r, s) - v(x) D(r, t) \leq \sigma(x; r, t)$ and $v(x) D(r, t) - v(y) D(r, s) \leq \sigma(y; r, s)$'

1. $v(y) D(r,s) - v(x) D(r,t) > \sigma$, or

2. $|v(y) D(r,s) - v(x) D(r,t)| \leq \sigma$, and $x < y$, or

3. $|v(y) D(r,s) - v(x) D(r,t)| \leq \sigma$, $x = y$ and $s < t$.

Obviously, if $x$ and $y$ are vectors, then extra criteria can be added. Utility values whose difference is less than $\sigma$ are regarded as not significantly different. This could be because, for example, the decision maker is not sure of the appropriate value function or discount function to use. Therefore, the decision maker does not want the decision to depend too critically on the choice of these functions. On the other hand, the decision maker might be absolutely sure that more is better than less and sooner is better than later. Example 6, below, shows how this theory can explain the common difference effect.

**Example 6** *: Consider the choice between receiving \$1000 now and \$1100 next year and the choice between these two sums received 10 and 11 years from now, respectively. As with the examples in subsection 5.2, we use the SIE value function (3.6), so that 5.13 holds. We use the exponential discount function (2.9) with* $\beta = 0.1$ *and the reference time* $r = 0$, $D(0,t) = e^{-0.1t}$. *We take* $\sigma = 3$, *so that present utility values whose difference is less than 3 are regarded as not significantly different. Using these values, we get* $v(1000) - v(1100) e^{-0.1} = 1000.5 - 1100.5e^{-0.1} = 4.7264 > 3$. *Hence, the primary criterion holds and the decision maker prefers \$1000 now to \$1100 next year. Next,* $|v(1100) e^{-1.1} - v(1000) e^{-1}| = |1100.5e^{-1.1} - 1000.5e^{-1}| = 1.7388 < 3$. *Hence, the primary criterion fails, and the decision maker considers the second criterion. Since* $1000 < 1100$, *the second criterion holds. The decision maker prefers \$1100 received 11 years from now to \$1000 received 10 years from now. We have an illustration of the common difference effect.*

Recall, from subsection 2.7 above, that the experimental results of Roelofsma and Read (2000) supported 'sooner is better than larger' against 'larger is better than sooner'. However, if the order of the secondary criteria is reversed, so that sooner is better than larger (in agreement with the experimental results of Roelofsma and Reed, 2000), then \$1000 received 10 years from now would be better than \$1100 received 11 years from now, and we would not get a common difference effect.

However, whether MM's explanation of the common difference effect is acceptable or not, to us the main contribution of their paper lies in the use of primary and secondary criteria. This appears to us to be a more accurate description of actual decision making compared to the assumption of a single criterion.

### 5.4. Intransitive preferences and relative discounting

Ok and Masatlioglu (2007) (henceforth OM) accommodate (apparent) intransitivity, such as (2.21), by regarding it as *real*. Thus, they develop a theory of intransitive time preferences. At time 0, a decision maker has a binary relationship, $\preceq$, on the set $\chi = X \times [0, \infty)$, where $X$ is a non-empty set. Let $x, y \in X$ and $s, t \in [0, \infty)$, then $(x, s) \preceq (y, t)$ is to be interpreted as '$y$ received at time $t$ is (weakly) preferred to $x$ received at time $s$'.

Let $\sim$ and $\prec$ be the symmetric and asymmetric parts of $\preceq$, respectively. For each $t \in [0, \infty)$, $\preceq_t$ is the $t$-th time projection of $\preceq$ onto $X$, i.e., $x \preceq_t y$, if, and only if, $(x, t) \preceq (y, t)$. In particular, $\preceq_0$ is the projection of $\preceq$ onto $X$ at time 0 (and, similarly, for $\sim_t$ and $\prec_t$).

If $X$ is a metric space, then further structure can be imposed on $\preceq$. In particular (OM, p218):

**Definition 14** *(time preferences): Let $X$ be a metric space, then $\preceq$ is a time preference on $\chi$ if*
*(i) $\preceq$ is complete and continuous,*
*(ii) $\preceq_0$ is complete and transitive,*
*(iii) $\preceq_t = \preceq_0$ for each $t \in [0, \infty)$.*

In Definition 14, note that transitivity is imposed on $\preceq_0$ (and, hence, also on $\preceq_t$) but transitivity is not imposed on $\preceq$. Hence, neither $\prec$ nor $\sim$ are, necessarily, transitive. In particular, $\sim$ is *not*, in general, an equivalence relationship.

Let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}_+$ the set of non-negative reals and $\mathbb{R}_{++}$ the set of positive reals. Recall that a homeomorphism is a mapping that is 1-1, onto, continuous and its inverse is also continuous. Then

**Proposition 33** *(OM, Theorem 1)*[25]*: Let $X$ be an open interval in $\mathbb{R}$ and $\preceq$ a binary relation on $\chi$. $\preceq$ is a time preference on $\chi$ that satisfies properties (A1)-(A6) if, and only, there exit an increasing homeomorphism $U : X \to \mathbb{R}_{++}$ and a continuous map $D : \mathbb{R}^2_+ \to \mathbb{R}_{++}$ such that, for all $x, y \in X$ and $s, t \in [0, \infty)$,*

$$(x, s) \preceq (y, t) \text{ iff } U(x) \leq U(y) D(s, t), \tag{5.24}$$

*while (i) for given $s$, $D(s, t)$ is decreasing in $t$ with $D(s, \infty) = 0$, and (ii) $D(t, s) = 1/D(s, t)$.*

[25] Assumption A1 to A6 of OM are plausible and clearly stated. We do not reproduce them here because we only need the theorems that OM derive from them. Also note that $D(s, t)$ here corresponds to $\eta(t, s)$ in OM.

Suppose $s \leq t$. Then (5.24) says that $y$ received at time $t$ is (weakly) preferred to $x$ received at time $s$ if, and only if, the (undated) utility of $x$ is less or equal to the (undated) utility of $y$ discounted from time $t$ back to time $s$ by the discount factor $D(s,t)$. In this case, part (i) of Proposition 33 implies the following. Fix the time, $s$, at which $x$ is received. Let the time, $t$, at which $y$ is received, recede into the future. Then the value of the utility of $y$, discounted back to time $s$, decreases. In the limit, as the receipt of $y$ is indefinitely postponed, the value of its utility, discounted back to time $s$, approaches zero. Part (ii) of Proposition 33 says that compounding forward, from time $s$ to time $t$, is the inverse of discounting backwards from time $t$ to time $s$.

For each $r \in [0,\infty)$, let $\preceq^r$ be the restriction of $\preceq$ to $X \times [r,\infty)$, i.e., to times $t \geq r$. Thus, for $r \leq s$ and $r \leq t$, $(x,s) \preceq^r (y,t)$ if, and only if, $(x,s) \preceq (y,t)$.

We can now point to the main differences between RT and OM.

First, note that $U$ in Proposition 33 can take only positive values while $v$ in (2.1)-(2.2) takes both positive and negative values.[26] To bypass this problem, we consider only the domain of strictly positive gains. Let $w_0$ be the reference point for wealth. Take $X = \{w - w_0 : w > w_0\} = (0,\infty)$ and let $\preceq$ satisfy the conditions of Proposition 33. Let $(U,D)$ be the representation of $\preceq$ guaranteed by that Proposition.

From subsection 2.2 recall that, for each $r \in [0,\infty)$, $\preceq_{w_0,r}$ is a complete transitive order on $(-\infty,\infty) \times [r,\infty)$ and, hence, also on $X \times [r,\infty)$. The second point we wish to make is that, in general, $\preceq_{w_0,r}$, unlike $\preceq^r$, is *not* the restriction to $X \times [r,\infty)$ of some complete binary relationship on $X \times [0,\infty)$. Thus OM and RT are not compatible and neither is a special case of the other.

Third, $\preceq_{w_0,r}$ is *transitive* while, in general, $\preceq^r$ is *not transitive*. To elaborate this point, consider $(x,r)$, $(y,s)$ and $(z,t)$, where $x,y,z \in X$ and $s,t \in [r,\infty)$. Suppose $(x,r) \preceq_{w_0,r} (y,s)$ and $(y,s) \preceq_{w_0,r} (z,t)$. Since $\preceq_{w_0,r}$ is transitive, we can conclude that $(x,r) \preceq_{w_0,r} (z,t)$. Now, suppose that $(x,r) \preceq^r (y,s)$ and $(y,s) \preceq^r (z,t)$. Since $\preceq^r$ is not, in general, transitive, we *cannot* conclude that $(x,r) \preceq^r (z,t)$.[27] More generally, given a compact subset $C \subset X \times [r,\infty)$, there is no guarantee in OM that it has a maximum under $\preceq^r$ (i.e., an $m \in C$ such that $c \preceq^r m$ for all $c \in C$). This, obviously, will cause great difficulty for any economic theory formulated in the OM framework. On the other hand, in RT theory, and if $D$ is continuous, $C$ will always have a maximum under $\preceq_{w_0,r}$.

Fourth, and finally, these problems with OM can all be resolved in the special case where $\preceq$ is transitive. But then $\preceq$ would also be additive. In this case, OM would reduce to the standard discounting model.

[26] Hence, in its present formulation, OM cannot explain *gain-loss asymmetry*. However, it can explain the *magnitude effect* using the SIE value function (3.6).

[27] As OM clearly explain, it is for this reason that we should think of $D(s,t)$ in their theory as the *relative* discount function between times $s$ and $t$.

## 6. Appendix: Proofs

**Proof of Proposition 1**: Let $r \in [0,\infty)$ and $t \in [r,\infty)$. Let $\{t_n\}_{n=1}^{\infty}$ be a sequence in $[r,\infty)$ converging to $t$. We want to show that $\{D(r,t_n)\}_{n=1}^{\infty}$ converges to $D(r,t)$. It is sufficient to show that any monotone subsequence of $\{D(r,t_n)\}_{n=1}^{\infty}$ converges to $D(r,t)$. In particular, let $\{D(r,t_{n_i})\}_{i=1}^{\infty}$ be a decreasing subsequence of $\{D(r,t_n)\}_{n=1}^{\infty}$. Since $\{D(r,t_{n_i})\}_{i=1}^{\infty}$ is bounded below by $D(r,t)$, it must converge to, say, $q$, where $D(r,t) \leq q \leq D(r,t_{n_i})$, for all $i$. Since $D$ is onto, there is a $p \in [r,\infty)$ such that $D(r,p) = q$. Moreover, $t_{n_i} \leq p \leq t$, for each $i$. Suppose $D(r,t) < q$. Then $t_{n_i} < p$, for each $i$. Hence also $t_{n_i} < t$, for each $i$. But this cannot be, since $\{t_{n_i}\}_{i=1}^{\infty}$, being a subsequence of the convergent sequence $\{t_n\}_{n=1}^{\infty}$, must also converge to the same limit, $t$. Hence, $D(r,t) = q$. Hence, $\{D(r,t_{n_i})\}_{i=1}^{\infty}$ converges to $D(r,t)$. Similarly, we can show that any increasing subsequence of $\{D(r,t_n)\}_{n=1}^{\infty}$ converges to $D(r,t)$. Hence, $\{D(r,t_n)\}_{n=1}^{\infty}$ converges to $D(r,t)$. Hence, $D(r,t)$ is continuous in $t$. □

**Proof of Proposition 2**: It is straightforward to check that each of (2.9), (2.11), (2.12) and (2.13) is a continuous discount function in the sense of Definition 2. It is also straightforward to check that (2.10) is a discount function. The reason the latter is not a *continuous* discount function is that $\lim_{t\to 0+} D(0,t) < D(0,0)$. □

**Proof of Proposition 3** (Time sensitivity): Let $D(r,t)$ be a continuous discount function and $r \geq 0$. Suppose $0 < x \leq y$. From (2.1) and (2.2), it follows that $0 < v(x) \leq v(y)$ and, hence, $0 < \frac{v(x)}{v(y)} \leq 1$. Since, by Definition 2(i), $D(r,t) : [r,\infty) \overset{\text{onto}}{\to} (0,1]$, it follows that $\frac{v(x)}{v(y)} = D(r,t)$ for some $t \in [r,\infty)$. A similar argument applies if $y < x < 0$. □

**Proof of Proposition 4** (Existence of present values): Let $r \leq t$ and $y \geq 0$. Then, $0 < D(r,t) \leq 1$. Hence, $0 = v(0) \leq v(y)\,D(r,t) \leq v(y)$. Since $v$ is continuous and strictly increasing, it follows that $v(y)\,D(r,t) = v(x)$ for some $x \in [0,y]$. Similarly, if $y \leq 0$, then $v(y)\,D(r,t) = v(x)$ for some $x \in [y,0]$. □

**Proof of Proposition 5**: (a) Suppose that $D(r,t) = [\varphi(r)]^{-1}\varphi(t)$ for some strictly decreasing real valued function, $\varphi : [0,\infty) \to (0,1]$. We first check that $D$ satisfies all parts of Definition 2. Clearly, $D(r,t)$ is strictly decreasing in $t$ and strictly increasing in $r$. Let $r \in [0,\infty)$. $D(r,r) = [\varphi(r)]^{-1}\varphi(r) = 1$. Hence, for fixed $r \in [0,\infty)$, $t \mapsto D(r,t)$ maps $[r,\infty)$ into $(0,1]$. Next, let $t \in [0,\infty)$. $D(t,t) = [\varphi(t)]^{-1}\varphi(t) = 1$. Hence, for fixed $t \in [0,\infty)$, $r \longmapsto D(r,t)$ maps $[0,t]$ into $(0,1]$. Thus Definition 2(ii) also holds. Hence, $D$ is a discount function. If, in addition, $\varphi(0) = 1$, then $D(0,t) = [\varphi(0)]^{-1}\varphi(t) = \varphi(t)$.

Now suppose that $\varphi$ is onto. Let $p \in (0,1]$. Hence, also, $\varphi(r)\,p \in (0,1]$. Since $\varphi$ is onto $(0,1]$, we get $\varphi(t) = \varphi(r)\,p$ for some $t \in [0,\infty)$. But $\varphi(t) = \varphi(r)\,p \leq \varphi(r)$. Hence, $t \geq r$. We also have $D(r,t) = [\varphi(r)]^{-1}\varphi(t) = [\varphi(r)]^{-1}\varphi(r)\,p = p$. Hence, for each $r \in [0,\infty)$, $t \longmapsto D(r,t)$ maps $[r,\infty)$ onto $(0,1]$. Thus Definition 2(i) holds with 'into' replaced by

'onto'. Hence, $D$ is a continuous discount function. Since $\varphi : [0,\infty) \overset{\text{onto}}{\to} (0,1]$ is strictly decreasing, we must have $\varphi(0) = 1$. Hence, $D(0,t) = [\varphi(0)]^{-1}\varphi(t) = \varphi(t)$.

For all $r$, $s$ and $t$, $D(r,s)D(s,t) = [\varphi(r)]^{-1}\varphi(s)[\varphi(s)]^{-1}\varphi(t) = [\varphi(r)]^{-1}\varphi(t) = D(r,t)$. Hence, $D$ is additive.

(b) Suppose that $D$ is an additive discount function. Then, for all $r$, $s$, $t$, where $0 \leq r \leq s \leq t$, $D(r,s)D(s,t) = D(r,t)$. From this, it follows that, for any $s$ and any $t$ $(0 \leq s \leq t)$: $\frac{D(r,t)}{D(r,s)} = D(s,t)$, which is independent of $r$, for all $r \in [0,s]$. Similarly, for any $r$ and any $s$ $(0 \leq r \leq s)$: $\frac{D(r,t)}{D(s,t)} = D(r,s)$, which is independent of $t$, for all $t \in [s,\infty)$. This can only hold if $D(r,t) = F(r)\Phi(t)$, for all $r$ and $t$ $(0 \leq r \leq t)$. In particular, $F(r)\Phi(r) = D(r,r) = 1$. Hence, $F(r) = [\Phi(r)]^{-1}$. Hence, $D(r,t) = [\Phi(r)]^{-1}\Phi(t)$. Set $\varphi(t) = \Phi(t)/\Phi(0)$. Then $D(r,t) = [\varphi(r)]^{-1}\varphi(t)$. In particular, $D(0,t) = [\varphi(0)]^{-1}\varphi(t) = \varphi(t)$. Hence, $\varphi$ is a strictly decreasing function from $[0,\infty)$ into $(0,1]$. If $D$ is continuous, so that $D(0,t)$ is onto, then $\varphi$ is also onto. □

**Proof of Proposition 6**: Exponential: $\varphi(t) = e^{-\beta t}$. PPL: $\varphi(0) = 1$ and $\varphi(t) = e^{-\delta-\beta t}$ for $t > 0$. LP: $\varphi(t) = (1+\alpha t)^{-\frac{\beta}{\alpha}}$. Generalized RS: $\varphi(t) = e^{-Q[w(t)]}$. □

**Proof of Proposition 7** (Invariance to the choice of reference time):

$$\begin{aligned}
&((x_1,s_1),(x_2,s_2),...,(x_m,s_m)) \preceq_{w_0,r} ((y_1,t_1),(y_2,t_2),...,(y_n,t_n)),\\
&\Leftrightarrow V_r((x_1,s_1),(x_2,s_2),...,(x_m,s_m)) \leq V_r((y_1,t_1),(y_2,t_2),...,(y_n,t_n)),\\
&\Leftrightarrow D(0,r)V_r((x_1,s_1),(x_2,s_2),...,(x_m,s_m)) \leq D(0,r)V_r((y_1,t_1),(y_2,t_2),...,(y_n,t_n)),\\
&\Leftrightarrow D(0,r)\Sigma_{i=1}^m v(x_i)D(r,s_i) \leq D(0,r)\Sigma_{i=1}^n v(y_i)D(r,t_i),\\
&\Leftrightarrow \Sigma_{i=1}^m v(x_i)D(0,r)D(r,s_i) \leq \Sigma_{i=1}^n v(y_i)D(0,r)D(r,t_i),\\
&\Leftrightarrow \Sigma_{i=1}^m v(x_i)D(0,s_i) \leq \Sigma_{i=1}^n v(y_i)D(0,t_i), \text{ by additivity,}\\
&\Leftrightarrow V_0((x_1,s_1),(x_2,s_2),...,(x_m,s_m)) \leq V_0((y_1,t_1),(y_2,t_2),...,(y_n,t_n)),\\
&\Leftrightarrow ((x_1,s_1),(x_2,s_2),...,(x_m,s_m)) \preceq_{w_0,0} ((y_1,t_1),(y_2,t_2),...,(y_n,t_n)). \square
\end{aligned}$$

To facilitate the proof of Propositions 8, below, and 22, later, we first establish Lemmas 1 and 2.

**Lemma 1** *: Let $x \geq 0$ and $y \geq 0$. Then:*
*(a) $\rho \geq 1 \Rightarrow x^\rho + y^\rho \leq (x+y)^\rho$.*
*(b) $0 < \rho \leq 1 \Rightarrow x^\rho + y^\rho \geq (x+y)^\rho$.*

**Proof of Lemma 1**: Clearly, the results hold for $x = 0$. Suppose $x > 0$. Let $z = \frac{y}{x}$ and $f(z) = (1+z)^\rho - 1 - z^\rho$. Then $f(z) = 0$ and $f'(z) = \rho\left[(1+z)^{\rho-1} - z^{\rho-1}\right]$, for $z > 0$. Suppose $\rho \geq 1$. Then $f'(z) \geq 0$. Since $f$ is continuous, it follows that $f(z) \geq 0$ for $z \geq 0$. Part (a) follows from this. Now suppose $0 < \rho \leq 1$. Then $f'(z) \leq 0$. Since $f$ is continuous, it follows that $f(z) \leq 0$ for $z \geq 0$. Part (b) follows from this. □

**Lemma 2** : *Let $\tau > 0$, $0 \leq s < t$ and $r > 0$. Let $f(r) = (t+r)^{\tau} - (s+r)^{\tau} - (t^{\tau} - s^{\tau})$. Then:*
*(a)* $\tau > 1 \Rightarrow f(r) > 0$.
*(b)* $0 < \tau < 1 \Rightarrow f(r) < 0$.

**Proof of Lemma 2**: Clearly, $f(0) = 0$. Also, $f'(r) = \tau\left[(t+r)^{\tau-1} - (s+r)^{\tau-1}\right]$. If $\tau > 1$, then $f'(r) > 0$ for $r > 0$. Since $f$ is continuous, it follows that $f(r) > 0$ for $r > 0$. This establishes part (a). If $0 < \tau < 1$, then $f'(r) < 0$ for $r > 0$. Since $f$ is continuous, it follows that $f(r) < 0$ for $r > 0$. This establishes part (b). □

**Proof of Proposition 8**: (a) Suppose $0 < \rho \leq 1$. Let $0 \leq r < s < t$. From (2.12), we get $D(r,s) = [1 + \alpha(s^{\tau} - r^{\tau})^{\rho}]^{-\frac{\beta}{\alpha}}$, $D(s,t) = [1 + \alpha(t^{\tau} - s^{\tau})^{\rho}]^{-\frac{\beta}{\alpha}}$ and $D(r,t) = [1 + \alpha(t^{\tau} - r^{\tau})^{\rho}]^{-\frac{\beta}{\alpha}}$. Hence,

$$\begin{aligned} D(r,s)\,D(s,t) &= \left\{1 + \alpha\left[(s^{\tau} - r^{\tau})^{\rho} + (t^{\tau} - s^{\tau})^{\rho}\right] + \alpha^2 (s^{\tau} - r^{\tau})^{\rho} (t^{\tau} - s^{\tau})^{\rho}\right\}^{-\frac{\beta}{\alpha}}, \\ &< \left\{1 + \alpha\left[(s^{\tau} - r^{\tau})^{\rho} + (t^{\tau} - s^{\tau})^{\rho}\right]\right\}^{-\frac{\beta}{\alpha}}, \text{ since } \alpha^2 (s^{\tau} - r^{\tau})^{\rho} (t^{\tau} - s^{\tau})^{\rho} > 0 \text{ and } -\frac{\beta}{\alpha} < 0, \\ &\leq \left[1 + \alpha(s^{\tau} - r^{\tau} + t^{\tau} - s^{\tau})^{\rho}\right]^{-\frac{\beta}{\alpha}}, \text{ by Lemma 1b and since } -\frac{\beta}{\alpha} < 0, \\ &= \left[1 + \alpha(t^{\tau} - r^{\tau})^{\rho}\right]^{-\frac{\beta}{\alpha}} = D(r,t). \end{aligned}$$

(b) It is sufficient to give an example. Let $\alpha = \tau = 1$ and $\rho = 2$. Hence, $D(0,1)\,D(1,2) = 4^{-\beta} > 5^{-\beta} = D(0,2)$. Hence, for $\alpha = \tau = 1$ and $\rho = 2$, $D$ cannot be additive or subadditive. However, for the same parameter values, we have $D(0,10)\,D(10,20) = 10201^{-\beta} < 401^{-\beta} = D(0,20)$. Hence, $D$ cannot be supper additive either.[28]

(c) Let $0 \leq s < t$. (ii) is obvious from inspecting (2.12). Let $r > 0$. (iii) For $\tau > 1$, Lemma 2(a) gives $D(s+r, t+r) = \left\{1 + \alpha\left[(t+r)^{\tau} - (s+r)^{\tau}\right]^{\rho}\right\}^{-\frac{\beta}{\alpha}} < \left\{1 + \alpha\left[t^{\tau} - s^{\tau}\right]^{\rho}\right\}^{-\frac{\beta}{\alpha}}$ $= D(s,t)$. (i) For $0 < \tau < 1$, Lemma 2(b) gives $D(s+r, t+r) = \left\{1 + \alpha\left[(t+r)^{\tau} - (s+r)^{\tau}\right]^{\rho}\right\}^{-\frac{\beta}{\alpha}}$ $> \left\{1 + \alpha\left[t^{\tau} - s^{\tau}\right]^{\rho}\right\}^{-\frac{\beta}{\alpha}} = D(s,t)$. □

**Proof of Proposition 9** (Properties of a delay function): Let $D$ be a discount function and $\Psi$ and $\Phi$ two corresponding delay functions. Let $s, t \in [0, \infty)$. Then $D(0, \Phi(s,t)) = D(0,s)\,D(0,t) = D(0, \Psi(s,t))$. Since $D(0,r)$ is strictly decreasing in $r$, we must have $\Phi(s,t) = \Psi(s,t)$. This establishes (a). Using Definition 2, it is straightforward to check that properties (b) to (d) follow from Definition 7. Now, suppose $v(x) = v(y)\,D(0,t)$. Multiply both sides by $D(0,s)$ to get $v(x)\,D(0,s) = v(y)\,D(0,s)\,D(0,t) = v(y)\,D(0, \Psi(s,t))$. Conversely, suppose $v(x)\,D(0,s) = v(y)\,D(0, \Psi(s,t))$. Then $v(x)\,D(0,s) = v(y)\,D(0,s)\,D(0,t)$. Since $D(0,s) > 0$, we can cancel it to get $v(x) = v(y)\,D(0,t)$. This establishes (e) and completes the proof. □

[28] Other examples can be given to show that there is nothing special about $r = 0$, $\alpha = 1$, $\tau = 1$, or $\rho = 2$, as long as $\rho > 1$.

**Proof of Proposition 10** (Existence of a delay function): Let $D$ be a continuous discount function. Let $s,t \in [0,\infty)$. Then $D(0,s), D(0,t) \in (0,1]$. Hence also $D(0,s)\,D(0,t) \in (0,1]$. Since $r \mapsto D(0,r)$ is onto $(0,1]$, there is some $T \in [0,\infty)$ such that $D(0,s)\,D(0,t) = D(0,T)$. Since $D(0,r)$ is strictly decreasing in $r$, this $T$ is unique. Set $T = \Psi(s,t)$. The function, $\Psi(s,t)$, thus defined, is a delay function corresponding to $D$. □

**Proof of Proposition 11**: (Representation Theorem 1) Let $\alpha > 0$ and $\beta > 0$. It is easy to verify that $D(0,t) = [1+\alpha\psi(t)]^{-\frac{\beta}{\alpha}}$ is an $(\alpha,\beta)$-representation of the discount function $D(r,t)$ if, and only if, $\psi(t) = \frac{1}{\alpha}\left[D(0,t)\right]^{-\frac{\alpha}{\beta}} - 1$. From this, using Definition 2, it is straightforward to verify that part (a) holds. It follows that the inverse of $\psi$ exists and $\psi^{-1} : [0,\infty) \overset{\text{onto}}{\to} [0,\infty)$ is strictly increasing with $\psi(0) = 0$, this establishes part (b). We now turn to part (c). We have $[1+\alpha\psi(\Psi(s,t))]^{-\frac{\beta}{\alpha}} = D(0,\Psi(s,t)) = D(0,s)\,D(0,t) = [1+\alpha\psi(s)]^{-\frac{\beta}{\alpha}}[1+\alpha\psi(t)]^{-\frac{\beta}{\alpha}} = [1+\alpha\psi(s)+\alpha\psi(t)+\alpha^2\psi(s)\psi(t)]^{-\frac{\beta}{\alpha}}$. Hence, $[1+\alpha\psi(\Psi(s,t))]^{-\frac{\beta}{\alpha}} = [1+\alpha\psi(s)+\alpha\psi(t)+\alpha^2\psi(s)\psi(t)]^{-\frac{\beta}{\alpha}}$. From the latter it follows that $\psi(\Psi(s,t)) = \psi(s)+\psi(t)+\alpha\psi(s)\psi(t)$. Hence, $\Psi(s,t) = \psi^{-1}[\psi(s)+\psi(t)+\alpha\psi(s)\psi(t)]$. □

**Proof of Proposition 12** (Representation Theorem 2): Let

$$\psi : [0,\infty) \overset{\text{onto}}{\to} [0,\infty) \text{ be strictly increasing (hence, } \psi(0) = 0), \tag{6.1}$$

and let

$$\Psi(s,t) = \psi^{-1}(\psi(s)+\psi(t)+\alpha\psi(s)\psi(t)),\ \alpha > 0. \tag{6.2}$$

Let $D$ be a continuous discount function with delay function, $\Psi$. Then

$$D(0,s)\,D(0,t) = D(0,\Psi(s,t)),\ s \geq 0,\ t \geq 0. \tag{6.3}$$

From (6.2) and (6.3), we get

$$D(0,s)\,D(0,t) = D\left(0,\psi^{-1}(\psi(s)+\psi(t)+\alpha\psi(s)\psi(t))\right),\ s \geq 0,\ t \geq 0. \tag{6.4}$$

From (6.1), we get that $\psi^{-1}$ exists and $\psi^{-1} : [0,\infty) \overset{\text{onto}}{\to} [0,\infty)$ is strictly increasing with $\psi^{-1}(0) = 0$. Bear these facts in mind for when the functions $G$ and $h$ are defined, below ((6.7) and (6.10)).

Let

$$X = 1+\alpha\psi(s),\ Y = 1+\alpha\psi(t). \tag{6.5}$$

Hence $\psi(s) = \frac{X-1}{\alpha}$, $\psi(t) = \frac{Y-1}{\alpha}$, $s = \psi^{-1}\left(\frac{X-1}{\alpha}\right)$, $t = \psi^{-1}\left(\frac{Y-1}{\alpha}\right)$, $\psi(s)+\psi(t)+\alpha\psi(s)\psi(t) = \frac{XY-1}{\alpha}$ and

$$\psi^{-1}(\psi(s)+\psi(t)+\alpha\psi(s)\psi(t)) = \psi^{-1}\left(\frac{XY-1}{\alpha}\right). \tag{6.6}$$

Define the function $G:[1,\infty)\rightarrow(0,\infty)$ by

$$G(X)=D\left(0,\psi^{-1}\left(\frac{X-1}{\alpha}\right)\right). \tag{6.7}$$

Hence,

$$G(Y)=D\left(0,\psi^{-1}\left(\frac{Y-1}{\alpha}\right)\right),\ G(XY)=D\left(0,\psi^{-1}\left(\frac{XY-1}{\alpha}\right)\right). \tag{6.8}$$

From (6.4), (6.6), (6.7), (6.8) we get $G(XY)=D\left(0,\psi^{-1}\left(\frac{XY-1}{\alpha}\right)\right)=D\left(0,\psi^{-1}\left(\psi(s)+\psi(t)+\alpha\psi(s)\psi(t)\right)\right)$ $=D(0,s)D(0,t)=D\left(0,\psi^{-1}\left(\frac{X-1}{\alpha}\right)\right)D\left(0,\psi^{-1}\left(\frac{Y-1}{\alpha}\right)\right)$ and, hence,

$$G(XY)=G(X)G(Y). \tag{6.9}$$

Define[29] the function $h:[0,\infty)\rightarrow(0,\infty)$ by

$$h(y)=G(e^{y}),y\geq 0 \tag{6.10}$$

Hence, and in the light of Definition 2(i), $h$ satisfies[30]:

$$h:[0,\infty)\rightarrow(0,\infty)\ \text{is strictly decreasing and}\ h(x+y)=h(x)h(y). \tag{6.11}$$

As is well known, see for example Corollary 1.4.11 in Eichhorn (1978) or Theorem 1, page 38, of Aczel (1966), the unique solution to (6.11) is the exponential function

$$h(y)=e^{cy},\ y\geq 0,\ c<0, \tag{6.12}$$

(6.5), (6.6), (6.8), (6.10), (6.12) give, in succession, $h(y)=(e^{y})^{c}$, $G(e^{y})=(e^{y})^{c}$, $G(Y)=Y^{c}$, $D\left(0,\psi^{-1}(\psi(t))\right)=[1+\alpha\psi(t)]^{c}$,

$$D(0,t)=[1+\alpha\psi(t)]^{c}. \tag{6.13}$$

Let $\beta=-\alpha c$. Then (6.13) gives $D(0,t)=[1+\alpha\psi(t)]^{-\frac{\beta}{\alpha}}$, $\alpha>0$, $\beta>0$, $t\geq 0$, where $\beta>0$ because $\alpha>0$ and $c<0$. □

**Proof of Proposition 13** (Extension Theorem 1): (a) Let $f$ and $g$ be extension functions corresponding to the discount function $D$. Let $r\in[0,\infty)$ and $t\in[r,\infty)$. Then, by Definition 11(ii), $D(0,f(r,t))=D(r,t)=D(0,g(r,t))$. Since $D(r,s)$ is strictly decreasing in $s$, it follows that $f(r,t)=g(r,t)$. Hence, $f=g$. (b) We have $D(0,t)=D(0,f(0,t))$. Since $D(0,s)$ is strictly decreasing in $s$, it follows that $f(0,t)=t$. We

[29]It is tempting, at this stage, to take a shortcut and conclude, from the fact that $G(XY)=G(X)G(Y)$, that, necessarily, $G(X)=X^{c}$. However, the relevant theorem (Theorem 1.9.13 in Eichhorn (1978) or Theorem 3, page 41, in Aczel (1966)) requires that $G(X)$ be defined for *all* $X>0$. However, $\psi^{-1}(t)$ is not defined for $t<0$ and, hence, $G(X)$ is not defined for $X<1$.

[30]It is sufficient that $h$ be strictly decreasing in some interval: $(a,a+\delta)$, $a\geq 0,\delta>0$.

also have $D(0, f(r,r)) = D(r,r) = 1 = D(0,0)$, i.e., $D(0, f(r,r)) = D(0,0)$. Again, since $D(0,t)$ is strictly decreasing in $t$, it follows that $f(r,r) = 0$. Since $D(0, f(r,t)) = D(r,t)$ and $D(r,t)$ is strictly decreasing in $t$, it follows that $f(r,t)$ must be strictly increasing in $t$. Since $D(r,t)$ is strictly increasing in $r$, it follows that $f(r,t)$ must be strictly decreasing in $r$. This completes the proof that $f$ is an extension function. (c) Since $D(r,t) = D(0, f(r,t))$, it follows that $v(x) = v(y) D(r,t)$ if, and only if, $v(x) = v(y) D(0, f(r,t))$. □

**Proof of Proposition 14** (Extension Theorem 2): Let $r \in [0,\infty)$ and $t \in [r,\infty)$. Then $D(r,t) \in (0,1]$. Since $D$ is a continuous discount function, it follows that $s \longmapsto D(0,s)$ is onto $(0,1]$. Hence, for some $T \in (0,1]$, $D(r,t) = D(0,T)$. Since $D(r,s)$ is strictly decreasing in $s$, it follows that this $T$ is unique. Set $f(r,t) = T$. The function, $f$, thus defined has the property $D(r,t) = D(0, f(r,t))$. Hence, $f$ is an extension function corresponding to $D$. Let $r \in [0,\infty)$. Let $t \in [0,\infty)$. Then $D(0,t) \in (0,1]$. Since, $D$ is a continuous discount function, it maps $[r,\infty)$ onto $(0,1]$. Hence $D(r,s) = D(0,t)$, for some $s \in [r,\infty)$. But since $f$ is an extension function corresponding to $D$, we have $D(r,s) = D(0, f(r,s))$. Hence, $D(0, f(r,s)) = D(0,t)$. Since $D(0,q)$ is strictly decreasing in $q$, it follows that $f(r,s) = t$. Hence, $t \longmapsto f(r,t)$ maps $[r,\infty)$ onto $[0,\infty)$. □

**Proof of Proposition 15** (Extension Theorem 3): (a) Since $f : \Delta \to [0,\infty)$ and $\varphi : [0,\infty) \to (0,1]$, it follows that $\varphi \circ f : \Delta \to (0,1]$. Let $r \in [0,\infty)$. Then $t \mapsto \varphi(f(r,t))$ maps $[r,\infty)$ into $(0,1]$ and is strictly decreasing with $D(r,r) = \varphi(f(r,r)) = \varphi(0) = 1$. Let $t \in [0,\infty)$. Then $r \mapsto \varphi(f(r,t))$ maps $[0,t]$ into $(0,1]$ and is strictly decreasing. Thus, $D(r,t) = \varphi(f(r,t))$ satisfies all parts of Definition 2 and, hence, is a discount function. (b) $D(0, f(r,t)) = \varphi(f(0, f(r,t))) = \varphi(f(r,t)) = D(r,t)$. Hence, $f$ is the extension function corresponding to $D$ and $D(r,t)$ is the $f$-extension of $D(0,t)$. (c) Let $r \in [0,\infty)$. Let $p \in (0,1]$. Since $\varphi$ is onto, there exists $t \in [0,\infty)$ such that $p = \varphi(t)$. Since $s \mapsto f(r,s)$ maps $[r,\infty)$ onto $[0,\infty)$, there exists $s \in [r,\infty)$ such that $f(r,s) = t$. Hence, $D(r,s) = \varphi(f(r,s)) = \varphi(t) = p$. Hence, $s \mapsto D(r,s)$ maps $[r,\infty)$ onto $[0,\infty)$. Hence, $D$ is a continuous discount function. □

**Proof of Proposition 16** (Characterization Theorem 1): Let $D$ be a continuous discount function. Then, by Representation Theorem 1 (Proposition 11) and Extension Theorem 2 (Proposition 14) $D(r,t) = D(0, f(r,t)) = [1 + \alpha\psi(f(r,t))]^{-\frac{\beta}{\alpha}}$, where $\alpha > 0$, $\beta > 0$, $\psi$ is a continuous seed function and $f$ is a continuous extension function. Conversely, if $D(r,t) = [1 + \alpha\psi(f(r,t))]^{-\frac{\beta}{\alpha}}$, where $\alpha > 0$, $\beta > 0$, $\psi$ is a continuous seed function and $f$ is a continuous extension function, then $D$ satisfies all the conditions of Definition 2, with 'onto' replacing 'into' in part (i). Hence, $D$ is a continuous discount function. Uniqueness of $f$ follows from Extension Theorem 1 (Proposition 13). □

**Proof of Proposition 17** (Characterization Theorem 2): Follows from Definition 6 and Characterization Theorem 1 (Proposition 16). □

**Proof of Proposition 18** (Characterization Theorem 3): Let $f(r,t) = \psi^{-1}\left(\frac{\psi(t)-\psi(r)}{1+\alpha\psi(r)}\right)$. Then $D(r,t) = [1+\alpha\psi(f(r,t))]^{-\frac{\beta}{\alpha}} = \left[1+\alpha\psi\left(\psi^{-1}\left(\frac{\psi(t)-\psi(r)}{1+\alpha\psi(r)}\right)\right)\right]^{-\frac{\beta}{\alpha}} = \left[1+\alpha\frac{\psi(t)-\psi(r)}{1+\alpha\psi(r)}\right]^{-\frac{\beta}{\alpha}}$ $= \left[\frac{1+\alpha\psi(t)}{1+\alpha\psi(r)}\right]^{-\frac{\beta}{\alpha}}$, which is additive, by Proposition 5(aiii). Conversely, suppose $D$ is additive. Then, by Proposition 5(b), $D(r,t) = [\varphi(r)]^{-1}\varphi(t)$ for some strictly decreasing real valued function, $\varphi : [0,\infty) \overset{\text{onto}}{\to} (0,1]$, $\varphi(0) = 1$. Let $D(r,t) = [1+\alpha\psi(f(r,t))]^{-\frac{\beta}{\alpha}}$. Then $\varphi(t) = D(0,t) = [1+\alpha\psi(t)]^{-\frac{\beta}{\alpha}}$. Hence, $[1+\alpha\psi(f(r,t))]^{-\frac{\beta}{\alpha}} = \left[\frac{1+\alpha\psi(t)}{1+\alpha\psi(r)}\right]^{-\frac{\beta}{\alpha}}$ and, hence, $1+\alpha\psi(f(r,t)) = \frac{1+\alpha\psi(t)}{1+\alpha\psi(r)}$. From which it follows that $f(r,t) = \psi^{-1}\left(\frac{\psi(t)-\psi(r)}{1+\alpha\psi(r)}\right)$. □

**Proof of Proposition 19**: All the claims can be verified by straightforward calculations. However, when dealing with PPL, do not use Proposition 11, as PPL is not continuous. So, for example, instead of using part b of Proposition 11, check directly that $D(0,s)D(0,t) = D(0,\Psi(s,t))$. □

**Proof of Proposition 20** (Representation Theorem 3): Since $\psi(0) = 0$ and $\psi$ is strictly increasing (Proposition 11a), it follows that $\psi(t) > 0 \Leftrightarrow t > 0$. Let $s > 0$ and $t > 0$, then $D(0,s)D(0,t) < D(0,s+t) \Leftrightarrow [1+\alpha\psi(s)]^{-\frac{\beta}{\alpha}}[1+\alpha\psi(t)]^{-\frac{\beta}{\alpha}} < [1+\alpha\psi(s+t)]^{-\frac{\beta}{\alpha}}$ $\Leftrightarrow [1+\alpha\psi(s)][1+\alpha\psi(t)] > 1+\alpha\psi(s+t) \Leftrightarrow \psi(s+t) < \psi(s)+\psi(t)+\alpha\psi(s)\psi(t)$ $\Leftrightarrow s+t < \psi^{-1}[\psi(s)+\psi(t)+\alpha\psi(s)\psi(t)] \Leftrightarrow s+t < \Psi(s,t)$. From this chain it follows that (c) ⇔ (a) ⇔ (b). □

**Proof of Proposition 21** (Characterization Theorem 4): Assume $\psi_+$ is $\alpha_+$-subadditive. Let $0 < x < y$, $v(x) = v(y)D_+(0,t)$ and $s > 0$. Then $0 < v(x) < v(y)$. It follows that $D_+(0,t) < 1$ and, hence, $t > 0$. From Representation Theorem 3(c) (Proposition 20), it follows that $D_+(0,s)D_+(0,t) < D_+(0,s+t)$. Hence, $v(y)D_+(0,t)D_+(0,s) < v(y)D_+(0,s+t)$. Hence, $v(x)D_+(0,s) < v(y)D_+(0,s+t)$. Hence, the common difference effect holds for gains.

Conversely, assume that the common difference effect holds for gains. Let $s > 0$ and $t > 0$. Hence, $0 < s < s+t$. Hence, by Proposition 2.7, for some $x$, $0 \le x \le 1$, $v(x) = v(1)D_+(0,t)$. Hence, $x > 0$. Hence, $v(x)D_+(0,s) < v(1)D_+(0,s+t)$. Hence, $v(1)D_+(0,s)D_+(0,t) < v(1)D_+(0,s+t)$. Hence, $D_+(0,s)D_+(0,t) < D_+(0,s+t)$. From Representation Theorem 3 (Proposition 20), it follows that $\psi_+$ is $\alpha_+$-subadditive.

Hence, preferences exhibit the common difference effect for gains if, and only if, the seed function for gains, $\psi_+$, is $\alpha_+$-subadditive. A similar argument shows that preferences exhibit the common difference effect for losses if, and only if, the seed function for losses, $\psi_-$, is $\alpha_-$-subadditive. □

**Proof of Proposition 22**: Let $\psi_+(t) = t^{\gamma_+}$, where $0 < \gamma_+ \le 1$. Then, for $s > 0$ and $t > 0$, $\psi_+(s+t) = (s+t)^{\gamma_+} \le t^{\gamma_+} + s^{\gamma_+} < s^{\gamma_+} + t^{\gamma_+} + \alpha_+ s^{\gamma_+} t^{\gamma_+} = \psi_+(s) + \psi_+(t) + \alpha_+\psi_+(s)\psi_+(t)$. Hence, by Characterization Theorem 4 (Proposition 21), preferences

that exhibit $\gamma$-delay (Definition 13), also exhibit the common difference effect for gains. Similarly, they exhibit the common difference effect for losses. $\square$

**Proof of Proposition 23**: Let preferences with the continuous discount function, $D$, exhibit $\gamma$-delay. Then, by Definition 13, its delay function is $\Psi(s,t) = (s^\gamma + t^\gamma + \alpha s^\gamma t^\gamma)^{\frac{1}{\gamma}}$, $\alpha > 0$, $0 < \gamma \leq 1$. Hence, by Representation Theorem 2 (Proposition 12), we must have $D(0,t) = (1+\alpha t^\gamma)^{-\frac{\beta}{\alpha}}$, for some $\beta > 0$. $\square$

**Proof of Proposition 24**: By Proposition 23 we have, necessarily, $D(0,t) = (1+\alpha t^\gamma)^{-\frac{\beta}{\alpha}}$, for some $\beta > 0$. Hence, by Characterization Theorem 3 (Proposition 18) we must have $D(r,t) = \left(\frac{1+\alpha t^\gamma}{1+\alpha r^\gamma}\right)^{-\frac{\beta}{\alpha}}$. $\square$

**Proof of Proposition 25**: By Proposition 23 we have, $D(0,t) = (1+\alpha t^\gamma)^{-\frac{\beta}{\alpha}}$, for some $\beta > 0$. Hence, $D(r,t) = D(0, f(r,t)) = [1+\alpha[f(r,t)]^\gamma]^{-\frac{\beta}{\alpha}} = \left[1+\alpha\left[(t^\tau - r^\tau)^{\frac{1}{\tau}}\right]^\gamma\right]^{-\frac{\beta}{\alpha}}$ $= \left[1+\alpha(t^\tau - r^\tau)^{\frac{\gamma}{\tau}}\right]^{-\frac{\beta}{\alpha}} = [1+\alpha(t^\tau - r^\tau)^\rho]^{-\frac{\beta}{\alpha}}$. $\square$

**Proof of Proposition 26:** See LP, p583. For a more detailed proof, see al-Nowaihi and Dhami (2008).

**Proof of Proposition 27:** See LP, p584. For a more detailed proof, see al-Nowaihi and Dhami (2008).

**Proof of Proposition 28**: See al-Nowaihi and Dhami (2008).

**Proof of Proposition 29**: Consider the two consumption streams: $((0,0),(c,s),(0,s+t))$ and $((0,0),(0,s),(c,s+t))$, where $c > 0$. If the consumer receives $((0,0),(0,s),(c,s+t))$ when he was expecting $((0,0),(c,s),(0,s+t))$ then, according to prospect theory, he codes the postponement of $c$ as a loss in period 2 but a gain in period 3. According to (2.6), $V_0((0,0),(0,s),(c,s+t)) = v(0)D(0,0) + v(-c)D(0,s) + v(c)D(0,s+t)$ $< -v(c)D(0,s) + v(c)D(0,s+t)$, where the last inequality comes from $v(0) = 0$ and $v(-c) < -v(c)$. On the other hand, if the consumer receives $((0,0),(c,s),(0,s+t))$ when he was expecting $((0,0),(0,s),(c,s+t))$, he codes the bringing forward of consumption as a gain in period 2 but a loss in period 3. Hence, $V_0((0,0),(c,s),(0,s+t))$ $= v(0)D(0,0) + v(c)D(0,s) + v(-c)D(0,s+t) < v(c)D(0,s) - \lambda v(c)D(0,s+t)$. We thus have: $V_0((0,0),(c,s),(0,s+t)) + V_0((0,0),(0,s),(c,s+t)) < v(c)D(0,s) - v(c)D(0,s+t) - v(c)D(0,s) + v(c)D(0,s+t) = 0$. Hence, $V_0((0,0),(c,s),(0,s+t)) < -V_0((0,0),(0,s),(c,s+t))$. $\square$

**Proof of Proposition 30**: See al-Nowaihi and Dhami (2008).

**Proof of Proposition 31**: See al-Nowaihi and Dhami (2008).[31]

**Proof of Proposition 32**: These can be verified by direct calculation using (2.10), (2.11), (2.12), (2.13) and (3.6). $\square$

[31] al-Nowaihi and Dhami (2008) also provide a scheme for generating further increasing elasticity value functions.

## References


[1] Aczel, J., 1966. Lectures on Functional Equations and their Applications. Academic Press, New York and London.

[2] al-Nowaihi, A. and Dhami, S., 2008. A value function that explains the magnitude and sign effects. University of Leicester Discussion Paper in Economics No. 08/31.

[3] Andersen, S., Harrison, G. W., Lau, M. I., Rutstrom, E.E., 2007. Eliciting risk and time preferences. Forthcoming, Econometrica.

[4] Eichhorn, W., 1978. Functional Equations in Economics, Addison-Wesley.

[5] Frederick, S., Loewenstein, G., O'Donoghue, T., 2002. Time discounting and time preference: a critical review. Journal of Economic Literature 40, 351-401.

[6] Halevy, Y., 2007. Strotz meets Allais: Diminishing impatience and the certainty effect. Forthcoming in American Economic Review.

[7] Kahneman, D., Tversky, A., 1979. Prospect theory : an analysis of decision under risk. Econometrica 47, 263-291.

[8] Kahneman D., Tversky A., 2000. Choices, Values and Frames, Cambridge University Press, Cambridge.

[9] Laibson, D., 1997. Golden eggs and hyperbolic discounting. Quarterly Journal of Economics 112, 443-477.

[10] Loewenstein G., 1988a. Frames of mind in intertemporal choice. Management Science 34, 200-214.

[11] Loewenstein G., 1988b. The weighting of waiting: response mode effects in intertemporal choice. Working paper, Center for Decision Research, University of Chicago.

[12] Loewenstein G., Prelec D., 1992. Anomalies in intertemporal choice : evidence and an interpretation. The Quarterly Journal of Economics 107, 573-597.

[13] Manzini, P. and M. Mariotti, 2006. A vague theory of choice over time. Advances in Theoretical Economics. Berkeley Electronic Press vol. 6(1), Article 6.

[14] Manzini, P. and M. Mariotti, 2008. Choice over time, in Anand, P., Pattanaik, P. and Puppe, P., The Handbook of Rational and Social Choice, Oxford, Oxford University Press.

[15] Noor, J., 2008a. Hyperbolic discounting and the standard model. Mimeo. Boston University.

[16] Noor, J., 2008b. Anomalous money-discounting. Mimeo. Boston University.

[17] Ok, E. A., and Y. Masatlioglu, 2007. A theory of (relative) discounting. Journal of Economic Theory 137, 214-245.

[18] Phelps, E.S., Pollak, R., 1968. On second-best national saving and game-equilibrium growth. Review of Economic Studies 35, 185-199.

[19] Prelec D. (1998). The probability weighting function. *Econometrica* 60, 497-528.

[20] Read, D., 2001. Is time-discounting hyperbolic or subadditive? Journal of Risk and Uncertainty, 23, 5-32.

[21] Roelofsma, P. and Read, D. 2000. Intransitive intertemporal choice. Journal of Behavioral Decision Making 13, 161-177.

[22] Rubinstein, A. 2003. Is it economics and psychology? The case of hyperbolic discounting, International Economic Review, 44 (4), 1207-1216.

[23] Samuelson, P., 1937. A note on measurement of utility. Review of Economic Studies, 4, 155-161.

[24] Scholten, M. and Read, D. 2006a. Discounting by Intervals: A generalized model of intertemporal choice. Management Science, 52, 1424-1436.

[25] Scholten, M. and Read, D., 2006b. Beyond Discounting: The Tradeoff Model of Intertemporal. Choice. Working Paper LSE-OR 06.88. London: London School of Economics and Political Science.

[26] Thaler, R.H., 1981. Some empirical evidence on dynamic consistency. Economic Letters 8, 201-207.

[27] Thaler, R.H., 1999. Mental accounting matters. Journal of Behavioral Decision Making 12, 183-206.

[28] Tversky, A., and Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297-323.