Descriptive Models of Intertemporal Choice Part 1:

Accuracy and Applicability

Marc Scholten
ISPA University Institute,
Lisbon, Portugal

and

Daniel Read
Durham Business School,
Durham University
Durham, United Kingdom

and

Yale School of Management,
Yale University
Abstract

Two comparative models of intertemporal choice have recently been proposed as an alternative to the standard hyperbolic discounting model. One, the interval discounting model, retains the notion that intertemporal choice is governed by discounting, but proposes that discounting involves direct comparisons between the options along the time attribute. The other, the tradeoff model, discards the notion of discounting, and proposes that intertemporal choice is governed by direct comparisons along the time attribute and the money attribute.

These comparative models have a broader coverage than the standard hyperbolic discounting model, but it is yet to be shown how much more accurate they are in quantitative analyses, and whether one is more accurate than the other. We parameterize all three models, and apply them to both primary and secondary data. The practical applicability of the two discounting models is limited, and their application is problematic. Moreover, the hyperbolic discounting model performs poorly, and the interval discounting model performs worse than the tradeoff model. We conclude that, both in terms of practical applicability and in terms of predictive accuracy, the tradeoff model is the best tool for empirical analyses of intertemporal choice. While the specific parameterizations of the respective models limit the generality of this result, they were the only ones that we could justify theoretically and estimate empirically.
Descriptive Models of Intertemporal Choice Part 1:
Accuracy and Applicability

Intertemporal choices involve tradeoffs between costs and benefits that occur at different points in time. These include choices such as taking a job now or getting an education and having a chance at a better job later, and spending money now or saving it and having more to spend later. Theories of intertemporal choice have been tested on elementary choices between smaller-sooner and larger-later amounts of money, such as receiving $100 now or $150 in 3 months. Economics has provided a normative standard for these choices, and psychology has discovered preference patterns that are anomalous to that standard. This paper is about descriptive models of intertemporal choice, and the accuracy with which they predict anomalies in choices between smaller-sooner and larger-later amounts of money.

The normative standard is Samuelson’s (1937) discounted utility model. In this model, each outcome is first integrated with the baseline consumption level, a utility is assigned to the consumption level resulting from this asset integration, the utility is discounted at a constant rate as a function of the delay to the outcome, and the outcome with the highest discounted utility is chosen. The anomalies to this model have come in three waves. The first wave, initiated by Thaler (1981), showed that the rate at which an outcome is discounted is not constant, but varies with the delay to an outcome, the magnitude and sign of the outcome, and changes in its timing (i.e., delay or speedup). These anomalies were accommodated by Loewenstein and Prelec’s (1992) hyperbolic discounting model (after Ainslie, 1975, who coined the term “hyperbolic” discounting). In this model, a value is assigned to each outcome, the value is discounted at a decreasing rate as a function of the delay to the outcome, and the outcome with the highest discounted value is chosen. Although a major development, the hyperbolic discounting model agrees with the discounted utility model that choice is
alternative-based: The available options are independently assigned an overall value, these values are compared, and the option with the highest value is chosen.

The second wave of anomalies showed that discount rates vary not only with the delay to the outcomes, but also with the interval between them (Read, 2001). A model that accommodates these anomalies is the interval discounting model (Scholten & Read, 2006). In this model, the time horizon is partitioned into a series of intervals. The first interval is the time to the first outcome on the horizon; the second is the time between the first outcome and the second; and so on. The value of an outcome is then discounted as a function of all intervals that precede it. The interval discounting model therefore incorporates attribute-based choice, because the options are directly compared along the time attribute before a value is assigned. However, it also incorporates alternative-based choice, because the options are assigned an overall value, even if, in contrast with the discounted utility model and the hyperbolic discounting model, these values are not independently assigned. An outcome at a given delay can therefore be worth more or less, depending on the delays to the other outcomes.

The third wave of anomalies, initiated by Rubinstein (2001), showed that discount rates are affected by comparisons between outcomes as well as delays. These “similarity” effects have been treated by Rubinstein (2001, 2003) and Leland (2002), but their proposals do not also address the first two waves of anomalies. A proposal that does this is the tradeoff model (Scholten and Read 2010a). In this model, choice is purely attribute-based: The options are directly compared along both attributes (i.e., time and money), and the option favored by these comparisons is chosen.

The tradeoff model has been shown to offer a qualitative account of the evidence that discounting models can and cannot address (Scholten & Read, 2010a), but it is yet to be shown to what extent it can also offer a quantitative account of that evidence. In this paper, we
examine whether the tradeoff model at least matches the interval discounting model in accommodating the anomalies covered by both models (i.e., all anomalies except the similarity effects). We also examine to what extent both models improve on the hyperbolic discounting model.

We are interested in whether the tradeoff model “at least” matches the interval discounting model, because there is an argument that favors the tradeoff model over the interval discounting model, all else equal. It is the argument of *practical applicability*. Problems in the application of discounting models arise when discount rates are inversely related to outcome magnitude, a robust anomaly called the ‘absolute magnitude effect.’ In the hyperbolic discounting model and the interval discounting model, this anomaly is captured by a complex parametric specification of the value function (see al-Nowaihi & Dhami, 2009, and the specification proposed in this paper). In the tradeoff model, on the other hand, the absolute magnitude effect is captured by a simple model equation. The bottom line is that the tradeoff model can be estimated following standard procedures, whereas discounting models often force the user to resort to complex estimation techniques that may actually prevent these models from being estimated.

We proceed as follows. First, we present the *benchmark model*, according to which outcomes are discounted at a constant rate. Then we discuss how the three descriptive models under consideration account for variation in discount rates. After this theoretical discussion, we deal with practical issues in the application of descriptive models. Then we apply the three models to both primary and secondary data. The hyperbolic discounting model performs poorly, and the interval discounting model performs worse than the tradeoff model. We conclude with methodological considerations in the formal analysis of intertemporal choice.
The Benchmark Model

We focus on intertemporal choices between pairs of single dated outcomes, one smaller-but-sooner (SS), the other larger-but-later (LL). An example is the choice between $100 in 1 month and $150 in 4 months. The outcomes are designated as \( x_S \) and \( x_L \) ($100 and $150), and their respective delays as \( t_S \) and \( t_L \) (1 and 4 months). Given such a choice, a person may prefer either SS or LL, or be indifferent between them. If indifferent, the delays and outcomes can be transformed into an empirical measure of discounting. Commonly, the model used for conducting this transformation assumes exponential (or ‘constant’) discounting of outcomes. According to this model, which we will call the benchmark model, the person is indifferent between SS and LL when

\[
\delta^{t_S} x_S = \delta^{t_L} x_L, \tag{1}
\]

where \( \delta \) is the discounting over any unit interval \( t \to t + 1 \). A lower \( \delta \) indicates more discounting: A lower proportion of \( x \) remains if \( x \) is delayed by one unit of time. An alternative measure of discounting used in the literature is \( \rho = (1 - \delta) / \delta \), a lower value of which indicates less discounting. Solving Equation 1 for \( \delta \), we get

\[
\delta = \left( \frac{x_S}{x_L} \right)^{1/(t_L - t_S)}, \tag{2}
\]

which is the average discounting over the interval \( t_S \to t_L \). Under the benchmark model, \( \delta \) is constant. Solving Equation 1 for \( \delta^{t_L - t_S} \), we get

\[
\Delta = \delta^{t_L - t_S} = \frac{x_S}{x_L},
\]

which is the total discounting over the interval \( t_S \to t_L \).
Descriptive Models

In this section, we discuss how a choice between SS and LL is addressed by three descriptive models of intertemporal choice: The hyperbolic discounting, interval discounting, and tradeoff models.

The hyperbolic discounting model

In the hyperbolic discounting model (Loewenstein & Prelec, 1992), the discounted values of SS and LL are given as

\[ V(x_S; t_S) = d(t_S)v(x_S) \]
\[ V(x_L; t_L) = d(t_L)v(x_L), \]

where \( V(x; t) \) is the present value of \( x \), or its value given that it will be received after a wait of \( t \), \( v(x) \) is the value \( x \) will have when it is received, and \( d(t) \) is a discount factor decreasing in \( t \). The decision maker is indifferent between SS and LL when

\[ \frac{V(x_S; t_S)}{V(x_L; t_L)} = d(t_S)v(x_S) = d(t_L)v(x_L). \]

The total discounting of value over the interval \( t_S \to t_L \) is

\[ \frac{d(t_L)}{d(t_S)} = \frac{v(x_S)}{v(x_L)}. \]

The discount function \( d \) is given as

\[ d(t) = (1 + \alpha t)^{-\beta/\alpha}, \]

where \( \beta > 0 \) is the degree of discounting and \( \alpha > 0 \) is hyperbolic discounting. If \( \alpha \downarrow 0 \), the hyperbolic discount function reduces to the exponential discount function \( d(t) = e^{-\beta t} \). This is the discount function of the benchmark model, where \( e^{-\beta t} = \delta' \). Hyperbolic discounting is motivated by the ‘common difference effect,’ which is that there is less discounting over an interval that begins later than over one that begins earlier (Bartels & Rips, 2010; Chapman & Weber, 2006; Green, Fristoe, & Myerson, 1994; Green, Myerson, & Macaux, 2005; Holt,
Green, Myerson, & Estle, 2008; Keren & Roelofsma, 1995; Kirby & Herrnstein, 1995; McAlvanah, 2010; Scholten & Read, 2006). For instance, someone who is indifferent between $150 in 4 months and $100 in 1 month will prefer $150 in 12 months to $100 in 9 months.

Given the hyperbolic discount function in Equation 4, the total discounting of value over the interval $t_S \rightarrow t_L$ is

\[
\frac{d(t_L)}{d(t_S)} = \left(\frac{1 + \alpha t_S}{1 + \alpha t_L}\right)^{\beta/\alpha} = \left(\frac{1 + \alpha t_S}{1 + \alpha t_S + \alpha(t_L - t_S)}\right)^{\beta/\alpha} = \frac{v(x_S)}{v(x_L)}.
\]

The breakdown of $t_L$ in the third term of the equation clarifies how hyperbolic discounting leads to the common difference effect: If we add a constant $a$ to both $t_S$ and $t_L$ (i.e., if we hold interval length constant and increase the front-end delay), we have $(a + t_L) - (a + t_S) = t_L - t_S$ in the denominator, and $a + t_S$ in both the numerator and the denominator, so that the constant $a$ drives the ratio of discount factors upward, closer to one. As a result, we have

\[
\frac{d(a + t_L)}{d(a + t_S)} = \frac{v(x_S)}{v(x_L)} \text{ if } a > 0.
\]

Thus, to preserve indifference between $SS$ and $LL$, either $x_S$ has to increase or $x_L$ has to decrease, which leads to a higher $\Delta$ (less total discounting of outcomes), and, because interval length is held constant, to a higher $\delta$ as well (less average discounting of outcomes).

Loewenstein and Prelec (1992) do not specify the value function $v$, but they do identify five of its properties. Three are taken from prospect theory (Kahneman & Tversky, 1979), the other two are specific to the hyperbolic discounting model.

1. **Reference dependence**: Outcomes are evaluated as gains and losses relative to a neutral reference point, i.e., $v(0) = 0$.

2. **Diminishing absolute sensitivity**: For constant absolute increases of $x > 0$, $v(x)$ increases by decreasing absolute amounts, i.e., $v$ is concave over gains. Similarly, for constant
absolute decreases of $x < 0$, $v(x)$ decreases by decreasing absolute amounts, i.e., $v$ is convex over losses.

3. **Loss aversion**: Losses loom larger than gains, i.e., $v$ is steeper for losses than for gains. Kahneman and Tversky (1979) define this as $-v(-x) > v(x)$ for $x > 0$ (for alternative definitions of loss aversion, see Abdellaoui, Bleichrodt, & L’Haridon, 2008; Abdellaoui, Bleichrodt, & Parashiv, 2007). Tversky and Kahneman (1991) discuss constant loss aversion, which is that reversing the sign of an outcome from positive to negative increases the magnitude of its value by a multiplicative constant, i.e., $-v(-x) = \Lambda v(x)$, where $x \geq 0$ and $\Lambda > 1$.

4. **Increasing elasticity**: The value function is more elastic for large outcomes than for small ones. That is, for constant proportional increases of $x > 0$, $v(x)$ increases by increasing proportional amounts, and, for constant proportional decreases of $x < 0$, $v(x)$ decreases by increasing proportional amounts. Increasing elasticity means that, if the magnitude of both outcomes increases by the same multiplicative constant, the ratio between the values of $x_S$ and $x_L$ decreases:

$$\frac{v(mx_S)}{v(mx_L)} < \frac{v(x_S)}{v(x_L)} \text{ if } m > 1.$$

Increasing elasticity is motivated by the ‘absolute magnitude effect,’ which is that discounting over a given interval is inversely related to outcome magnitude (Thaler, 1981, and countless other studies). For instance, someone who is indifferent between $150 in 4 months and $100 in 1 month will prefer $1,500 in 4 months to $1,000 in 1 month.

5. **Loss amplification** (see Prelec & Loewenstein, 1991): The value function is more elastic for losses than for gains. That is, the proportional decrease in value that results from doubling a loss of $100 is greater than the proportional increase in value that results from doubling a gain of $100. Loss amplification means that, if the sign of the outcomes is reversed from positive to negative, the ratio between the values of $x_S$ and $x_L$ decreases:
\[
\frac{v(-x_S)}{v(-x_L)} < \frac{v(x_S)}{v(x_L)} \text{ if } x_S, x_L > 0.
\]

Loss amplification is motivated by the ‘gain-loss asymmetry,’ which is that there is less discounting over a given interval for losses than for gains (Baker, Johnson, & Bickel, 2003; McAlvanah, 2010; Murphy, Vuchinich, & Simpson, 2001; Yates & Watts, 1975). For instance, someone who is indifferent between receiving $100 in 1 month and receiving $150 in 4 months will prefer to pay $100 in 1 month rather than pay $150 in 4 months.

A parametric specification of the Loewenstein and Prelec (1992) value function has to be a compromise between two goals: The theoretical goal of incorporating its five properties and the methodological goal of specifying a model than can be estimated. The problem is how to combine increasing elasticity with diminishing absolute sensitivity (see Scholten & Read, 2010a, Appendix B). To achieve this, we specify the value function as a linear combination of two power functions, each exhibiting constant elasticity, but one exhibiting greater elasticity than the other, so that, as outcome magnitude increases, the combination of the two power functions becomes increasingly determined by the one with the greater elasticity:

\[
v(x) = \begin{cases} 
\frac{1}{1 + \gamma} + \mu \frac{\gamma}{1 + \gamma} x^{1+\gamma} & \text{if } x \geq 0 \\
-\Lambda \left[ \frac{1}{1 + \gamma} (-x)^{1+\gamma} + (\mu + \sigma) \frac{\gamma}{1 + \gamma} (-x)^{1+\gamma} \right] & \text{if } x < 0,
\end{cases}
\]

where \(\Lambda > 1\) is loss aversion (see below), \(\gamma > 0\) is diminishing absolute sensitivity, \(\mu > 0\) (\(\mu\) for ‘magnitude’) is increasing elasticity, and \(\sigma > 0\) (\(\sigma\) for ‘sign’) is loss amplification. The two parts have elasticity \(1 / (1 + \gamma)\) and \(\gamma / (1 + \gamma)\). Constant elasticity arises under two circumstances. First, it arises when \(\mu = \sigma = 0\). In this case, we obtain a value function that approaches constant sensitivity as \(\gamma\) goes to zero, and insensitivity as \(\gamma\) goes toward infinity. Second, constant elasticity arises when \(\mu, \sigma > 0\) and \(\gamma = 0\) or \(\gamma\) goes toward infinity. In either case, we obtain a value function with constant sensitivity and elasticity equal to one. Elasticity
increases with $\mu$ and $\sigma$ as long as $\gamma > 1$, i.e., as long as $\mu$ and $\sigma$ are associated with the more elastic part of the value function.

Loss aversion is that the value function is steeper for losses than for gains. In the above value function, this is driven not only by $\Lambda$, but also by $\sigma$. That is, when $\Lambda = 1$ and $\sigma > 0$, we also have loss aversion, i.e., $-v(-x) > v(x)$, and, because $\sigma$ affects not only the steepness of the value function but also its curvature, we have *increasing*, not constant, loss aversion, i.e., $-v(mx)/v(mx) > -v(-x)/v(x)$ for $m > 1$. To simplify our exposition, we refer to $\Lambda$ as ‘loss aversion’ and to $\sigma$ as ‘loss amplification.’ In the absence of loss amplification, i.e., $\sigma = 0$, we refer to $\Lambda$ as ‘*constant* loss aversion.’

**The interval discounting model**

In the hyperbolic discounting model, the values of outcomes are discounted *only* as a function of the delay to the outcomes. The implication is ‘additive discounting.’ To illustrate, suppose that two outcomes are separated by a single day, and we use two different procedures to assess the discounting over that day. In one procedure, we obtain a single measure of discounting over the whole day. In the other, we obtain separate measures of discounting over the two day segments (morning and afternoon), and then combine them into a single measure. Discounting is *additive*, when the two procedures yield the same result:

$$\frac{d(t_M)}{d(t_S)} \cdot \frac{d(t_I)}{d(t_M)} = \frac{d(t_I)}{d(t_S)},$$

where the undivided interval is denoted $t_S \rightarrow t_L$ and the two subintervals are denoted $t_S \rightarrow t_M$ and $t_M \rightarrow t_I$. Discounting is *superadditive* when there is more discounting over the whole day, and *subadditive* when there is more discounting over the two segments. Nonadditive discounting has been verified in many studies (Kinari, Ohtake, & Tsutsui, 2009; McAlvanah, 2010; Read 2001; Read & Roelofsma, 2003; Scholten & Read, 2006; Zauberman, Kim, Malkoc, & Bettman, 2009). The pattern that emerges from these studies is that subdivision of
an interval yields more discounting (subadditivity) up to a point, beyond which it yields less discounting (superadditivity).

The interval discounting model (Scholten & Read, 2006) accommodates a succession of superadditive and subadditive discounting over intervals of increasing length. In this model, the value of an outcome is discounted as a function of all intervals that precede it. The discounted values of $SS$ and $LL$ are given as

$$V(x_s;0,t_s) = D(0,t_s)v(x_s)$$

$$V(x_L;0,t_s,t_L) = D(0,t_s)D(t_s,t_L)v(x_L).$$

The decision maker is indifferent between $SS$ and $LL$ when $V(x_s;0,t_s) = V(x_L;0,t_s,t_L)$, or

$$v(x_s) = D(t_s,t_L)v(x_L).$$ \hspace{2cm} (6)

The total discounting of value over the interval $t_S \rightarrow t_L$ is

$$D(t_S,t_L) = \frac{v(x_S)}{v(x_L)},$$

where $v$ is the Loewenstein and Prelec (1992) value function and $D$ is the following interval discount function:

$$D(t_S,t_L) = \left(1 + \alpha \left(\frac{w(t_L) - w(t_S)}{\vartheta}\right)^\vartheta\right)^{-\beta/\alpha},$$ \hspace{2cm} (7)

where $\beta > 0$ is the degree of discounting, $\alpha > 0$ is subadditive discounting, $\vartheta > 1$ is superadditive discounting, and $w(t)$ is a time-weighing function that captures diminishing absolute sensitivity to delays.

The interval discount function is an inverse S-shaped discount function over differences between weighted delays. However, it differs in two ways from the one proposed by Scholten and Read (2006). First, the differences between weighted delays are divided by $\vartheta$ before being raised to the power $\vartheta$. Without the division, the inflection point of the inverse S-shaped discount function is bounded from above by 1, so that superadditivity can occur only
over a restricted range of intervals. With the division, the inflection point is bounded from above by $\beta$, so that superadditivity can occur over any range of intervals. Second, the time-weighing function is not a power function but a logarithmic function:

$$w(t) = \frac{1}{\tau} \log(1 + \tau t),$$

(8)

where $\tau > 0$ is diminishing absolute sensitivity to delays. When $t_S = 0$ and $\beta = 1$, so that diminishing absolute sensitivity to delays is confounded with subadditivity, the same discount function arises when either vanishes (i.e., either $\tau \downarrow 0$ or $\alpha \downarrow 0$). In both cases, the interval discount function reduces to the hyperbolic discount function.

Figure 1 shows average discounting under the interval discount function, in the absence of diminishing absolute sensitivity and in the presence of subadditive discounting (dotted line, less discounting over longer intervals), superadditive discounting (dashed line, more discounting over longer intervals), and both subadditive and superadditive discounting (solid line, more discounting followed by less discounting over longer intervals).

<Insert Figure 1 about here>

The tradeoff model

The interval discounting model partially retreats from the discounting paradigm, because the options are directly compared along the time attribute before value is discounted. The tradeoff model (Scholten & Read, 2010a) abandons discounting altogether by proposing that the options are directly compared along the money attribute as well.

In the tradeoff model, the advantage that a smaller-sooner gain or a larger-later loss has along the time attribute, denoted $f(t_S, t_L) > 0$, is weighted against the advantage that a larger-later gain or a smaller-sooner loss has along the money attribute, denoted $g(x_S, x_L) > 0$. Thus, the decision maker is indifferent between SS and LL when

$$f(t_S, t_L) = g(x_S, x_L).$$

(9)
The advantages can be decomposed into three weighing functions: Two intra-attribute weighing functions, \( w \) and \( v \), and one inter-attribute weighing function, \( Q \). Thus, indifference arises when

\[
Q(w(t_L) - w(t_S)) = \begin{cases} 
& v(x_L) - v(x_S) \quad \text{if } x_L > x_S > 0, \\
& v(x_S) - v(x_L) \quad \text{if } 0 > x_S > x_L. 
\end{cases}
\] (10)

The time-weighing function \( w \) weighs delays against one another, and the value function \( v \) weighs outcomes against one another. The tradeoff function \( Q \) weighs the absolute difference between weighted delays (the effective interval) against the absolute difference between valued outcomes (the effective compensation). This tradeoff function can be given as follows (see also Scholten & Read, 2010a, caption to Figure 3):

\[
Q(w(t_L) - w(t_S)) = \frac{\kappa}{\alpha} \log \left( 1 + \alpha \left( \frac{w(t_L) - w(t_S)}{\varrho} \right)^\varrho \right),
\] (11)

where \( \kappa > 0 \) is a currency parameter that translates effective intervals and effective compensations into a common unit, \( \alpha > 0 \) is subadditivity, \( \varrho > 1 \) is superadditivity, and \( w(t) \) is the time-weighing function from Equation 8. When \( t_S = 0 \) and \( \varrho = 1 \), so that diminishing absolute sensitivity to delays is confounded with subadditivity, the same tradeoff function arises when either vanish (i.e., either \( \tau \downarrow 0 \) or \( \alpha \downarrow 0 \)). Like the interval discount function \( D \), the tradeoff function \( Q \) accommodates a succession of superadditivity and subadditivity over intervals of increasing length.

The time-weighing function and the value function share three properties: Reference dependence, diminishing absolute sensitivity, and augmenting proportional sensitivity. The value function has a fourth property of its own: Constant loss aversion.

1. **Reference dependence.** This is a prerequisite for the other properties.

2. **Diminishing absolute sensitivity.** This property accounts for the common difference effect, because
Descriptive Models of Intertemporal Choice

3. **Augmenting proportional sensitivity**: For constant proportional increases of \( t \) and \( x > 0 \), \( w(t) \) and \( v(x) \) increase by increasing absolute amounts, and, for constant proportional decreases of \( x < 0 \), \( v(x) \) decreases by increasing absolute amounts. Augmenting proportional sensitivity accounts for the absolute magnitude effect, because
\[
v(mx) - v(mx) > v(x) - v(x) \quad \text{if} \quad xS, xL > 0 \quad \text{and} \quad m > 1.
\]

4. **Constant loss aversion**: This property accounts for the gain-loss asymmetry, because
\[
v(-x) - v(-x) = -\Lambda v(x) - [\Lambda v(x)] > v(x) - v(x) \quad \text{if} \quad xS, xL > 0 \quad \text{and} \quad \Lambda > 1.
\]

Scholten and Read (2010a) proposed a logarithmic value function, the formal analogue of the logarithmic time-weighing function in Equation 8, to capture the above properties and to satisfy some additional requirements:

\[
v(x) = \begin{cases} 
\frac{1}{\gamma} \log(1 + \gamma x) & \text{if} \quad x \geq 0 \\
-\frac{\Lambda}{\gamma} \log(1 + \gamma(-x)) & \text{if} \quad x < 0,
\end{cases}
\]

where \( \Lambda > 1 \) is constant loss aversion and \( \gamma > 0 \) is diminishing absolute sensitivity. This value function approaches constant sensitivity as \( \gamma \) goes to zero, and insensitivity as \( \gamma \) goes to infinity. In contrast with the value function of the discounting models, the logarithmic value function is *decreasingly* elastic.

This concludes our exposition of the three competing models. We next turn to methodological issues in estimating and evaluating them.

**Modeling Indifference**

**Data collection**

The point of indifference between SS and LL is usually determined by fixing the two delays and one outcome, and then locating the variable outcome that yields indifference.
between SS and LL, denoted \( x \). We will analyze data from two choice-based matching studies, in which \( x \) is adjusted in response to a series of choices until the largest amount that yields preference for one option is sufficiently close to the smallest amount that yields preference for the other option. The midpoint between the two amounts is taken as the indifference point.

**Model estimation and evaluation**

If possible, the models will be estimated explicitly, but if they cannot be solved for \( x \), they will be estimated implicitly. All estimation will be done with the Hooke-Jeeves and quasi-Newton routine in the nonlinear estimation module of Statistica (StatSoft 2003). Explicit estimation is done by minimizing the least-squares loss function

\[
\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 ,
\]

where \( y_i \) and \( \hat{y}_i \) are the observed and predicted values of the dependent variable, and \( n \) is the number of data points (option pairs). The null model predicts a constant value of the dependent variable. The choice of the dependent variable determines the meaningfulness of the null model.

Discounting models are usually estimated and evaluated on \( x \) (e.g., Green et al., 2005; Kirby 1997; Johnson & Bickel, 2002; McKerchar et al., 2009; Murphy et al., 2001; Rachlin, Raineri, & Cross, 1991; Scholten & Read, 2006; Simpson & Vuchinich, 2000), which does not yield a meaningful null model. When \( x \) serves as dependent variable, the null model predicts a constant value of \( x \), which is not the prediction of the benchmark model. To clarify this, we solve the benchmark model for \( x_S \):

\[
x_S = \delta^{t_S - t_L} x_L .
\]

The benchmark model predicts two general discounting effects. First, larger outcomes are discounted by a greater amount than smaller ones. For instance, if $10 is discounted by $5 over a given period, $100 will be discounted by more than $5 over the same period. This is the ‘proportional magnitude effect’ (Scholten & Read, 2010a), which is accommodated by all
discounting models through their multiplicative model equations, and by the tradeoff model through diminishing absolute sensitivity. Second, outcomes are discounted by a greater amount over longer delays than over shorter ones. For instance, if $10 is discounted by $5 over 1 period, it will be discounted by more than $5 over 2 periods. This effect is so basic that it is accommodated both by the multiplicative model equations of discounting models and by the additive model equation of the tradeoff model. In sum, the benchmark model predicts that \( x_S \) will vary as a function of the magnitude of \( x_L \) and the length of \( t_L - t_S \). Both the benchmark model and the alternative models can come out favorably by correctly predicting these general discounting effects.

The solution is to estimate and evaluate models on \( \delta \). The null model predicts a constant value of \( \delta \), which is precisely the prediction of the benchmark model. Alternative models predict that \( \delta \) will vary. Their predictions can be derived by solving Equation 3 (hyperbolic discounting model), Equation 6 (interval discounting model), and Equation 10 (tradeoff model) for \( x \), and then inserting \( x \) into Equation 2, together with the delays and the fixed outcome, to yield the predicted values of \( \delta \).

To evaluate the predictive accuracy of the alternative models, we consider two aspects of model complexity: Number of free parameters and functional form (Pitt & Myung, 2002). Number of free parameters is taken into account by the adjusted goodness-of-fit measure

\[
R^2_{\text{adj}} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \cdot \frac{n-1}{n-k},
\]

where \( \bar{y} \) is the average value of the dependent variable, and \( k \) is the number of free parameters. A larger \( k \) yields a larger downward adjustment in fit, meaning that \( R^2_{\text{adj}} \) penalizes less parsimonious models. The average value of \( y \) is the best ordinary least-squares estimate of a constant value of \( y \), so that, for the null model, \( \hat{y}_i = \bar{y} \) and \( R^2_{\text{adj}} = 0 \). This means that,
when \( \delta \) serves as the dependent variable, \( R^2_{\text{adj}} \) directly compares the predictive accuracy of the alternative models with that of the benchmark model.\(^7\)

Both number of free parameters and functional form are taken into account by the cross-validated goodness-of-fit measure

\[
R^2_{\text{CV}} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},
\]

where predicted values, denoted \( \hat{y}_i \), are obtained through cross-validation. We apply \( n \)-fold cross-validation, perhaps better known as the leave-one-out method (Efron & Tibshirani, 1993; for an application to intertemporal choice, see Keller & Strazzera, 2002). This method consists in estimating a model on \( n - 1 \) data points, employing the estimated model to predict the value of the data point that was left out, and repeating this for all \( n \) data points.

A complication arises when applying the discounting models to indifference data that contain the absolute magnitude effect. To accommodate this effect, discounting models must incorporate an increasingly elastic value function. Increasing elasticity, in combination with diminishing absolute sensitivity, yields a model that cannot be estimated explicitly on indifference data. Specifically, the identities \( d(t_S)v(x_S) = d(t_L)v(x_L) \) in Equation 3 (hyperbolic discounting model) and \( v(x_S) = D(t_S, t_L)v(x_L) \) in Equation 6 (interval discounting model) can no longer be solved explicitly for either \( x_S \) or \( x_L \). Thus, the discounting models must be estimated \textit{implicitly}, which is done by minimizing the following loss functions:

\[
\text{Hyperbolic discounting model: } \sum \left( \frac{d(t_L)v(x_L)}{d(t_S)v(x_S)} - 1 \right)^2. \quad (13)
\]

\[
\text{Interval discounting model: } \sum \left( \frac{D(t_S, t_L)v(x_L)}{v(x_S)} - 1 \right)^2. \quad (14)
\]

This is implicit estimation because the observed value of \( x \), the variable outcome, is now ‘inside’ the model to be estimated. Once the models are estimated, Equations 3 and 6 must be
solved numerically: For each data point, the model equations must be solved for the predicted value of $x$, given the estimated values of the parameters and the stated values of the independent variables. This will be done with the fsolve routine of Maple (MapleSoft, 2002). The predicted values of $x$ and the stated values of the independent variables are then inserted into Equation 2 to yield predicted values of $\delta$.

Implicit estimation poses several problems for discounting models. First, it can be done only on $x$, and not on $\delta$. Second, it can produce degenerate solutions, which are parameter configurations that ‘minimize’ the loss function to zero, regardless of the data. Given the discount functions in Equations 4 and 7 and the value function in Equation 5, the loss functions will reach their minimum value of zero by setting $\beta = \mu = \sigma = 0$, and letting $\gamma$ go toward infinity. The precise specification of the value function in Equation 5 was motivated by the attempt to avoid degenerate solutions. In sum, because discounting models can only be estimated implicitly on indifference data, their practical applicability is limited, and their application is problematic.

The tradeoff model can continue to be estimated explicitly on $x$ or $\delta$. However, when the value function is not a constant sensitivity function, explicit and implicit estimation on $x$ are not equivalent. Therefore, the tradeoff model will also be estimated implicitly on $x$, so that its predictive accuracy can be directly compared with that of the discounting models. This is done by minimizing the following loss function:

$$\sum \left( \frac{g(x_s, x_L)}{f(t_s, t_L)} - 1 \right)^2.$$  

Data aggregation

The models will be estimated and evaluated on aggregate data. In the literature, aggregate indifference data have been obtained by taking arithmetic means of $\rho$ (e.g., Benzion, Rapoport, & Yagil, 1989), geometric means of $\rho$ (e.g., Chapman, 1996), arithmetic
means of $\delta$ (e.g., Read & Roelofsma, 2003), or geometric means of $\delta$ (e.g., Scholten & Read, 2006). The advantage of using geometric means of $\delta$ is that the geometric mean of $\delta$ is the same as $\delta$ computed from the geometric mean of $x$. However, we will use the arithmetic mean of $\log(\delta)$ instead, because equal intervals on the $\log(\delta)$ scale correspond to equal intervals on the $\log(x)$ scale. Therefore, $R_{\text{adj}}^2$ for $\log(\delta)$ and $R_{\text{adj}}^2$ for $\log(x)$ will differ only because $\log(x)$ contains general discounting effects whereas $\log(\delta)$ does not.

**Modeling Preference**

Indifference between SS and LL is a special case in which the two options are equally preferred. The models may be generalized to analyze any strength of preference between SS and LL. In our analyses, strength of preference is defined as the odds of choosing LL (gains) or SS (losses), i.e., $\Omega = P / (1 - P)$, where $P$ is the probability of choosing LL (gains) or SS (losses). To predict choice odds, we use the following specifications of the three models:

Hyperbolic discounting model:

$$\Omega = \left( \frac{d(t_L)v(x_L)}{d(t_S)v(x_S)} \right)^{1/\varepsilon}.$$  \hspace{1cm} (16)

Interval discounting model:

$$\Omega = \left( \frac{D(t_S,t_L)v(x_L)}{v(x_S)} \right)^{1/\varepsilon}.$$  \hspace{1cm} (17)

Tradeoff model:

$$\Omega = \left( \frac{g(x_S,x_L)}{f(t_S,t_L)} \right)^{1/\varepsilon}.$$  \hspace{1cm} (18)

In these equations, $\varepsilon > 0$ is a “noise parameter” (Andersen, Harrison, Lau, & Rutström, 2010). As $\varepsilon$ approaches zero, choice is entirely determined by the model, i.e., $P = 0, 1$. As $\varepsilon$ approaches infinity, choice is not at all determined by the model, and is only noise, i.e., $P = \frac{1}{2}$. The special case of indifference arises when the numerator is equal to the denominator, so that $P = \frac{1}{2}$ or $\Omega = 1$, regardless of the value of $\varepsilon$. In this case, Equations 16, 17, and 18 reduce to Equations 3, 6, and 9, respectively. To obtain a preference scale with equal intervals, we take $\log(\Omega)$ as the dependent variable.
A disadvantage of choice odds is that they contain general discounting effects. For instance, someone who is equally likely to choose “$5 now” and “$10 in 1 month” is more likely to choose “$95 now” than “$100 in 1 month,” and more likely to choose “$5 now” than “$10 in 2 months.” Because the benchmark model predicts general discounting effects, \( R_{adj}^2 \) for the benchmark model will not be zero, so that \( R_{adj}^2 \) for the alternative models does not compare the predictive accuracy of these models with that of the benchmark model. To compare the predictive accuracy of an alternative model with that of the benchmark model, we compute generalized adjusted \( R^2 \) (see Anderson-Sprecher 1994):

\[
GR_{adj}^2 = \frac{R_{adj}^2 (\text{alternative model}) - R_{adj}^2 (\text{benchmark model})}{1 - R_{adj}^2 (\text{benchmark model})}.
\]

Thus, \( GR_{adj}^2 \) expresses the additional variance accounted for by the alternative model as a proportion of the variance that is not accounted for by the benchmark model. When applying the leave-one-out method, we compute \( GR_{CV}^2 \) from \( R_{CV}^2 \) (alternative model) and \( R_{CV}^2 \) (benchmark model).

This concludes our discussion of methodological issues in estimating and evaluating the three competing models. We next apply the models to several data sets.

**Indifference Data from Choice-Based Matching:**

**A Re-Analysis of Scholten and Read (2006)**

This was a choice-based matching study, which yielded indifference points for nine pairs of delayed gains, corresponding to the nine intervals displayed in Figure 2. These comprised six short intervals of 1 week, two medium-length intervals of 3 weeks, and one long interval of 17 weeks. There was a medium-length interval at the beginning and end of the long interval. Each medium-length interval spanned three short intervals. We obtained an
indifference point for a £500 gain at the beginning of each interval and a larger gain at the end.

There were three results: A higher $\delta$ over the four late intervals than over the four early ones (common difference effect), a higher $\delta$ over the six short intervals than over the two medium-length ones (superadditivity), and a higher $\delta$ over the long interval than over the two medium-length ones (subadditivity).

All models were estimated explicitly on the arithmetic means of $\log(\delta)$ and $\log(x_L)$. The hyperbolic discounting model and the interval discounting model could be estimated explicitly because $x_S$ was always a gain of the same magnitude (£500), so that the diminishing-sensitivity parameter ($\gamma$) could not be identified, and the value function could be given as $v(x) = x$.

Table 1 reports the goodness-of-fit measures. The benchmark model achieved a 91.5% fit to $\log(x_L)$, which, by comparison with its 0% fit to $\log(\delta)$, suggests that general discounting effects eclipsed any anomalies to the benchmark model. When estimated and evaluated on $\log(\delta)$, the tradeoff model and the interval discounting model yielded, on average, a 36.5% better fit than the hyperbolic discounting model. The tradeoff model yielded a 2% better fit than the interval discounting model.

Application of the leave-one-out method was problematic because of the small number of data points. By leaving out the long interval of 17 weeks, there was no data point on which to estimate the subadditivity parameter ($\alpha$) of the interval discount function and the tradeoff function, so that the long interval was not left out, and, by leaving out one medium-length interval of 3 weeks, there was only one data point (the other medium-length interval) on which to estimate the superadditivity parameter ($\vartheta$), which penalized all models. When cross-
validated on log($x_L$), the hyperbolic discounting model performed worse than the benchmark model, while the interval discounting model and the tradeoff model performed better. The tradeoff model yielded a 4.5% better fit than the interval discounting model. Its advantage increased to 7.5% when the predictive accuracy of both models was compared to that of the benchmark model. When cross-validated on log($\delta$), the tradeoff model and the interval discounting model yielded, on average, an 18% better fit than the hyperbolic discounting model. The tradeoff model yielded a 9.5% better fit than the interval discounting model.

Table 2 reports parameter estimates and significance tests upon estimating the models explicitly on log($\delta$). The significance tests are applied to the differences between parameter estimates and neutral parameter values. The parameter estimates show diminishing absolute sensitivity ($\alpha$ in the hyperbolic discounting model, $\tau$ in the interval discounting model and the tradeoff model), subadditivity ($\alpha$ in the interval discounting model and the tradeoff model), and superadditivity ($\vartheta$). The subadditivity parameter ($\alpha$) is not very reliable, but it could be estimated from only one data point (the long interval of 17 weeks).

Figure 3 depicts observed and predicted values of $\delta$. There are two major inaccuracies in the predictions made by the hyperbolic discounting model. First, $\delta$ is predicted to be the same for short intervals and medium-length intervals. This is invalidated by superadditivity. Second, $\delta$ is predicted to be lower for the long interval than for the medium-length interval that ends at the same time. This is invalidated by subadditivity.

In our next analysis, we broaden the scope to a differential weighing of outcomes as well as delays, and to states of differential as well as equal preference between SS and LL.
Indifference and Preference Data from Choice-Based Matching

Participants

The participants were 34 students (nine females and 25 males) from the London School of Economics and 18 students (15 females and three males) from ISPA University Institute, in Lisbon. The participants were paid $10 in their local currency (£5 in London, €7.50 in Lisbon).

Design and procedure

The design included three within-participant factors: The delays to the outcomes (standard, additively increased, or multiplicatively increased; see Figure 4), the magnitude of the outcomes (small or large), and their sign (positive or negative). Orthogonal manipulation of the three factors yielded 12 option pairs.

<Insert Figure 4 about here>

The presentation order of the option pairs was randomized across participants. For each pair, we conducted a choice-based matching procedure, as described by Scholten and Read (2006), to close in on the indifference point. The outcome chosen on the first trial was adjusted. The adjustment on the second trial was always in a downward direction, yielding a smaller gain or a larger loss, so as to preclude that the participants could aim at an infinitely large gain or a zero loss. The direction of the adjustment on subsequent trials depended on the participants’ choices. For instance, if SS was chosen on trial 1, and then chosen again on trial 2, the adjustment on trial 3 was in a downward direction, but, if LL was chosen on trial 2, the adjustment on trial 3 was in an upward direction.

Indifference data

For each participant and each option pair, we computed $\log(\delta)$. A repeated measures analysis of variance on $\log(\delta)$ showed a higher $\delta$ for large outcomes than for small ones, $F(1, 51) = 25.00, p = .00$ (absolute magnitude effect), a higher $\delta$ for losses than for gains, $F(1, 51)$
Descriptive Models of Intertemporal Choice 25

= 29.06, \( p = .00 \) (gain-loss asymmetry), a higher \( \delta \) for the additively increased delays than for the standard ones, \( t(51) = 4.87, p = .00 \) (common difference effect), and a higher \( \delta \) for the multiplicatively increased delays than for the additively increased ones, \( t(51) = 5.29, p = .00 \)

The latter result, a higher \( \delta \) for the multiplicatively increased delays than for the additively increased ones, indicates that the common difference effect was outweighed by subadditivity. As illustrated in Figure 4, the interval spanned by the multiplicatively increased delays, denoted \( M \), can be divided into an early interval \( E \), an intermediate interval spanned by the additively increased delays, denoted \( A \), and a late interval \( L \). The common difference effect is that, because \( E \) is longer than \( L \), \( \delta \) will be higher for \( A \) than for \( M \).\(^{10}\) Conversely, subadditivity is that, because \( M \) is longer than \( A \), \( \delta \) will be higher for \( M \) than for \( A \). We obtained the latter result, showing that the common difference effect was outweighed by subadditivity.

The indifference data were aggregated by taking, for each option pair, the geometric means of the outcomes, \( x_S \) and \( x_L \), at the indifference point. Mean outcomes, and the delays to the outcomes, are given in Table 3.

<Insert Table 3 about here>

The benchmark model and the tradeoff model were estimated explicitly on \( \log(x_L) \) and \( \log(\delta) \). The hyperbolic discounting model and the interval discounting model had to be estimated implicitly on \( x_L \), and so the tradeoff model was estimated implicitly on \( x_L \) as well. The superadditivity parameter (\( \vartheta \) in the tradeoff model and the interval discounting model) could not be identified. We report the results for \( \vartheta = 1 \), but other parameter configurations, with \( \vartheta > 1 \), are equally predictive of the dependent variables, i.e., \( \log(x_L) \) and \( \log(\delta) \).

Table 4 reports the goodness-of-fit measures. When cross-validated on \( x_L \) and evaluated on \( \log(\delta) \), the interval discounting model and the tradeoff model yielded, on average, a 45% better fit than the hyperbolic discounting model. The tradeoff model yielded a
1% better fit than the interval discounting model. Its advantage increased to 2% when it was cross-validated on log(\(\delta\)) rather than \(x_L\).

Table 5 reports parameter estimates and significance tests upon estimating the models implicitly on \(x_L\). The hyperbolic discounting model shows the signs of a degenerate solution: No reliable parameter estimates, but, instead, parameter estimates that ‘minimize’ the loss function to zero, i.e., \(\beta\), \(\mu\), and \(\sigma\) close to zero, and \(\gamma\) far removed from zero. The parameter estimates of the other two models show diminishing absolute sensitivity to delays (\(\tau\)), subadditivity (\(\alpha\)), diminishing absolute sensitivity to outcomes (\(\gamma\)), increasing elasticity (\(\mu\) in the interval discounting model), loss amplification (\(\sigma\) in the interval discounting model), and constant loss aversion (\(\Lambda\) in the tradeoff model). The two models differ mainly with respect to the parameters that capture the gain-loss asymmetry: In the interval discounting model, loss amplification (\(\sigma\)) is the least reliable parameter, whereas, in the tradeoff model, constant loss aversion (\(\Lambda\)) is the most reliable one.

In research on risky choice, estimates of prospect theory tend to show loss aversion in combination with symmetric elasticity (in which case we have constant loss aversion) or loss amplification (for a review, see Booij, van Praag, & van de Kuilen, 2010). In our research on intertemporal choice, the hyperbolic discounting model does not allow us to estimate loss aversion in addition to loss amplification, but the tradeoff model does allow us to estimate asymmetric elasticity, either alone or in addition to loss aversion. Both specifications of the tradeoff model were estimated explicitly on log(\(\delta\)). Asymmetric elasticity slightly reduced the adjusted goodness-of-fit, both when it was estimated alone (-1.04%) and when it was estimated in addition to loss aversion (-0.13%). When estimated alone, asymmetric elasticity took on the form of loss amplification. That is logical, because it replaced loss aversion in
creating a steeper value function for losses than for gains, thus generating the gain-loss asymmetry. When estimated in addition to loss aversion, however, asymmetric elasticity took on the form of ‘gain amplification’ (instances of which have also emerged in risky choice; see Wakker, Köbberling, & Schwieren, 2007, Note 1, and Scholten & Read, 2010b). Thus, the tradeoff model reversed two properties of the discounting models: Gain rather than loss amplification, and decreasing rather than increasing elasticity.

Figure 5 depicts observed and predicted values of $\delta$. While the predictions of the tradeoff model and the interval discounting model are on target, those of the hyperbolic discounting model are inaccurate. The most striking inaccuracy is that $\delta$ is predicted to be higher for the additively increased delays than for the multiplicatively increased ones. This is because the hyperbolic discounting model accommodates the common difference effect but not subadditivity (which, in these data, outweighs the common difference effect).

Preference data

For each option pair, we computed the proportion of participants who chose LL (gains) or SS (losses) on the first trial of the choice-based matching procedure. The option pairs and choice probabilities are given in Table 6.

The models were estimated explicitly on log($\Omega$). However, the subadditivity parameter ($\alpha$ in the tradeoff model and the interval discounting model) could not be identified. We report the results for $\alpha \downarrow 0$, in which case the interval discounting model reduces to the hyperbolic discounting model, but other parameter configurations, with $\alpha > 0$, are equally predictive of log($\Omega$), so that we refer to the resulting model as the ‘discounting model.’
Table 7 reports the goodness-of-fit measures. The predictions of log(Ω) are less accurate than the predictions of log(xL) or even log(δ). Upon cross-validation, the tradeoff model yielded a 7% better fit than the discounting model.

<Insert Table 7 about here>

Table 8 reports parameter estimates and significance tests. Again, the models differ mainly with respect to the parameters that capture the gain-loss asymmetry: In the discounting model, loss amplification (σ) is the least reliable parameter, whereas, in the tradeoff model, constant loss aversion (Λ) is the most reliable one.

<Insert Table 8 about here>

Figure 6 depicts observed and predicted values of Ω. The absolute magnitude effect for standard delays and multiplicatively increased delays is fairly irregular, for which the discounting model and the tradeoff model do not have an explanation.

<Insert Figure 6 about here>

We have thus far analyzed small data sets from choice-based matching studies. We next analyze a larger data set from a choice study.

**Preference Data from Choice**

*Participants*

The participants were 128 on-line visitors to eLab, the Yale School of Management virtual laboratory. There were 52 men and 76 women, averaging 34 years of age. The great majority had some college or a Bachelor’s degree. They were paid by being entered in a lottery in which they had a 1 in 50 chance of winning a $50 Amazon.com gift certificate.

*Design and procedure*

The design included four within-participant factors: The delay to the outcomes (standard or additively increased), the interval between the outcomes (undivided or divided...
Descriptive Models of Intertemporal Choice

into three subintervals), the magnitude of the outcomes (small or large), and their sign (positive or negative). Orthogonal manipulation of these factors yielded 32 option pairs.

One master list of option pairs was created under the restriction that none of their features ($t_S, t_L, x_S,$ and $x_L$) would be repeated from one trial to the other. In the master list, the four features were minimally correlated with their ordinal position ($r = .05, .01, .04,$ and $.03,$ respectively). The program cycled four times through the master list to determine where in the list a particular participant would begin and end. Thus, for instance, the first participant would begin with the first option pair in the list and end with the thirty-second, whereas the second participant would begin with the second option pair in the list and end with the first, and so on.

**Data**

We determined whether the participants chose $LL$ (gains) or $SS$ (losses). A repeated measures analysis of variance on these choices showed a higher probability of choosing $LL$ (gains) or $SS$ (losses), denoted $P$, for large outcomes than for small ones, $F(1, 127) = 71.01, p = .00$ (absolute magnitude effect), a higher $P$ for losses than for gains, $F(1, 127) = 35.85, p = .00$ (gain-loss asymmetry), a higher $P$ for the additively increased delays than for the standard delays, $F(1, 127) = 27.03, p = .00$ (common difference effect), and a lower $P$ for the earliest subinterval (.66) than for the later subintervals (.70 and .72, respectively) and the undivided interval (.71), $t(127) = -4.19, p = .00$. Option pairs and choice probabilities are given in Table 9.

The models were estimated explicitly on $\log(\Omega)$. In all analyses, the loss function had a flat minimum and the parameters had large standard errors, meaning that different parameter configurations yielded almost the same fit. However, large differences between the models
arose in the degree to which they minimized the loss function, and in the degree to which they suffered from the parameter identification problem.

The benchmark model performed worse than the null model, yielding a 17% higher Root Mean Square Deviation (RMSD) between predictions and observations. The hyperbolic discounting model, the interval discounting model, and the tradeoff model achieved a 17.5%, 61.5%, and 80% fit, respectively. Upon cross-validation, the hyperbolic discounting model performed worse than the null model, yielding a 5.5% higher RMSD, while the interval discounting model and the tradeoff model achieved a 37.5% and 76.5% fit, respectively.

Overall, the tradeoff model minimized the loss function much better than the discounting models, especially upon cross-validation. Because, in the leave-one-out method, predictive accuracy declines with parameter instability across the data points left out, we can conclude that the tradeoff model suffered much less from the parameter identification problem than the discounting models.

The tradeoff model also offered a solution to the parameter identification problem. The solution was to declare the noise parameter $\varepsilon$ at the value of 1, in which case the choice rule in Equation 18 reduces to Restle's (1961) model. Adjusted goodness-of-fit (also adjusted for the $\Psi$-parameter) increased slightly, but still rounded to 80%, and cross-validated goodness-of-fit increased slightly to 77%. Table 10 reports parameter estimates and significance tests (in which the $\Psi$-parameter also consumed one degree of freedom). The parameter estimates show diminishing absolute sensitivity to delays ($\tau$) and outcomes ($\gamma$), and constant loss aversion ($\Lambda$). According to the tradeoff model, there is neither subadditivity ($\alpha$) nor superadditivity ($\varphi$) in these data. The discounting models did not produce this result, but they performed poorly, and perhaps it is because they failed to produce the null result that they performed poorly.
Figure 7 depicts observed and predicted values of $\Omega$. The tradeoff model closely reproduces the qualitative patterns in the data. However, it underestimates the gain-loss asymmetry for small outcomes, and overestimates the gain-loss asymmetry for large outcomes.

General Discussion

The goal of this paper was to compare three descriptive models of intertemporal choice: The standard hyperbolic discounting model (an alternative-based model), the interval discounting model (a partly-alternative-partly-attribute-based model), and the tradeoff model (an attribute-based model). In three quantitative analyses, the hyperbolic discounting model performed poorly, and the interval discounting model performed worse than the tradeoff model. Thus, in terms of predictive accuracy, the tradeoff model is the best tool for empirical analyses of intertemporal choice. While the specific parameterizations of the respective models limit the generality of this result, they were the only ones that we could justify theoretically and estimate empirically.

That the hyperbolic discounting model performed worse than the two competing models comes as no surprise, because the data contained preference patterns that the hyperbolic discounting model does not accommodate. However, our purpose was exactly to determine how poor the performance of the hyperbolic discounting model would be in this kind of circumstances.

Concerning the comparison between the interval discounting model and the tradeoff model, our goal was to determine whether the tradeoff model “at least” matches the interval discounting model in terms of predictive accuracy, because, in terms of practical applicability, discounting models are at a disadvantage. The tradeoff model can be estimated explicitly or implicitly on any data, whereas discounting models, when applied to indifference data, can
only be estimated implicitly, which, given the danger of degenerate solutions, may actually prevent these models from being estimated. Thus, also in terms of practical applicability, the tradeoff model is the best tool for empirical analyses of intertemporal choice.

The question arises why the hyperbolic discounting model performed poorly, when it is often considered the “gold standard,” and it has previously been shown to perform very well. There are several reasons. First, the hyperbolic discounting model is usually applied to data in which the effects of the interval between outcomes (i.e., subadditivity, superadditivity) cannot be disentangled from the effect of the delay to outcomes. When applied to sufficiently rich data, from a design in which interval and delay are manipulated separately, the hyperbolic discounting model has low predictive accuracy. Second, the hyperbolic discounting model has so far not been applied as it should. As proposed by Loewenstein and Prelec (1992), it has an increasingly elastic value function to accommodate the absolute magnitude effect. However, the typical approach is (1) to assume hyperbolic discounting of outcomes, and not values, (2) to apply the hyperbolic discounting model within, and not across, outcomes of different magnitude, and (3) to accommodate the absolute magnitude effect with multiple, outcome-dependent discounting parameters, and not a single value function over all outcomes. Third, the hyperbolic discounting model is usually estimated on $x$ (the outcome that yields indifference between the options), and not on $\delta$ or $\rho$ (one-period discount measures). The former includes general discounting effects, which inflate ‘predictive accuracy.’

The latter point merits emphasis: A high $R^2$ does not necessarily signal high predictive accuracy. It may simply signal an overly naïve null model. In all our analyses conducted on $x$ (the outcome that yields indifference between $SS$ and $LL$), even the benchmark model, with its constant discounting of outcomes, yielded a goodness-of-fit that was well in the nineties. This is because it captures general discounting effects, which appear to be pervasive in indifference data. However, the benchmark model misses all ‘anomalous’ preference patterns, which are
the focus of the alternative models. Given this focus, the ‘correct’ measure for comparing the predictive accuracy of the alternative models on indifference data is one that removes general discounting effects. Using such a measure, the predictive accuracy of the benchmark model is zero.

We also considered preference data, a practice that is not very common in quantitative analyses of intertemporal choice. Preference data also include general discounting effects, but they cannot be removed. In our analysis of preference data, we assumed that choice probabilities derive from the ratio between discounted values (discounting models) or pairwise advantages (tradeoff model), not between their differences (Stott 2006; see also Worthy, Maddox, & Markman, 2008). In discounting models, the difference rule means that the probability of choosing $LL$ is given as $P = F(V_L - V_S)$. Prominent instances are Thurstone’s (1927) Case V and Luce’s (1959) choice axiom (difference interpretation), in which $F$ is the normal and logistic distribution function, respectively. The ratio rule means that the probability of choosing $LL$ (gains) or $SS$ (losses) is given as $P = 1 / (1 + G(V_S) / G(V_L))$. Herrnstein’s (1961, 1970) matching law and Luce’s (1959) choice axiom (ratio interpretation) arise when $G$ is the identity and power function, respectively (Stott, 2006).\footnote{We used Luce’s ratio rule. One reason was that the discounting models would have to generate the absolute magnitude effect and the gain-loss asymmetry on the basis of ratios between discounted values, just as they do with indifference data. The other reason was simply that all models performed poorly with a difference rule.}

In this paper, we only analyzed indifference data from choice-based matching. We did not analyze indifference data from matching, in which the participant directly states the outcome that yields indifference between $SS$ and $LL$. Choice-based matching provides high-quality data, but it is more labor-intensive than matching, and therefore typically provides fewer data points. Moreover, whereas choice involves an evenhanded comparison between the
options, matching involves a focus on one of the options while the other option is mentally adjusted so as to establish equivalence with the focal option (see also Abdellaouei, Attema, & Bleichrodt, 2010). As a result, matching \( LL \) to \( SS \) is seen as receiving or paying a compensation for delaying \( x_S \), and matching \( SS \) to \( LL \) is seen as receiving or paying a compensation for speeding up \( x_L \). This typically gives rise to a ‘delay-speedup asymmetry,’ which is that a positive outcome is discounted more, and a negative outcome discounted less, when it is delayed than when it is “sped up” over the same interval (Benzion et al., 1989; Shelley, 1993; see also Loewenstein, 1988, Experiment 3; Malkoc & Zauberman, 2006; Weber et al., 2007). Scholten and Read (2010a) proposed a version of the tradeoff model in which people partially adapt to the outcome to be delayed or “sped up,” meaning that this outcome is evaluated from another reference point than the outcome that compensates for the delay or speedup. Partial adaptation could also be incorporated into discounting models. We take on the modeling of data containing the delay-speedup asymmetry in a companion paper (Scholten & Read, 2011).

Bleichrodt and Johannesson (2001, p. 280) argued that “an important thing to keep in mind in the development of alternative discounting models is the trade-off between theoretical soundness and practical applicability. One reason that many researchers continue to use constant rate discounting in spite of the observed violations is its practical appeal. The danger of developing alternative intertemporal models is that they quickly become too complicated to be useful in practical research. The true challenge for future research is to develop models that strike a good balance between theoretical soundness and practical applicability.” We suggest that the tradeoff model strikes the right balance between the two criteria, and that discounting models do not: Either they fail on the first criterion (benchmark model) or on the second (interval discounting model) or on both (hyperbolic discounting model). However, we do not suggest that ‘constant rate discounting’ should be abolished altogether: We recommend that it
be used, and used more systematically, as a *benchmark model*. But, for practical research to be useful itself, we *also* need a model that is theoretically sound. The tradeoff model was the only model that did strike a good balance between theoretical soundness and practical applicability.

A natural disclaimer is that our analyses were restricted to choices between *SS* and *LL*, which are the focus of most experimental literature on intertemporal choice. Our conclusions cannot be readily extended, for instance, to choices between *sequences* of outcomes. The development of the tradeoff model for such choices is nontrivial. Discounting models can readily be applied to such choices, but they arrive at the wrong prediction that people will prefer decreasing over increasing sequences (e.g., Loewenstein & Prelec, 1993). Clearly, our research cannot be treated as a definitive statement on the relative merits of comparative and non-comparative models of intertemporal choice. However, we do conclude that, for the most elementary of intertemporal choices, i.e., choices between *SS* and *LL*, discounting models are a problem, and the tradeoff model is a solution.
References


Authors’ Note

Marc Scholten, associate professor with habilitation at ISPA University Institute, Lisbon, Portugal. Daniel Read, professor of behavioral economics at Durham Business School, Durham University, Durham, United Kingdom, and visiting professor of behavioral economics at Yale School of Management, Yale University.

We are very grateful for the financial support received from the Fundação para a Ciência e Tecnologia (FCT), program POCI 2010, and projects POCI/PSI/56030/2004, PPCDT/PSI/56030/2004, and PTDC/PSIPCO/101447/2008. We are indebted to Joaquim Judice (Universidade de Coimbra) and João Patrício (Instituto Politécnico de Tomar) for the generosity with which they helped us solve the estimation problems posed by the discounting models.

Correspondence concerning this paper should be addressed to Marc Scholten, ISPA University Institute, Rua Jardim do Tabaco 34, 1149-041 Lisboa, Portugal. E-Mail: scholten@ispa.pt.
## TABLE 1

*Indifference data from Scholten and Read (2006): Adjusted goodness-of-fit ($R^2_{adj} \times 100$) and cross-validated goodness-of-fit ($R^2_{CV} \times 100$) upon estimating the models explicitly on $\log(x_L)$ and $\log(\delta)$.*

<table>
<thead>
<tr>
<th>Model</th>
<th>Leave-all-in</th>
<th>Leave-one-out</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\log(x_L)$</td>
<td>$\log(\delta)$</td>
</tr>
<tr>
<td>Benchmark model</td>
<td>91.29</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Hyperbolic discounting model</td>
<td>94.63</td>
<td>37.68</td>
</tr>
<tr>
<td></td>
<td>(38.37)</td>
<td></td>
</tr>
<tr>
<td>Interval discounting model</td>
<td>99.56</td>
<td>72.98</td>
</tr>
<tr>
<td></td>
<td>(94.94)</td>
<td></td>
</tr>
<tr>
<td>Tradeoff model</td>
<td>99.65</td>
<td>74.96</td>
</tr>
<tr>
<td></td>
<td>(96.02)</td>
<td></td>
</tr>
</tbody>
</table>

* $GR^2_{adj} \times 100$ (leave-all-in) and $GR^2_{CV} \times 100$ (leave-one-out) in parentheses.*
### Table 2

Indifference data from Scholten and Read (2006): Parameter estimates and statistical tests for the hyperbolic discounting model, the interval discounting model, and the tradeoff model, upon estimating the models explicitly on \( \log(\delta) \).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Hyperbolic discounting model</th>
<th>Interval discounting model</th>
<th>Tradeoff model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>0.08</td>
<td>6.88</td>
<td>0.00</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.05</td>
<td>1.67</td>
<td>0.07</td>
</tr>
<tr>
<td>( \tau )</td>
<td>0.03</td>
<td>2.61</td>
<td>0.02</td>
</tr>
<tr>
<td>( \vartheta )</td>
<td>1.79</td>
<td>2.04</td>
<td>0.05</td>
</tr>
</tbody>
</table>

*a* The models were estimated on nine data points collected from 42 participants.

*b* Testing whether \( \beta, \kappa, \alpha, \text{ and } \tau \) are reliably greater than zero, and whether \( \vartheta \) is reliably greater than one (one-tailed t-tests).
<table>
<thead>
<tr>
<th>$t_s^a$</th>
<th>$t_l^a$</th>
<th>$x_s^b$</th>
<th>$x_l^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9</td>
<td>15.94</td>
<td>33.11</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>-29.27</td>
<td>-44.45</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>184.72</td>
<td>296.61</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>-301.87</td>
<td>-430.45</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>19.19</td>
<td>29.00</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>-32.27</td>
<td>-40.62</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>191.52</td>
<td>261.15</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>-345.80</td>
<td>-400.95</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>13.55</td>
<td>36.46</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>-28.79</td>
<td>-44.26</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>176.18</td>
<td>331.57</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>-282.65</td>
<td>-445.06</td>
</tr>
</tbody>
</table>

$^a$Delays in months.

$^b$Geometric means of outcomes, in pounds or euros.
TABLE 4
Indifference data from the choice-based matching study: Adjusted goodness-of-fit \((R_{adj}^2 \times 100)\) and cross-validated goodness-of-fit \((R_{CV}^2 \times 100)\).\(^a\)

<table>
<thead>
<tr>
<th>Model</th>
<th>Implicit estimation on (x_L)</th>
<th>Explicit estimation on (\log(x_L)) and (\log(\delta))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Leave-all-in</td>
<td>Leave-one-out</td>
</tr>
<tr>
<td></td>
<td>(\log(x_L))</td>
<td>(\log(\delta))</td>
</tr>
<tr>
<td>Hyperbolic discounting model</td>
<td>97.43</td>
<td>45.24</td>
</tr>
<tr>
<td>Interval discounting model</td>
<td>99.88</td>
<td>96.80</td>
</tr>
<tr>
<td>Tradeoff model</td>
<td>99.83</td>
<td>97.27</td>
</tr>
<tr>
<td>Benchmark model</td>
<td>94.92</td>
<td>0.00</td>
</tr>
<tr>
<td>Tradeoff model</td>
<td>99.85</td>
<td>98.11</td>
</tr>
</tbody>
</table>

\(^a\) \(GR_{adj}^2 \times 100\) (leave-all-in) and \(GR_{CV}^2 \times 100\) (leave-one-out) in parentheses.
### TABLE 5

Indifference data from the choice-based matching study: Parameter estimates and statistical tests for the hyperbolic discounting model, the interval discounting model, and the tradeoff model, upon estimating the models implicitly on \( x^L \).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>( t(7)^b )</th>
<th>( p )</th>
<th>Parameter</th>
<th>Estimate</th>
<th>( t(6)^b )</th>
<th>( p )</th>
<th>Parameter</th>
<th>Estimate</th>
<th>( t(7)^b )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>0.03</td>
<td>0.54</td>
<td>0.30</td>
<td>( \beta )</td>
<td>0.22</td>
<td>2.56</td>
<td>0.02</td>
<td>( \kappa )</td>
<td>4.50</td>
<td>1.92</td>
<td>0.05</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.26</td>
<td>0.74</td>
<td>0.24</td>
<td>( \alpha )</td>
<td>1.37</td>
<td>1.93</td>
<td>0.05</td>
<td>( \alpha )</td>
<td>1.31</td>
<td>1.56</td>
<td>0.08</td>
</tr>
<tr>
<td>( \tau )</td>
<td>0.19</td>
<td>2.38</td>
<td>0.03</td>
<td>( \tau )</td>
<td>0.17</td>
<td>1.99</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma )</td>
<td>7.66</td>
<td>1.44</td>
<td>0.14</td>
<td>( \gamma )</td>
<td>3.64</td>
<td>6.13</td>
<td>0.00</td>
<td>( \gamma )</td>
<td>0.09</td>
<td>3.44</td>
<td>0.01</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( 6 \times 10^{-6} )</td>
<td>0.06</td>
<td>0.48</td>
<td>( \mu )</td>
<td>0.01</td>
<td>2.58</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma )</td>
<td>( 3 \times 10^{-4} )</td>
<td>0.28</td>
<td>0.39</td>
<td>( \sigma )</td>
<td>0.12</td>
<td>1.03</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Lambda )</td>
<td></td>
<td></td>
<td></td>
<td>( \Lambda )</td>
<td>1.60</td>
<td>5.21</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The models were estimated on 12 data points collected from 52 participants.

*Testing whether \( \beta, \kappa, \alpha, \tau, \gamma, \mu, \) and \( \sigma \) are reliably greater than zero, and whether \( \Lambda \) is reliably greater than one (one-tailed \( t \)-tests).*
TABLE 6
Preference data from the choice-based matching study: Delays, outcomes, and choice probabilities
(N = 52).

<table>
<thead>
<tr>
<th>$t_S^a$</th>
<th>$t_L^a$</th>
<th>$x_S^b$</th>
<th>$x_L^b$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9</td>
<td>20</td>
<td>40</td>
<td>.481</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>-20</td>
<td>-40</td>
<td>.731</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>200</td>
<td>400</td>
<td>.712</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>-200</td>
<td>-400</td>
<td>.769</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>20</td>
<td>40</td>
<td>.750</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>-20</td>
<td>-40</td>
<td>.904</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>200</td>
<td>400</td>
<td>.865</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>-200</td>
<td>-400</td>
<td>.981</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>20</td>
<td>40</td>
<td>.269</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>-20</td>
<td>-40</td>
<td>.692</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>200</td>
<td>400</td>
<td>.481</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>-200</td>
<td>-400</td>
<td>.712</td>
</tr>
</tbody>
</table>

$^a$Delays in months.

$^b$Outcomes in pounds or euros.
TABLE 7
Preference data from the choice-based matching study: Adjusted goodness-of-fit ($R_{adj}^2 \times 100$) and cross-validated goodness-of-fit ($R_{CV}^2 \times 100$), upon estimating the models explicitly on log($\Omega$).\textsuperscript{a}

<table>
<thead>
<tr>
<th>Model</th>
<th>Leave-all-in</th>
<th>Leave-one-out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark model</td>
<td>21.49</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Discounting model</td>
<td>86.48</td>
<td>70.59</td>
</tr>
<tr>
<td></td>
<td>(82.78)</td>
<td></td>
</tr>
<tr>
<td>Tradeoff model</td>
<td>88.07</td>
<td>77.68</td>
</tr>
<tr>
<td></td>
<td>(84.80)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} $GR_{adj}^2 \times 100$ in parentheses.
TABLE 8
Preference data from the choice-based matching study: Parameter estimates and statistical tests for the interval discounting model and the tradeoff model, upon estimating the models explicitly on log(Ω).\(^a\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Discounting model</th>
<th>Tradeoff model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>(t(6)^b)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.17</td>
<td>1.61</td>
</tr>
<tr>
<td>(\tau)</td>
<td>0.47</td>
<td>1.41</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>3.67</td>
<td>2.70</td>
</tr>
<tr>
<td>(\mu)</td>
<td>0.01</td>
<td>1.36</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>0.07</td>
<td>1.02</td>
</tr>
<tr>
<td>(\Lambda)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\varepsilon)</td>
<td>0.13</td>
<td>4.96</td>
</tr>
</tbody>
</table>

\(^a\)The models were estimated on 12 data points collected from 52 participants.

\(^b\)Testing whether \(\beta\), \(\kappa\), \(\tau\), \(\gamma\), \(\mu\), \(\sigma\), and \(\varepsilon\) are reliably greater than zero and whether \(\Lambda\) is reliably greater than one (one-tailed \(t\)-tests).
TABLE 9
Preference data from the choice study: Delays, outcomes, and choice probabilities (N = 128).

<table>
<thead>
<tr>
<th>$t_S^a$</th>
<th>$t_L^a$</th>
<th>$x_S^b$</th>
<th>$x_L^b$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9</td>
<td>35</td>
<td>50</td>
<td>.469</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>35</td>
<td>39</td>
<td>.344</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>39</td>
<td>44</td>
<td>.422</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>44</td>
<td>50</td>
<td>.445</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>35</td>
<td>50</td>
<td>.539</td>
</tr>
<tr>
<td>23</td>
<td>25</td>
<td>35</td>
<td>39</td>
<td>.422</td>
</tr>
<tr>
<td>25</td>
<td>27</td>
<td>39</td>
<td>44</td>
<td>.469</td>
</tr>
<tr>
<td>27</td>
<td>29</td>
<td>44</td>
<td>50</td>
<td>.539</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>350</td>
<td>500</td>
<td>.687</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>350</td>
<td>390</td>
<td>.648</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>390</td>
<td>440</td>
<td>.750</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>440</td>
<td>500</td>
<td>.719</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>350</td>
<td>500</td>
<td>.781</td>
</tr>
<tr>
<td>23</td>
<td>25</td>
<td>350</td>
<td>390</td>
<td>.727</td>
</tr>
<tr>
<td>25</td>
<td>27</td>
<td>390</td>
<td>440</td>
<td>.773</td>
</tr>
<tr>
<td>27</td>
<td>29</td>
<td>440</td>
<td>500</td>
<td>.812</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$t_S^a$</th>
<th>$t_L^a$</th>
<th>$x_S^b$</th>
<th>$x_L^b$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9</td>
<td>-35</td>
<td>-50</td>
<td>.734</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>-35</td>
<td>-39</td>
<td>.727</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>-39</td>
<td>-44</td>
<td>.758</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>-44</td>
<td>-50</td>
<td>.781</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>-35</td>
<td>-50</td>
<td>.844</td>
</tr>
<tr>
<td>23</td>
<td>25</td>
<td>-35</td>
<td>-39</td>
<td>.758</td>
</tr>
<tr>
<td>25</td>
<td>27</td>
<td>-39</td>
<td>-44</td>
<td>.805</td>
</tr>
<tr>
<td>27</td>
<td>29</td>
<td>-44</td>
<td>-50</td>
<td>.820</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>-350</td>
<td>-500</td>
<td>.742</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>-350</td>
<td>-390</td>
<td>.805</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>-390</td>
<td>-440</td>
<td>.781</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>-440</td>
<td>-500</td>
<td>.781</td>
</tr>
<tr>
<td>23</td>
<td>29</td>
<td>-350</td>
<td>-500</td>
<td>.875</td>
</tr>
<tr>
<td>23</td>
<td>25</td>
<td>-350</td>
<td>-390</td>
<td>.836</td>
</tr>
<tr>
<td>25</td>
<td>27</td>
<td>-390</td>
<td>-440</td>
<td>.852</td>
</tr>
<tr>
<td>27</td>
<td>29</td>
<td>-440</td>
<td>-500</td>
<td>.844</td>
</tr>
</tbody>
</table>

$^a$Delays in months.

$^b$Outcomes in dollars.
TABLE 10

Preference data from the choice study: Parameter estimates and statistical tests for the tradeoff model, upon estimating the model explicitly on log(Ω).a

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>t(25)b</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>κ</td>
<td>1.96</td>
<td>4.00</td>
<td>.00</td>
</tr>
<tr>
<td>α</td>
<td>0.04</td>
<td>0.48</td>
<td>.35</td>
</tr>
<tr>
<td>τ</td>
<td>0.03</td>
<td>2.34</td>
<td>.01</td>
</tr>
<tr>
<td>θ</td>
<td>1.00</td>
<td>0.00</td>
<td>.50</td>
</tr>
<tr>
<td>γ</td>
<td>0.02</td>
<td>4.17</td>
<td>.00</td>
</tr>
<tr>
<td>Λ</td>
<td>2.59</td>
<td>5.57</td>
<td>.00</td>
</tr>
<tr>
<td>ε</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The model was estimated on 32 data points collected from 128 participants.

Testing whether κ, α, τ, and γ are reliably greater than zero, and whether θ and Λ are reliably greater than one (one-tailed t-tests).
Figure Captions

Figure 1. Average discounting under the interval discount function. The dotted line represents subadditivity only ($\beta = \alpha = \theta = 1$ and $\tau \downarrow 0$), in which case $\delta$ increases with interval length. The dashed line represents superadditivity only ($\beta = 1$, $\alpha \downarrow 0$, $\theta = 1.5$, and $\tau \downarrow 0$), in which case $\delta$ decreases with interval length. The solid line represents both subadditivity and superadditivity ($\beta = 1$, $\alpha = 1$, $\theta = 1.5$, and $\tau \downarrow 0$), in which case $\delta$ first decreases and then increases with interval length.

Figure 2. The nine intervals used by Scholten and Read (2006).

Figure 3. Indifference data from Scholten and Read (2006): The effect of interval length and delay to interval onset. Predicted values of $\delta$ obtained by estimating the models explicitly on log($\delta$). Values of $\delta$ displayed along a logarithmic scale.

Figure 4. The six delays used in the choice-based matching study: Standard delays (3 and 9), additively increased delays (23 and 29), and multiplicatively increased delays (12 and 36). The interval spanned by the multiplicatively increased delays, denoted $M$, is divided into an early interval $E$, an intermediate interval spanned by the additively increased delays, denoted $A$, and a late interval $L$.

Figure 5. Indifference data from the choice-based matching study: The effect of outcome magnitude and outcome sign for standard delays ($t$), additively increased delays ($a + t$), and multiplicatively increased delays ($m \times t$). Predicted values of $\delta$ obtained by estimating the models implicitly on $x_L$. Values of $\delta$ displayed along a logarithmic scale.

Figure 6. Preference data from the choice-based matching study: The effect of outcome magnitude and outcome sign for standard delays ($t$), additively increased delays ($a + t$), and multiplicatively increased delays ($m \times t$). Predicted values of $\Omega$ obtained by estimating the models explicitly on log($\Omega$). Values of $\Omega$ displayed along a logarithmic scale.
Figure 7. Preference data from the choice study. The top panel shows the effect of outcome magnitude and sign for standard delays ($t$) and additively increased delays ($a + t$). The bottom panel shows the effect of interval length (on a magnified scale). Predicted values of $\Omega$ obtained by estimating the models explicitly on log($\Omega$). Values of $\Omega$ displayed along a logarithmic scale.
FIG. 1

\[ \delta = D(t_S, t_L)^{1/(t_L - t_S)} \]
FIG. 2

Long

Medium

Short

Weeks
FIG. 3

Delay to interval onset

- Short
- Long

Data
Hyperbolic discounting model
Interval discounting model
Tradeoff model

\( \delta \)
FIG. 4

Now  3  9  12  23  29  36

E   A   L

M
FIG. 5

Discounting or Tradeoffs?

Data

Hypermolic discounting model

Interval discounting model

Tradeoff model

Delays

- t
- a + t
- m × t
FIG. 6

Discounting or Tradeoffs? 60

Data
Discounting model
Tradeoff model

Delays

- t
- a + t
- m x t
Footnotes

1In the discounted utility model, what is discounted at a constant rate is not the outcome itself, but rather the utility of the consumption level that the person attains through the outcome. Thus, some evidence may seem consistent with, not anomalous to, the discounted utility model. For instance, losses are discounted less than gains. This observation, by itself, is consistent with a concave utility function over consumption levels. However, because the utility function is closer to linearity for small changes in consumption levels than for large ones, the asymmetry between losses and gains should be less pronounced for small outcomes than for large ones, contrary to what is usually observed (Loewenstein & Prelec, 1992).


3Killeen (2009) presents an additive discount function, which, unlike the multiplicative utility discount function in discounting models, subtracts the subjective time to an outcome from the utility of the outcome. His approach, however, does not address similarity effects and (all variants of) interval effects.

4Diminishing absolute sensitivity means that the elasticity of the value function lies between 0 and 1.

5In our original presentation of the interval discounting model (Scholten & Read, 2006), we designated the discounted values as $V(x_s,t_s)$ and $V(x_L,t_L)$. As pointed out by Ali al-Nowaihi and Sanjit Dhami (personal communication, April 11, 2008), this notation is imprecise, in that it does not identify all the delays that define the intervals over which an outcome is discounted.

6Many other measures attempt to strike a balance between parsimony and goodness-of-fit, such as Akaike’s Information Criterion, Amemiya’s Prediction Criterion, and Schwarz’s Bayesian Information Criterion, but $R^2_{adj}$ yields the smallest adjustment for a larger number of
Discounting or Tradeoffs? 63

parameters (Kennedy, 1998, p. 103). This is convenient for our purposes, because it yields the mildest penalization of the less parsimonious discounting models, and, therefore, the most conservative criterion for evaluating the comparative merits of the tradeoff model.

7In nonlinear estimation (and in linear estimation without an intercept), the total sum of squares (SST) from the null model can actually be smaller than the residual sum of squares (SSR) from the alternative model, in which case \( R^2 = 1 - \frac{SSR}{SST} \), and, therefore, \( R^2_{\text{adj}} \), become negative (see also Anderson-Sprecher, 1994). In this case, we report the Root Mean Square Deviation of each model, i.e., \( \sqrt{\text{SST}} \) and \( \sqrt{\text{SSR}} \).

8This method thus avoids the problem, discussed by Kirby (1997), that the arithmetic mean of individual discount rates is not the same as an aggregate discount rate computed from the arithmetic mean of \( x \).

9Taking the logarithm of \( \delta \), we get

\[
\log(\delta) = \frac{1}{t_L - t_S} \left[ \log(x_S) - \log(x_L) \right].
\]

Thus, with \( t_L - t_S \) and \( x_L \) held constant, equal differences between \( \log(\delta) \) correspond to equal differences between \( \log(x_S) \), or, with \( t_L - t_S \) and \( x_S \) held constant, equal differences between \( \log(\delta) \) correspond to equal differences between \( \log(x_L) \).

10Most models that account for the common difference effect, including the three models under investigation, imply that \( \delta \) will be higher for \( M \) than for \( U \) even if \( E \) is as long as \( L \), because \( \delta \) increases by a greater proportion from \( E \) to \( M \) than from \( M \) to \( L \), i.e., \( \delta_M / \delta_E > \delta_L / \delta_M \) or \( \delta_M > \sqrt{\delta_E / \delta_L} \).

11The ratio rule actually requires that \( G \) be a power function of the form \( G(V) = iV^j \) if \( V \geq 0 \) and \( G(V) = i(-V)^j \) if \( V < 0 \), where \( i, j > 0 \). The identity function arises when \( i = j = 1 \).