A General Mathematical Theory of Discounting

Robert   Axtell
Gregory   McRae

SANTA FE INSTITUTE

A GENERAL MATHEMATICAL THEORY
OF
BEHAVIORAL AND AGGREGATE DISCOUNTING [*]

Robert L. Axtell
Center for Social Complexity
Krasnow Institute for Advanced Study
George Mason University
Fairfax, VA 22030 USA

Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501 USA

Center for Social and Economic Dynamics
The Brookings Institution
1775 Massachusetts Avenue, NW
Washington, DC 20036 USA

Gregory J. McRae
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Abstract
A mathematical formalism is developed for the existence of unique invariants associated with wide classes of observed discounting behavior. These invariants are 'exponential discount rate spectra,' derived from the theory of completely monotone functions. Exponential discounting, the empirically important case of hyperbolic discounting, and so-called sub-additive discounting are each special cases of the general theory. This formalism is interpreted at both the individual and social levels. Almost every discount rate spectrum yields a discount function that is 'hyperbolic' with respect to some exponential. Such hyperbolic discount functions may not be integrable, and the implications of non-integrability for intertemporal valuation are assessed. In general, non-stationary spectra lead to discount functions that are not completely monotone. The same is true of discount rate spectra that are not proper measures. This formalism unifies theories of non-constant discounting, declining discount rates, 'gamma' discounting, and related notions.

Keywords: hyperbolic discounting, completely monotone discount function, sub-additive discounting, Bernstein's theorem, discount rate spectrum, time consistency, preference reversal, negative probability.

I   Introduction: Two Distinct Motivations

A growing body of empirical research has demonstrated that the conventional present value calculus is unsatisfactory as a description of how humans value the future [e.g., Frederick, Loewenstein et al. 2002]. The process of multiplying time-labeled monetary valuations by exponential discount factors and then summing from the present time to infinity—the normal formalism—amounts to making very specific assumptions about the character of impatience, the time separability and consistency of preferences, and the constancy of discount rates. Such assumptions, while theoretically useful are, unfortunately, unsupported by empirical evidence [Thaler 1981; Herrnstein and Mazur 1987]. Instead of discounting exponentially, people seem to discount the short run *more* than exponentially and the long run *less*. This so-called *hyperbolic discounting* has received significant attention in the experimental and behavioral economics literatures [Chung and Herrnstein 1961; Loewenstein and Thaler 1989; Ainslie 1992; Loewenstein and Prelec 1992; Loewenstein, Read et al. 2003].

Exponential discounting is also unsatisfactory as a description of *aggregate* behavior when the economic aggregate has components being discounted at heterogeneous exponential rates. For example, consider two individuals discounting the same good at different rates. Because the sum of two exponentials is not generally an exponential function, the combined discounting of the two will not be representable by an exponential function. Conventional practice is to average the two discount rates, then characterize the aggregate by exponential discounting at this average rate [e.g., Lave 1981], but this is merely a heuristic device, not a precise mathematical description of the actual discounting.

Below we review these two distinct motivations for non-exponential discounting. We then go on to propose an alternative formalism that subsumes exponential and hyperbolic discounting as special cases.

A. Behavioral Discounting

Experiments with a variety of subjects have clearly demonstrated that the way animals, including humans, value the future deviates significantly from the conventional constant exponential discounting that is widely used in a variety of economic and financial contexts . For example, Ainslie [1992] and co-workers [Ainslie and Herrnstein 1981] have demonstrated in experiments with pigeons that they discount the near term considerably more than exponentially, while discounting the distant future considerably less. This hyperbolic discounting behavior has been corroborated in other experiments on non-humans.

However, most of the experimental results on valuation over time have been performed on humans. Experimental results involve both monetary [e.g., Thaler 1981; Loewenstein 1987; Horowitz 1991; Green, Myerson et al. 1997; Harrison, Lau et al. 2002] and non-monetary

rewards [e.g., Horowitz and Carson 1990; Cairns and van der Pol 1999; van der Pol and Cairns 1999; van der Pol and Cairns 2001]. From this empirical work there have been attempts to develop a theoretical basis for choice behavior over time that relaxes the usual assumptions of preference separability, constant discounting, etc. [Harvey 1986; Loewenstein and Prelec 1992]. These efforts leave little doubt that people do not value the future in accord with simple compound interest.

Most recently, various brain imaging experiments in decision-making environments [Glimcher 2003] have demonstrated that different parts of the brain are active when short term and long run decisions are being made. This leaves little doubt that the underlying mechanisms responsible for discounting behavior are unlikely to be simple.

B. *Heterogeneous Discount Rates at the Social Level*

At any instant in time, a great variety of discount rates are at work in any real economy. A single household might have loans at different interest rates for their home mortgage, automobile, line of credit, and credit cards. They may have investments that are appreciating at fixed rates, like certificates of deposit and bonds, in addition to equities having market-determined returns. Over an entire economy there will be a fairly broad distribution of such rates, with an average and standard deviation.

Or consider a particular intertemporal valuation issue across households and individuals. Do people weight the present versus the future in the same way? Essentially all survey data of this type gives evidence of robust heterogeneity [Hausman 1979; e.g., Horowitz and Carson 1990; Moore and Viscusi 1990; Holden, Shiferaw et al. 1998]. Even among economists there is little consensus on how to discount the future, as in Weitzman's survey concerning the appropriate discount rate to use for cost-benefit analysis [Weitzman 2001].

Does the existence of an average discount rate mean that credit and investment activities in an economy are mathematically well-represented by the exponential discount function that corresponds to this average rate? Or is it instead the case that a combination of exponential discount functions is not well approximated by an exponential? This is the question we resolve in this paper.

In particular, with respect to agent heterogeneity, we will demonstrate that a population of exponential discounters is generally *not* well-represented by exponential discounting overall. Instead, a non-exponential discount function results from aggregation, an aggregate discount function whose explicit functional form depends on the distribution of exponential rates present in the population.

In the next section a mathematical formalism is described for a *generalized discount function* having exponential and hyperbolic

functions as special cases.[1] The subsequent section develops a methodology for rationalizing observed discounting behavior in terms of this formalism. Section IV gives applications of the generalized discount function, and section V extends the formalism in a variety of ways.

II Discount Functions Generated by Discount Rate Spectra

Let us take, for the moment, the notion that people discount the immediate future more than exponential and the distant future less than exponential. Call this 'weak' hyperbolic discounting, as opposed to following an exact hyperbolic function, what might be called 'strong' hyperbolicity. Now note that a simple combination of distinct exponential discount functions is 'weakly hyperbolic' with respect to an exponential function having an interior rate. That is, for $r_1 < r_2 < r_3$, $[\exp(-r_1 t) + \exp(-r_3 t)]/2$ is initially less than $\exp(-r_2 t)$, and then sometime later the two curves intersect, while for long times the relationship is reversed.

In lieu of such a composite discount function, it is more common to simply average r_1 and r_3—call it $\bar{r}$—and then perform exponential discounting at rate $\bar{r}$. Examples of the use of this approximation in economic theory and practice are manifold, from *macroeconomics* wherein 'social welfare functions' are discounted at the 'social discount rate' (see any standard text, e.g. Blanchard and Fischer [1989] or Sargent [1987]) to problems of *environmental economics* in which the 'contingent value' of public resources is ascertained by averaging individual 'willingness-to-pay' valuations and then discounting exponentially (cf. the special issue of the *Journal of Environmental Economics and Management* on the 'The Social Discount Rate' edited by Howe [1990], even to practical problems involving the use of discount rates in governmental *cost-benefit project analyses* [Lyons 1990]. In all of these areas the main empirical question regarding this approximation appears to involve finding the 'right' effective discount factor, $\bar{r}$, to use.[2]

To make somewhat more formal this comparison of an exponential average rate with an average of exponential rate, suppose that the economic entity engaged in discounting has several distinct discount rates it uses in order to value economic streams that run into the future. Specifically, write the discount function that obtains in such a case as

(1)
$$d(t) = \sum_{i=1}^{n} \sigma_i \exp(-r_i t),$$

---

[1] Independently, Jamison and Jamison [2003] have explored many of the same issues. They do not base their inquiry on discount rate spectra underlying discount functions, but rather on the properties of discount functions themselves. Ok and Masatlioglu [forthcoming] have similar objectives to ours—to provide a mathematical foundation for intertemporal valuation—but take an axiomatic approach through preferences.

[2] For now we treat r as constant. Time-varying discount rates [cf. Blanchard and Fischer 1989] are the subject of section V.

where the σ_i provide the appropriate (normalized) weighting, i.e., $\sum_{i=1}^{n} \sigma_i = 1$. In the limit of a large number of exponential rates, this expression is

$$(2) \qquad\qquad d(t) \approx \int_0^\infty S(r)\exp(-rt)dr$$

where $S(r)$ provides the weights, i.e., $\int_0^\infty S(r)dr = 1$. We call this function, *S(r)*, the (exponential) discount rate spectrum.[3] Note that for

$$S(r) = \sum_{i=1}^{n} \sigma_i \delta(r - r_i),$$

where $\delta(\cdot)$ represents the Dirac delta function, (2) becomes (1). In the remainder of this section we shall develop a formal rationale for representing discounting functions via (2).

A natural interpretation of this discount rate spectrum is via so-called *multiple selves*, in which a single individual is conceived of as having both myopic and farsighted interpersonal selves, with various ways for mediating between them [e.g., Schelling 1984; Ainslie and Haslam 1992]. As opposed to having just a few such 'selves,' the discount rate spectrum approach we adopt here facilitates an interpretation in which a single individual has a potentially large number of selves, perhaps even a continuum of such selves. In what follows we will be less concerned with such interpretations and more concerned with the economic meaning of the mathematical formalism.

A. *Necessary Conditions for a Discount Function*

In order for a function to be a discount function it must have certain properties, which we treat as axioms. These are simply:

Axiom 1: $d(0) = 1$, that is, no discounting of the present;
Axiom 2: $d(t)$ must be strictly monotone decreasing, $d'(t) < 0$, i.e., the value at any two distinct future times is larger at the nearer time.

Functions having these properties are many, of course. It turns out that conventional discount functions typically possess one additional property, *complete monotonicity*.

B. *Completely Monotone Discount Functions*

A function is said to be *completely monotone* (*CM*) when its derivatives alternate in sign [cf., Feller 1971: 439]. It is well known that:

Proposition 1: *Exponential discount functions are completely monotone.*

---

[3] An alternative and equivalent formulation of (2) reads $d(t) = \int_0^1 R(\beta)\beta^{\,t}d\beta$, with $\beta = \exp(-r)$ and $R(\beta) = S(ln(1/\beta))$.

<u>Proof</u>: The nth derivative of the exponential function is $d^{(n)}(t) = (-1)^n r^n \exp(-rt) = (-1)^n r^n d(t)$, which clearly alternates in sign. ❏

It is not difficult to see that a linear combination of *CM* functions is *CM*, meaning that (1), for instance, is *CM*. Next we show that hyperbolic discount functions have this same property:

Proposition 2: Hyperbolic discount functions are completely monotone.

<u>Proof</u>: Noting the discount function by $p/(p+t)^q$,

$$d'(t) = \frac{-pq}{\left(p + t\right)^{q+1}}, \quad d''(t) = \frac{pq(q + 1)}{\left(p + t\right)^{q+2}},$$

with the general case given by $d^{(n)}(t) = (-1)^n \dfrac{p\prod\limits_{i=0}^{n-1}(q + i)}{\left(p + t\right)^{q+n}}.$ ❏

Because these common discount functions have the *CM* property, we replace the second axiom above with the following:

Axiom 2': $d(t)$ is completely monotone and decreasing.
This sufficiently restricts the class of admissible discount functions such that we can now completely characterize this class, and do so in a way commensurate with the heuristic development that led to expression (2).

C. *Existence of Discount Rate Spectra*

Each *CM* function has associated with it a unique *spectrum,* which *generates* the function through an integral transformation:

Theorem 1 (Bernstein [1928]): *Every completely monotone discount function, d(t), can be represented as* $d(t) = \int_0^\infty S(r)\exp(-rt)dr$ *where S(r) is a probability density function having support on a subset of* $[0,\infty)$[4].

<u>Proof</u>: Feller [1971: 439-440] ❏

Notice that this representation is the same as (2). $S(r)$ acts as a kind of weighting function over the domain of all feasible discount rates. Stated differently, $d(t)$ has all discount rates 'in it' in varying degrees according to $S(r)$. The function $d(t)$ specifies a *generalized discount function.*[5]

---

[4] Negative discount rates could be admitted to the definition of $d(t)$—e.g., Ayres and Mori [1989] discuss negative discount rates in the context of a product life-cycle model—however, such rates invalidate the Laplace transform interpretation. See § 5.C below for more on this topic.

[5] Note that it is *not* a universal discount function since complete mono-tonicity (axiom 2') is a stronger condition than monotonicity (axiom 2).

D. Determination of Discount Functions

Given a discount rate spectrum, there are a variety of ways to determine the discount function, in addition to integration as in (2).

D.1. Discount functions as Laplace transforms

A function $g: \mathbf{R}_+ \mapsto \mathbf{R}$ has a Laplace transform, $G: \mathbf{R}_+ \mapsto \mathbf{R}$, defined[6]

$$L\big[g(x)\big] \equiv G(s) \equiv \int_0^\infty g(x)\exp(-sx)dx.$$

From (2) it can be seen that $d(t)$ is just the Laplace transform of the discount spectrum in the (real) transform variable t.[7] With this knowledge it is an easy matter to determine the generalized discount function, $d(t)$, that corresponds to a particular spectrum, $S(r)$, for so-called Laplace transform pairs are tabulated in a great many places [Doetsch 1961; Abramowitz and Stegun 1964; Doetsch 1970; LePage 1980; Wolfram 1991]. The only caveat to note regarding employment of such tables is that $S(r)$ must be a probability density function (*pdf*), defined on $r \geq 0$. We present a short table in the Appendix (table *A*).

D.2. Discount functions via characteristic functions

Calculating a particular $d(t)$ from a specified $S(r)$ by integration—as one needs to do if available Laplace transform tables are inadequate—can be a tedious undertaking. However, there is a rather more succinct way to obtain $d(t)$, through the characteristic function, Φ, of the discount rate spectrum $S(r)$. The characteristic function (*chf*) of a measure is its Fourier transform [Billingsley 1986], i.e., for a random variable X,

$$\Phi_X(s) \equiv E\big[\exp(isX)\big] = \int_{-\infty}^\infty \exp(isx)f_X(x)dx$$

where $i = \sqrt{-1}$ and $f(\cdot)$ is the *pdf* of X. Since the Laplace transform can be thought of as half a Fourier transform, or as a Fourier transform of a function defined only on $\mathbf{R}_+$, it can be seen that $d(t)$ is just the characteristic function of $S(r)$ in the variable $s = it$, with the integration performed over $[0,\infty)$. Thus, if the *chf* is available for a particular spectrum it is an easy matter to obtain the associated discount function.

Example 1: Characteristic function used to obtain the generalized discount function from a specified discount rate spectrum
For a Pearson Type III spectrum [Abramowitz and Stegun 1964], i.e.,

$$S(r) = \frac{1}{\beta\Gamma(p)}\left(\frac{r-\alpha}{\beta}\right)^{p-1}\exp\left(-\frac{r-\alpha}{\beta}\right)$$

---

[6] It is conventional to treat *s* as a complex variable but much of the theory of Laplace transforms goes through with *s* real-valued [Kreider, Kuller et al. 1966], which is closer to our purposes.

for $r > \alpha$, $\alpha \in \boldsymbol{R}$, and $\beta, p \in \boldsymbol{R}_{++}$, one simply looks up the *chf* as $\Phi(s) = \exp(i\alpha s) [1 - i\beta s]^{-p}$. Therefore, $d(t) = \Phi(it) = \exp(-\alpha t) [1 + \beta t]^{-p}$. Note that this is a combination of exponential and hyperbolic discounting.

D.3. Discount functions via moment-generating functions

Given the close relation between the characteristic function of a probability measure and its moment-generating function (*mgf*), it will come as no surprise that it is easy to determine the corresponding discount function from a measure's *mgf*. Since

$$M_X(s) \equiv E\big[\exp(sX)\big] = \int_{-\infty}^{\infty} \exp(sx) f_X(x) dx$$

we have that $d(t) = M(-t)$ for whatever discount rate spectrum is under consideration. Interestingly, the moments of the discount rate spectrum are readily available in this formalism as $-d'(0)$, $d''(0)$, and so on.

In the next section we investigate the problem of determining discount spectra from generalized discount functions. It is the inverse problem of that which we have analyzed above.

III   Determination of Discounting Spectra from Behavior

Given empirical data on discounting behavior, it is possible to reconstruct a discount rate spectrum. That is, an economic actor *reveals* its (unique) discount spectrum through its valuation over time. Formally, given $d(t)$ it is possible to obtain $S(r)$ from (2) by solving the inverse problem where the unknown function appears under the integral. This is a linear Fredholm integral equation of the first kind. In general, Fredholm equations of this type are notoriously difficult to solve: solutions may not exist, if they exist they may depend sensitively on parameters or not be unique, and so on. However, in the present case no such difficulties are encountered due to the monotone nature of the kernel—*exp(-rt)*. In particular, we have the following uniqueness result:

Proposition 3: *Distinct discount spectra have distinct discount functions.*

Proof: Feller [1971: 430].   ❏

What is more, an explicit inversion formula exists which is applicable to differentiable $d(t)$ functions.[8] In particular it can be shown [Feller 1971: 440]

$$\lim_{n \to \infty} \frac{(-1)^{n-1}}{(n-1)!}\left(\frac{n}{r}\right)^n d^{(n-1)}\left(\frac{n}{r}\right) = S(r).$$

---

[7] In discrete time one must use the *z*-transform [Cadzow 1973; Luenberger 1979] and $S(r)$ becomes a probability mass function.

Example 2: Use of the inversion formula

From table *A* we find that the exponential discount rate distribution in parameter p possesses generalized discount function $p/(p+t)$.[9] The kth derivative of $d(t)$ in this case is given by

$$\frac{(-1)^k k!\, p}{(p+t)^{k+1}}.$$

Inserting this into the inversion formula and taking the limit we obtain

$$S(r) = \lim_{n\to\infty} \frac{p}{\left(p+\dfrac{n}{r}\right)^n}\left(\frac{n}{r}\right)^n = \lim_{n\to\infty} p\left(1+\frac{pr}{n}\right)^{-n}$$

The goal is to show that this last expression is in fact $p\, exp(-pr)$. Clearly the prefactor p is right and the question that remains is whether the second term equals the negative exponential function in the limit. The traditional representation of the negative exponential is as follows

$$exp(-pr) = \lim_{n\to\infty}\left(1-\frac{pr}{n}\right)^{-n}.$$

But it is not difficult to see that this expression is the same as the last term of the previous one in the limit of large n—they just approach the limiting value from opposite directions. Thus we have shown what we started out to prove.[10]

Having demonstrated how to determine $d(t)$ from $S(r)$ and vice-versa, we next show that exponential discounting is a *very* special case.

A. Conditions for Exponential Discounting

It is easily inferred from all that has been said above that setting $d(t) = exp(-r_0 t)$, i.e., exponential discounting, is a very special case of the generalized discounting formalism.[11] Intuitively it would seem that discounting an aggregate value stream with this functional form would be valid only when all component value streams have the same discount rate, that is $r_i = r_0$ for all i. This intuition proves correct:

---

[8] The standard Laplace transform inversion formula will also produce $S(r)$ from $d(t)$. However, economists may not find this convenient since it involves contour integration, and because $d(t)$ is a function of a real variable.

[9] Harvey [1995] calls this proportional discounting; see also Kenley and Armstead [2003].

[10] The utility of the inversion formula for other discount functions depends on one's ability to recognize elementary functions from their representation as limits.

[11] The ubiquity of this discount formulation in economic theory is commonly ascribed to an influential paper of Samuelson [1937].

Proposition 4: *A discount rate spectrum having Dirac measure is necessary and sufficient for exponential discounting.*

<u>Proof</u>: Call $S_E(r; r_0)$ the discount spectrum that produces $d(t) = \exp(-r_0 t)$, It is obtained by solving the following equation:

$$\exp(-r_0 t) = \int_0^\infty S_E(r; r_0) \exp(-rt) dr.$$

To do this, consider the last entry of table *A*. Notice that the Dirac spectrum of $\delta(r-r_0)$ corresponds to $d(t) = \exp(-r_0 t)$, which proves sufficiency. Necessity then follows from proposition 3.   ❏

Summarizing, the *only* way to produce exponential discounting is with a discount rate spectrum having all its mass on the exponential rate in question. Any heterogeneity in the spectrum whatsoever leads to a departure from the exponential form.

B. Conditions for Hyperbolic Discounting

A version of this problem was solved as example 2: an exponential distribution of discount rates yields pure hyperbolic discounting. Mathematically, this result—that an exponential distribution having its parameter exponentially or Γ-Erlang distributed gives the Pareto distribution (hyperbolic law)—has been known for some time [Maguire, Pearson et al. 1952; Harris 1968]. It has been rediscovered several times in discounting contexts:

Proposition 5 [Axtell 1992; Sozou 1998; Weitzman 2001]: *A discount rate spectrum having Γ-Erlang measure is necessary and sufficient for hyperbolic discounting.*

<u>Proof</u>: Let $S(r) = p^q r^{q-1} \exp(-pr)/\Gamma(q)$ for $p, q > 0$. Performing the integration in equation (2) leads to $d(t) = p^q / (p + t)^q$.[12] Necessity follows directly from proposition 3.   ❏

Note that this result is shown on the third row of table *A*. Weitzman [2001] has termed this 'gamma discounting.'

Example 3: Hyperbolic discounting in the neoclassical growth model

In his extension of the neoclassical growth model to hyperbolic discounting, Barro [1999] employs the following functional form,

---

[12] This form of hyperbolic discounting is equivalent to, for example, Loewenstein and Prelec [1992], who write it as $1/(1 + \alpha t)^{\beta/\alpha}$; substituting $p = 1/\alpha$ and $q = \beta/\alpha$ produces their formulation from mine, although yields a non-standard representation of the underlying Γ-Erlang distribution. Furthermore, since the mean of this distribution, μ, is q/p and the variance, σ^2 is q/p^2, the hyperbolic discount function can be written in terms of these parameters as $d(t) = 1/(1 + \sigma^2 t/\mu)^{\mu^2/\sigma^2}$, as Weitzman [2001] does.

$$d_B(t) = \exp\left[-\rho t - \frac{b}{\gamma}\left(1 - \exp(-\gamma t)\right)\right],$$

where b, γ, and ρ are positive constants. The initial discount rate is ρ and it declines over time at rate γ. Elementary albeit tedious calculations reveal this function to be *CM*: $d_B'(t) = -\left(\rho + b\exp(-\gamma t)\right)d_B(t)$,

$$d_B''(t) = \left[\left(\rho + b\exp(-\gamma t)\right)^2 + b\gamma\exp(-\gamma t)\right]d_B(t),$$

$$d_B'''(t) = -\left\{\left[\left(\rho + b\exp(-\gamma t)\right)^2 + 3b\gamma\exp(-\gamma t)\right]\left(\rho + b\exp(-\gamma t)\right)\right\}d_B(t),$$

and so on. Therefore, it must be the case that there exists a discount rate spectrum, $S_B(r)$, that generates Barro's discount function; that is, $L[S_B(r)] = d_B(t)$. We want to solve this for $S_B(r)$ to see what spectrum underlies $d_B(t)$. To accomplish this we will use table *A*. First note that

$$L\left[S_B(r + \rho)\right] = \exp(\rho t)L\left[S_B(r)\right] = \exp\left[-\frac{b}{\gamma}\left(1 - \exp(-\gamma t)\right)\right].$$

Furthermore,

$$L\left[\exp\left(\frac{b}{\gamma}\right)S_B(r + \rho)\right] = \exp\left(\frac{b}{\gamma}\right)L\left[S_B(r + \rho)\right] = \exp\left(\frac{b}{\gamma}\exp(-\gamma t)\right).$$

Next, expand the RHS to read

$$L\left[\exp\left(\frac{b}{\gamma}\right)S_B(r + \rho)\right] = \sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{b}{\gamma}\exp(-\gamma t)\right)^j = \sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{b}{\gamma}\right)^j\exp(-j\gamma t).$$

From the earlier discussion of exponential discounting it is easy to see that the spectrum that generates the RHS function above is

$$\sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{b}{\gamma}\right)^j\delta(r - j\gamma),$$

a weighted sum of Dirac measures at equally-spaced intervals. Therefore, the spectrum underlying Barro's version of hyperbolic discounting amounts to $S_B(r) = \sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{b}{\gamma}\right)^{j-1}\delta(r - \rho - j\gamma)$ an infinite sum of Dirac measures at equal intervals of size γ—a so-called Dirac comb—beginning at $r = \rho$, with declining weights in b/γ.

C. Sub-Additive Discounting

So-called 'sub-additive' discounting has been suggested as an alternative to hyperbolic discounting [Read 2001; Scholten and Read 2006]. A variety of functional forms for sub-additive discount functions have been proposed, including $d(t) = \exp(-pt^s)$ with $0 < s < 1$. In this section we investigate this class of discount functions.

Proposition 6: *The discount function $d(t) = \exp(-pt^\alpha)$ with $\alpha \in (0, 1)$ is completely monotone.*

<u>Proof</u>: Calculations that are elementary but (again) tedious reveal:

$$d'(t) = -\alpha p t^{\alpha-1} d(t),$$

$$d''(t) = \alpha p t^{\alpha-2}\left[1 - \alpha\left(1 - pt^\alpha\right)\right]d(t),$$

and

$$d'''(t) = -\alpha p t^{\alpha-3}\left[2 - 3\alpha\left(1 - pt^\alpha\right) + \alpha^2\left(1 - 3pt^\alpha + p^2 t^{2\alpha}\right)\right]d(t).$$

It is difficult to write out the general case but not difficult to convince oneself that this discount function is indeed *CM*. ❏

What discount rate spectrum underlies this discount function? Perhaps surprisingly, there is a straightforward answer to this question:

Proposition 7: *A stable discount rate distribution, $S_\alpha(r)$, is necessary and sufficient to generate the discount function $d(t) = \exp(-pt^\alpha)$ with $\alpha \in (0, 1)$.*

<u>Proof</u>: Feller [1971: 448-9]

A *stable* distribution is one having the property that if independent random variables have that distribution then their sum, appropriately normalized, will also have that distribution [Samorodnitsky and Taqqu 1994]. For example, the normal and Cauchy are two well-known stable distributions, for $\alpha = 2$ and 1, respectively. However, they do not underlie the sub-additive discount function $d(t) = \exp(-pt^\alpha)$ due to the restriction of α to (0, 1). There is only one stable distribution in the feasible range whose *pdf* is known:

Example 4: The Lévy distribution of exponential discount rates

Consider the following discount rate spectrum, known as a Lévy distribution:

$$S(r) = \frac{p}{2\sqrt{\pi r^3}}\exp\left(-\frac{p^2}{4r}\right).$$

This is the stable distribution having $\alpha = 1/2$. Its moments do not exist. Inserting this into equation (2) and integrating reveals that it corresponds to the sub-additive discount function $d(t) = \exp\left(-p\sqrt{t}\right)$.

The Lévy spectrum is plotted in figure 1a for various values of p (and is the eighth entry in table *A*). In figure 1b the sub-additive discount function is compared with exponential discounting.
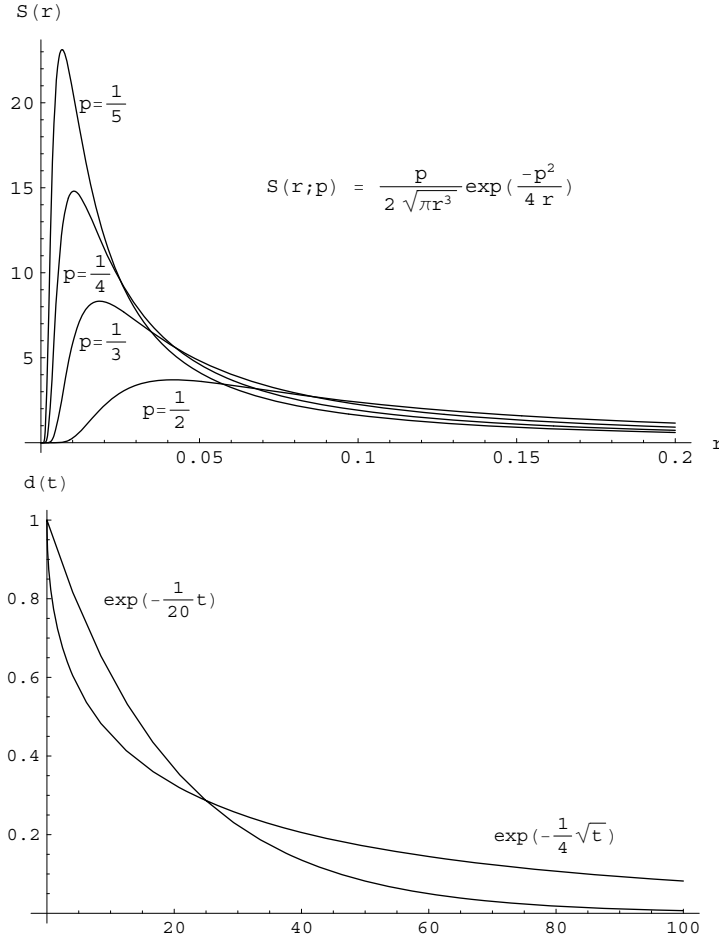
$$S(r;p) \;=\; \frac{p}{2\sqrt{\pi r^3}} \exp\left(\frac{-p^2}{4\,r}\right)$$

Figure 1, *a* and *b*: The discount rate spectrum that corresponds to sub-additive exponential discounting (top plot) and the resulting discount function compared to a conventional exponential discount function (bottom plot); the cross-over point is *t* = 25.

So far our main results have been analytical. We now investigate how to use empirical data to determine underlying discount rate spectra.

D. Computation of Discount Spectra from Valuation Data

The results of experiments with human subjects are not usually analytically-precise data. Rather, intertemporal valuation experiments typically generate noisy data at discrete times. How might one go about inferring an underlying discount rate spectrum from such data?

It turns out that there are well-known and reasonably robust procedures for computing the inverses of Laplace transforms from numerical data. These are particularly efficient when the spectrum can be assumed to be a probability measure [Abate and Whitt 1995], as in present circumstances. In fact, specialized algorithms are available when *CM* functions are present [Abate and Whitt 1999].

E. Generic 'Weak Hyberbolicity' of Discount Functions

Hyperbolic discounting has two distinct interpretations. There is

the literal interpretation, in which the discount function follows a hyperbolic function exactly. Then there is a more qualitative interpretation, corresponding to any discount function that gives more discounting than some exponential amount in the near term followed by less discounting than exponential in the long term. Let us call these different versions of hyperbolic discounting 'strong' and 'weak' hyperbolicity, respectively. So far we have been essentially concerned with 'strong hyperbolicity', focusing on explicitly hyperbolic functional forms. While this has been mathematically tractable, it is unlikely that human subjects, in departing from exponential discounting, will behave in ways that are literally hyperbolic. We now turn our attention to 'weakly hyperbolic' discounting:

Proposition 8: *Almost all discount rate spectra yield weakly hyperbolic discount functions.*

<u>Proof</u>: Consider a nondegenerate discount spectrum, *S(r)*, having support on the interval $[r_1, r_2]$, with $0 \leq r_1 < r_2 < \infty$. Pick some r_0 in this interval and form the discount function, $exp(-r_0 t)$. The mass of *S(r)* for $r < r_0$ corresponds to less discounting of the distant future than $exp(-r_0 t)$, while *S(r)* for $r > r_0$ corresponds to more discounting of the near term (see fig 2).
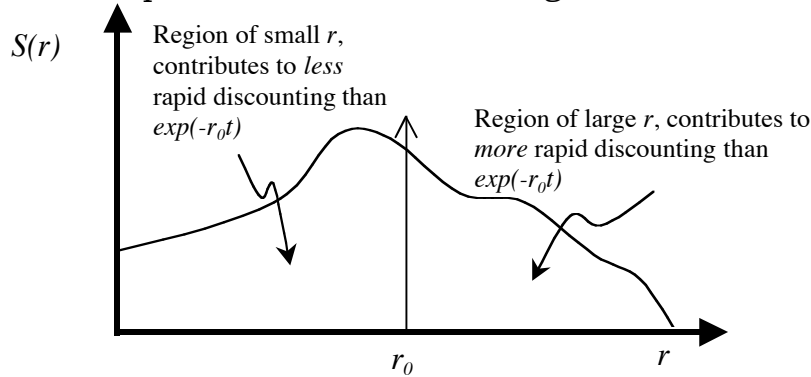

Figure 2: For any nondegenerate (non-Dirac) spectrum, *S(r)*, the corresponding discount function is 'weakly hyperbolic' with respect to some exponential discount function, $exp(-r_0 t)$; the mass of *S(r)* for $r < r_0$ corresponds to less discounting of the distant future than $exp(-r_0 t)$, while for $r > r_0$ there is more discounting of the near term.

Thus, the discount function generated by *S(r)* will be weakly hyperbolic with respect to $exp(-r_0 t)$. Furthermore, since the set of degenerate spectra has measure zero in the space of all probability measures, the discount function formed by an arbitrary discount rate spectrum is almost surely weakly hyperbolic with respect to some exponential discount function.❑

The two facets of 'weak hyperbolicity'—more discounting than exponential in the near term, less than exponential in the long term—is thus a generic property of completely monotone functions.[13]

---

[13] That only the first of these is a property of 'quasi-hyperbolic' discounting [Phelps and Pollak 1968; Laibson 1997] is problematic. For instance, the integrability properties of discount functions, investigated in § IV.B below, do not come into play for quasi-hyperbolic discounting.

IV   Some Properties of Generalized Discount Functions

In this section we describe two important issues associated with generalized discount functions, time (in)consistency and integrability.

A. Time Inconsistency and Preference Reversals

An important issue arising in inter-temporal valuation is the *time consistency* of optimal paths. Strotz [1956] investigated this in the context of recursive, time-additive preferences. Time consistency requires that economic decision-making be performed such that future recalculations of optimal policies will not generate different paths from the one decided upon at the start. Strotz proved that the only discount functions that are time consistent are exponential ones. Since then there has emerged a large literature on time consistency and preference reversals [e.g., Marschak 1964; Pollak 1968; Ainslie and Herrnstein 1981; Thaler 1981; Mischel 1984; Loewenstein and Thaler 1989; Horowitz 1992; Green, Fristoe et al. 1994; Kirby and Herrnstein 1995].

A.1. Micro Consistency Does Not Imply Macro Consistency

Consider an economic aggregate composed of many individuals, each of whom discounts exponentially. At the 'micro-level' of individuals everyone behaves in a time consistent fashion. However, non-exponential generalized discount functions obtain for the aggregate for nearly any distribution of micro-level exponential rates. These non-exponential aggregate discount functions are *not* time consistent paths. Thus we have something of a paradox: microeconomic time consistency, macro-economic time inconsistency. The only way in which microeconomic time consistency can result in macroeconomic consistency is when the components of the aggregate have identical discount behavior.

A.2. Macro Consistency Implies Micro Time Consistency

By macro-time consistency we mean exponential discounting in the aggregate. As has been shown above, this is possible iff all micro-level components are time consistent *and* have identical discount rates.[14]

B. Integrability

Exponential and hyperbolic discount functions are *CM*. However, in contrast to the exponential form, hyperbolic discount functions may not be elements of $L^1(0,\infty)$, the space of functions having finite norm over

---

[14] Horowitz [1991; 1992] has reported what is apparently the *exact opposite* result from a series of economic experiments. In particular, in the context of a portfolio choice experiment he observed inconsistent behavior among many agents but approximate time consistency in the aggregate. However, discounting behavior is not present in these experiments; his work is a test of *intertemporal stationarity* instead of time consistency.

[0,∞).[15] That is, generalized discount functions may not be integrable.

Basically, those functions having infinite L^1 norm do not approach zero 'fast enough' as time increases, this despite the fact that they are *CM*. The economic implication of this slow decrease is there is no point beyond which it is possible to neglect the future. Stated differently, for discount functions not having this norm one must explicitly select a time beyond which economic considerations are neglected.

B.1. Infinite Time Horizon

We want to establish the properties that $S(r)$ must possess in order for $d(t)$ to be integrable.[16] We have the following characterization of integrable discount functions:

Theorem 2: *Discount spectrum $S(r) \Rightarrow d(t) \in L^1(0,\infty)$ iff $\int_0^\infty S(r)\dfrac{dr}{r} \in L^1(0,\infty)$.*

Proof: Begin by writing the discount function in terms of the discount rate spectrum and compute the L^1 norm

$$\left|d(t)\right| \equiv \int_0^\infty \int_0^\infty S(r)\exp(-rt)drdt$$

Rearrange the order of integration to yield

$$\left|d(t)\right| = \int_0^\infty S(r)\int_0^\infty \exp(-rt)dtdr$$

Performing the integration yields $\left|d(t)\right| = \int_0^\infty S(r)\dfrac{dr}{r}$ ❏

Clearly, a necessary condition for the discount function to be integrable is that $S(r) \to 0$ as $r \to 0$. In words, every discount rate distribution that evaluates to a positive number at $r = 0$ will produce a generalized discount function which is not in L^1. Many of the distributions listed in table *A*, for instance, have $S(r) \neq 0$, at least for some set of parameters.

There are at least two important implications of this theorem. First, in the case of a public good, if a single individual believes that the good should not be discounted then any aggregation that takes account of this person's valuation must be not integrable. That is, the distribution of discount rates, *S(r)*, will have finite mass at $r = 0$ thus, no matter what distribution of discount rates other people report, it is as if everyone demands $r = 0$. The person with $r = 0$ is a kind of dictator.[17]

A second implication of the theorem is a formalization of the notion

---

[15] $L^1(0,\infty)$ is usually abbreviated simply as L^1 in what follows.

[16] In the mathematics literature, results establishing asymptotic properties of *S(r)* from those of its Laplace transform—*d(t)* herein—are called *Tauberian*. Alternatively, *Abelian* theorems describe the behavior of the transform in terms of properties of the spectrum.

[17] For example, in Weitzman's [2001] survey of some 2160 economists as to what discount rate should be applied to combat global warming, several respondents said 0% (and a few even gave negative discount rates!). His fit of a gamma distribution to the empirical data neglects this lower end of the reported discount rates, thus he obtains an integrable discount function despite the fact that the data do not support this.

of *existence values* for environmental and natural resources. An important topic in current discussions of biodiversity issues associated with global change, existence value refers to the intrinsic value of maintaining a resource in its unexploited state [Page 1977]. It is not an economic value and it is not even clear that it is monetizable in any well-defined way. It is most commonly invoked when an ecological object has no clear economic functionality—and cannot, therefore, be called a good—but is sufficiently appreciated, in some sense, by some fraction of the current generation. When an object has finite value but is not discounted ($r = 0$), the usual present value calculus must assign an infinite present value to the object, while if one integrates out only through T then the present value can be made arbitrarily large by selecting T large. It does not seem that such an enormous valuation could ever be overturned such that complete consumption or extinction of the object would be justified on economic grounds. Thus we propose that assigning a zero discount rate is nominally equivalent to saying that an object has existence value.

B.2. Finite Time Horizon

For generalized discount functions not in L^1 a more general definition of present value is needed, where the time horizon, T, is taken into account explicitly, i.e., $PV(T) = \int_0^T d(t)V(t)dt$. Here it is easy to see that PV can be an increasing function of T, i.e., $PV'(T) > 0$. Of course, for those functions that are in L^1 the two definitions coincide, either for all T beyond some finite one or in the limit as $T \to \infty$. To compare this new definition with the traditional one it is instructive to build a table for the *effective* time horizon, T, in the traditional method as a function of the fraction of total ($T \uparrow \infty$) present value achieved in time T, as well as of the discount rate. Representative calculations are given in table 1:

	Fraction of PV(∞)		
Discount Rate	**90%**	**99%**	**99.90%**
2%	115.13	230.26	345.39
4%	57.56	115.13	172.69
6%	38.38	76.75	115.13
8%	28.78	57.56	86.35
10%	23.03	46.05	69.08
20%	11.51	23.03	34.54

Table 1: Time for percent devaluation as a function of the discount rate

For example, at a rate of 8% a little more than 57 years are needed for extraction of all but one percent of the value stream. So depending on which fraction of *PV* one considers, employing an 8% discount rate is equivalent to having either a 29, 58 or 86 year time horizon.[18]

---

[18] Note that increasing the discount rate by a factor of 10 causes a decrease in the time horizon,

The present value function the requires explicit specification of *T* is very different from the usual one. Use of it in practice would generate very different results. But the methodology of its use is far from clear; e.g., *a priori* it is ambiguous how *T* would be specified for a given project.

In the specific case of a pure hyperbolic discount function, which is not integrable, a finite time horizon, *T*, must be specified if one is to make use of this functional form. In the case of the individuals behaving with effective hyperbolic discount functions perhaps a useful *T* is simply the expected lifetime of the individual.[19]

V   *Extensions: Beyond Complete Monotonicity*

Here we explore non-*CM* discount functions and their spectra, non-stationary spectra first followed by spectra that are improper measures.

A. *Non-Stationary Discount Rate Spectra*

In principle, there is no reason why the discount rate spectrum cannot vary with time, i.e., replace $S(r)$ with $S(r|t)$, where the distribution has been written as a conditional one. The defining relation for the generalized discount function is then $d(t) = \int_0^\infty S(r,t)\exp(-rt)dr$. If $d(t)$ is to be defined for all $t \in [0,\infty)$ then $S(r|t)$ must be a valid *pdf* at each *t*, including $t = 0$. Note that this formulation of generalized discounting includes time-varying discount rates [e.g., Epstein and Hynes 1983] as a special case. In this section we explore the effects of non-stationary discount rate spectra on generalized discount functions.

A.1. *Forward problem: Discount function from a spectrum*

It is straightforward to compute the generalized discount function from a non-stationary spectrum, as the following example illustrates.

Example 5: Non-stationary discount spectra from table *A*

Since for any feasible numerical value of the parameters *p* and *q* in table *A* the resulting spectra are well-defined *pdf*s, it is also true that one can substitute any positive function of *t* for the parameters and obtain a conditional *pdf*. (Note that it is not possible to merely substitute *t* for either *p* or *q* since these parameters cannot equal zero, while the discount function must be defined at $t = 0$.) For example, substituting γ/t for *p* in the exponential distribution there results

---

T, of the discount rate by a factor of 10, this due to the linearity of the present value integral. Also note that because of the way the fractional *PVs* were chosen, the last two columns are simply multiples of the second one, 2X and 3X, respectively.

[19] The relationship between the area under the exponential discount function and the *T* necessary to produce the same area under a hyperbolic discount function has been explored in Ayres and Axtell [1996], where the relationship between the parameters in these functions is also analyzed in detail.

$S(r|t) = \exp(-\gamma r/t)/\gamma t$. Since the mean and variance of the exponential distribution in table *A* are $1/p$ and $1/p^2$, respectively, these same moments of $S(r|t)$ are t/γ and t^2/γ^2. This distribution starts out as something like a Dirac measure centered at $r = 0$ and then shifts over time in the direction of positive r with its mean moving away from 0 such that, in the limit of $t \rightarrow \infty$ its density is completely dispersed along the r-axis. This is displayed in figure 3 for $\gamma = 1$, with *S* on the vertical axis, *t* increasing to the left and *r* increasing out of the page.
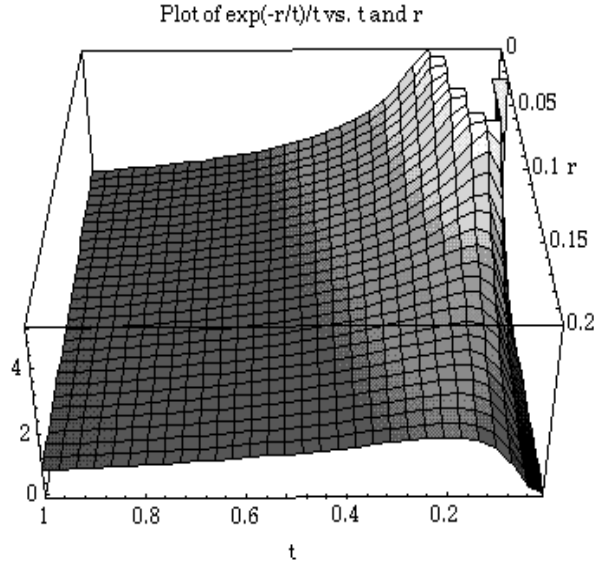

Figure 3: Example of a non-stationary discount rate spectrum

Carrying out the Laplace transform to reveal the generalized discount function is particularly easy in this case, but it is even easier to use table *A*, substituting for *p* in the expression for $d(t)$ to obtain

$$d(t) = \frac{\frac{\gamma}{t}}{\frac{\gamma}{t} + t} = \frac{\gamma}{\gamma + t^2}.$$

Furthermore, it is easy to see that for a Γ-Erlang distribution in parameter *q* and with γ/t substituted for *p* (see the table) that

$$d(t) = \frac{\gamma^q}{\left(\gamma + t^2\right)^q}.$$

We now treat an example that serves to illustrate the ways in which non-stationarity can complicate things.

Example 6: Distinct non-stationary discount rate spectra can generate the same generalized discount function

In the previous example it was shown that the generalized discount function generated by the non-stationary Γ-Erlang discount rate distribution in parameter *q* and with *p* replaced by γ/t was $d(t) =$

$\gamma^q/(\gamma+t^2)^q$. We now ask whether this discount function could be generated by some other non-stationary discount rate spectrum. In particular, does there exist some function of time, $\chi(t)$, which can be substituted for the parameter p in the exponential distribution, say, such that the same generalized discount function obtains? Start out by making the substitution in the parameterized form of $d(t)$ from Table *A* and equating the result with the desired function, like

$$\frac{\chi(t)}{\chi(t) + t} = \frac{\gamma^q}{\left(\gamma + t^2\right)^q}.$$

Now solve for $\chi(t)$ and after some algebra there results

$$\chi(t) = \frac{\gamma^q t}{\left(\gamma + t^2\right)^q - \gamma^q}.$$

Substituting this expression for p into the original discount rate spectrum one obtains the non-stationary spectrum

$$S(r|t) = \frac{\gamma^q}{\left(\gamma + t^2\right)^q - \gamma^q} \exp\left(\frac{-\gamma^q tr}{\left(\gamma + t^2\right)^q - \gamma^q}\right).$$

The non-stationary Γ-Erlang distribution can be written in a way that facilitates direct comparison

$$S(r|t) = \frac{\gamma^q r^{q-1}}{t^q \Gamma(q)} \exp\left(\frac{-\gamma r}{t}\right).$$

Algebraically, these appear to be very different, although notice that for $q = 1$ they are the same. They are compared graphically in figure 4 for $q = 1/2$ and $\gamma = 1$, where the conditional density is along the vertical axis, r increases to the lower right and t increases to the upper right:
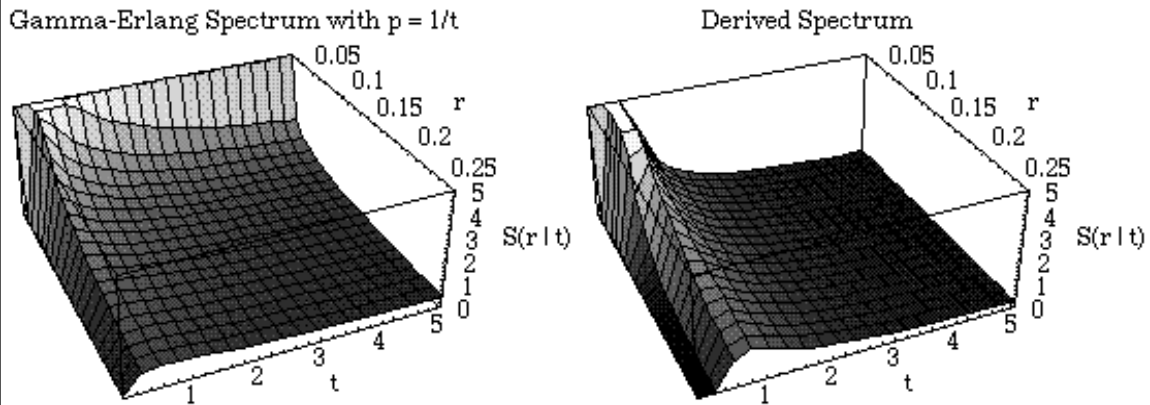

Figure 4, *a* and *b*: Comparison of two distinct discount rate spectra that generate the same generalized discount function

Clearly these are very different spectra. This can also be seen from the moments of the spectra. The mean and variance of the Γ-Erlang spectrum can be shown to be qt/γ and qt^2/γ^2, respectively, while for the derived spectrum these moments are more complicated,

$$\frac{\left(\gamma + t^2\right)^q - \gamma^q}{\gamma^q t}, \frac{\left[\left(\gamma + t^2\right)^q - \gamma^q\right]^2}{\gamma^{2q} t^2}$$

respectively. Once again, note that in the case of $q = 1$ the spectra share the same moments, i.e., they are one and the same spectrum.

The main implication of non-stationarity is now clear. While in the stationary case there is a one-to-one relationship between a discount rate spectrum and the generalized discount function it produces (proposition 3), no such relationship holds in the non-stationary case. In particular, there are an *infinite number* on non-stationary spectra which produce a given generalized discount function. Thus the *assumption* of stationarity in dealing with data is a very strong assumption indeed.

Non-stationary discount rate spectra are capable of producing generalized discount functions that are not only not *CM* but possibly not decreasing. To see this we have only to differentiate wrt time, obtaining

$$d'(t) = \int_0^\infty \frac{\partial}{\partial t}\left[S(r|t)\exp(-rt)\right]dr = \int_0^\infty \left[\frac{\partial S(r|t)}{\partial t} - rS(r|t)\right]\exp(-rt)dr$$

where use has been made of Leibniz's rule for differentiating an integral [Kreider, Kuller et al. 1966]. When the first term in the second integral vanishes, i.e., when the spectrum is not conditional on t, the previous results on stationary spectra apply. If this first term is small vis-a-vis the second term then it will turn out that $d(t)$ will at least be monotone decreasing. However, if the first term is large than it is possible for the generalized discount function to actually *increase* with time.

The ability of non-stationary spectra to generate discount functions that are not *CM* was actually displayed in example 5 above. For $d(t) = \gamma/\left(\gamma + t^2\right)$, $d'(t) \le 0$ for all t, but the 2nd derivative can be written as

$$d''(t) = 2\gamma \frac{3t^2 - \gamma}{\left(\gamma + t^2\right)^3}.$$

This expression is negative on $t \in [0, \sqrt{\gamma/3})$, and therefore $d(t)$ is not *CM*.

It would appear that *d(t)* functions having very complicated behavior over time can be adequately modeled with the non-stationary version of the discount rate spectrum approach described above. It remains for empirical work to determine whether the formalism is sufficiently powerful to handle most observed discounting behaviors. It may turn out

that discount functions that are not *CM* are relatively rare in practice.

A.2. *Inverse problem: Spectrum from a discount function*

When the spectrum is known to be non-stationary the problem of determining the discount spectrum from the generalized discount function is significantly more complicated. In this case the unknown function, $S(r|t)$, has two arguments and appears under the integral sign in equation (19) above. In general, we cannot expect to be able to solve this equation uniquely; in general there will be an infinite number of solutions. Non-uniqueness was illustrated in example 8 above, where two distinct spectra produced the same generalized discount function. That is, the addition of the second independent variable in the spectrum has imparted an additional degree of freedom to the problem. This non-uniqueness may manifest itself problematically in a number of ways. First, one may never be certain about the validity of the spectrum obtained unless it can be said without doubt that the spectrum is stationary. Second, small perturbations in the data may produce quite different spectra. Third, the set of non-stationary spectra that generates a particular discount function may be dense in the space of all non-stationary spectra and therefore individual spectra are not distinguishable.

One can study this equation under special assumptions on the non-stationary spectrum in order to obtain somewhat stronger results. For example, if the spectrum is *degenerate* then it can be written as $S(r|t) = S_1(t)\,S_2(r)$ and the second term can be taken outside the integral:

$$d(t) = S_1(t)\int_0^\infty S_2(r)\exp(-rt)dr$$

Insofar as we are treating *d(t)* as given throughout this section, we can consider $S_1(t)$ as a nonlinear transformation of the data—which we have the freedom to specify—and solve the resulting equation for the now unique[20] $S_2(r)$. That is, for each particular transformation of the data by dividing $d(t)$ by $S_1(t)$, there exists a discount rate spectrum $S_2(t)$ that solves the equation that defines the generalized discount function. So we recover uniqueness when we have separability.

Since the generalized discount rate functions produced by stationary spectra are *CM*, it must be the case that when a discount function is encountered which is not *CM* that it is the result of a non-stationary spectrum of rates. Alternatively, if the discount function is not completely monotone over some significant portion of the time domain it may be the case that the entire generalized discount function formalism developed herein is not well-suited to treat it. There is a definite sense, however, in which the generalized discounting apparatus is something like the minimum structure necessary for preserving 'exponential-like' discounting, since the generalized discount functions it produces can

---

[20] Uniqueness is a consequence of Proposition 3.

always be interpreted as possessing a 'mixture' of exponential rates. Some of these matters are illustrated in the following example.

Example 7: Cropper, Aydede and Portney on time preference for saving lives

    Cropper *et al.* [1992; 1994] used a direct survey to investigate individuals' implicit discount rates for saving lives. They found extensive heterogeneity in time preferences and observed discounting behavior that was apparently non-exponential. To model this latter phenomenon these authors assumed the discount function to be of standard exponential type with time-varying discount rate, i.e., $d(t) = $ exp$[-r(t)t]$. In particular, they postulated a rate linearly decreasing with time like $r(t) = r_0 - r_1 t$, producing $d(t) = $ exp$[-r_0 t + r_1 t^2]$. This yielded a significant fit of the data with one set of constants r_0 and r_1 over a short time horizon and a different set for a long horizon.

    Clearly this model is only valid over limited times for eventually the discount rate will become negative—i.e., for $t > r_0/r_1$—as the authors point out. This discount function is not *CM* over the range of t for which it displays a positive discount rate. The first derivative is

$$d'(t) = \left(-r_0 + 2r_1 t\right)\exp\left(-r_0 + r_1 t^2\right) = \left(-r_0 + 2r_1 t\right)d(t)$$

This expression is *positive* for all $t > r_0/2r_1$, i.e., $d(t)$ is *increasing* over the latter half of the range of t for which the discount rate is positive.

    The linear variation of discount rate with time is certainly just an analytical convenience and the authors hypothesize that the "discount rate follows a negatively sloped, convex pattern." Here we investigate whether or not a discount function produced by such a rate function can fit into the generalized discounting formalism. Start by checking for complete monotonicity with the derivatives of $d(t) = $ exp$[-r(t)t]$, with $r(t) > 0$, $r'(t) < 0$ and $r''(t) > 0$ for all t, by assumption. If $d(t)$ is *CM* then we are assured there exists a discount rate spectrum which produces it. The first derivative, which needs to be *negative* for complete CM, is

$$d'(t) = \left[-r(t) - tr'(t)\right]\exp\left[-r(t)t\right] = \left[-r(t) - tr'(t)\right]d(t).$$

Since $r'(t) < 0$ for all t, $d(t)$ will be completely monotone up through the first value of t, say t'—the 'cross-over' time—that solves $t = -r(t)/r'(t)$; indeed, there may be more than one t that makes this expression true.

    However, this is not the whole story. For certain functional forms of $r(t)$ the cross-over time, t', may not exist. To see this we can take advantage of the monotone structure of the problem to solve for the marginal rate function, call it $r^*(t)$, for which the discount function, $d(t)$, is never CM. That is, solve $-r(t) - tr'(t) = 0$ and the $r^*(t)$ obtained will be a kind of border of feasible $r(t)$. This equation is a linear homogeneous ODE, whose solution is $r^*(t) = k/t$, with k a constant.

    In particular, consider the class of $r(t)$ functions that scale like t^ζ. We will show that for $\zeta \in [0, 1)$, *CM* discount functions are produced. For $\zeta = 0$ the discount rate function, $r(t)$, is constant and the usual

exponential discount function obtains. For $\zeta < 0$ the discount function eventually increases with time—as happened above in the case of $r(t)$ varying linearly, i.e., $\zeta = -1$. For $\zeta > 1$ the cross-over time, t', exists and is finite. To see this first calculate $r'(t) \propto -\zeta t^{\zeta-1}$. Then $tr'(t) \propto -\zeta t^{\zeta}$ and $r(t)$ - $tr'(t) \propto -t^{\zeta} + \zeta t^{\zeta} = (\zeta-1)t^{\zeta}$. This last expression is clearly greater than 0 for $\zeta > 1$ and so implies that d(t) will not be *CM* in this part of the parameter space. The restrictions placed on $r(t)$ by Cropper *et al.*— that it be "negatively sloped" and "convex"—are clearly not sufficient to insure that $d(t)$ does not possess certain pathological and seemingly unreasonable properties, such as being an increasing function of time (i.e., $d'(t)$ positive). While there seems to be no *a priori* reason why discount functions need to be *CM*, they should at least be not increasing.

Analysis of higher derivatives of $r(t)$ produces a similar picture. We have shown that the set **R** of discount rate spectra that yield feasible $d(t)$ is somewhat smaller than the set **C** implicitly specified by Cropper *et al.*, i.e., **C** $\supset$ **R**. It is not apparent how to characterize **R** but it turns out that it is possible to completely specify the set **M** $\subset$ **R** of discount rate functions which produce CM discount functions. To do this succinctly, an additional result on complete monotonicity is necessary:

Proposition 9: *If y(x) is CM and z(x) is positive with CM derivatives then y(z(x)) is CM.*

Proof: Feller [1971: 441]   ❏

Let $1/exp(\cdot)$ assume the role of $y(\cdot)$ in the proposition; clearly it is CM. Then specify $R(t) \equiv r(t)t$ be $z(x)$. $R(t)$ is positive for all t and can be seen to have monotone derivatives whenever $r(t)$ scales like t^{ζ} with $\zeta \in [0, 1)$, since this makes $R(t)$ scale like t^{ξ} with $\xi \in (0,1]$ and so $R'(t)$ goes like $t^{\xi-1}$. Therefore, for this set of parameters $\exp[-r(t)t]$ is CM. The set of all $r(t)$ having elements that scale like t^{ζ} with $\zeta \in [0, 1)$ is **M** from above.

B. *Spectra that are not Measures*

The requirement that $d(0) = 1$, i.e., no discounting of the present, is formally equivalent to the requirement that $\int_0^{\infty} S(r)dr = 1$. Clearly, a *sufficient* condition for a function to be a valid discount rate spectrum is that it be a measure. However, this is not a *necessary* condition. Functions that are not everywhere positive, and therefore not standard measures, may have this property.

Complete monotonicity is the main casualty of spectra of this type. To see this compute the discount function derivative with respect to time:

$$d^{(k)}(t) = (-1)^k \int_0^{\infty} r^k S(r)\exp(-rt)dr$$

For $S(r)$ not everywhere positive it is no longer possible to guarantee that the derivatives have uniform sign over the whole range of t.

B.1. *Discount functions from non-measure spectra*

Spectra that are not valid measures are the subject we explore here. Specifically, spectra that are not everywhere positive—that have, in essence, negative probability—are considered.

Example 8: Variations on a theme of Aris [1991]

Consider the following spectrum composed of an infinite number of equally-spaced Dirac measures having hyperbolically declining weight:

$$S(r; r_0) = \frac{1}{\ln(2)} \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j} \delta(r - jr_0),$$

Note that the odd terms in the sum are positive while the even ones are negative. Because its components alternate in sign this spectrum is not a valid probability measure (in the conventional sense). However, it is a straightforward matter to show that

$$d(t; r_0) = \frac{1}{\ln(2)} \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j} \exp(-jr_0 t)$$

is a reasonably well-behaved discount function, albeit not *CM*.
Actually observing such discount rate spectra in reality would depend on careful transform inversion [Abate, Choudhury et al. 1994]. For instance, if one postulates that a normal probability measure underlies one's data then non-measure spectra would be excluded by assumption.

B.2. *Non-measure spectra from non-CM discount functions*

We have just derived a non-*CM* discount function from a peculiar (non-measure) spectrum. It is also possible to do the reverse, to derive properties of *S(r)* from discount functions that are not *CM*. For example:

Proposition 10: *If S(r) has n changes of sign then the derivatives $d^{(k)}(t)$ possess exactly n real zeroes for sufficiently large k. When S(r) has a change of sign at r_0 then $d^{(k)}(t)$ possess a zero t_k with*

$$\lim_{k \to \infty} \frac{k}{t_k} = r_0.$$

<u>Proof</u>: Doetsch [1961: 204-5].   ❏

Using this result may not be easy, especially when experimental discounting data are involved, as numerical differentiation is an inherently noisy process. However, if the data are first fit to some functional form, the derivatives could subsequently be computed.

C. *Negative Discount Rates*

Negative discount rates are an important feature, empirically [e.g., Loewenstein and Prelec 1991; Samwick 1998]. Consider a stationary

spectrum having some of its support in the negative half of the real line. Such a spectrum will not generally be integrable, because of the positive mass at the origin as well as for negative values of *r*. Therefore, the only way to deal with such spectra, and the potentially quite peculiar discount functions that result, is to associate with them a finite time horizon.

VI   Summary and Conclusions

Empirically, microeconomic agents do not normally discount exponentially. Socially, there are many practical situations in which there exists a distribution of discount rates over the components of an economic aggregate. These two distinct motivations have led us to develop a general mathematical formalism for non-exponential discounting. The main features of this formalism have been described.

It has been shown that exponential discounting obtains only in the special case of a Dirac measure of discount rates. Any discount rate heterogeneity creates non-exponential discounting that is 'weakly hyperbolic' with respect to some exponential discount function. In the special case of a Γ-Erlang distribution of discount rates, 'strongly hyperbolic' discounting obtains overall. Sub-additive discount functions also fall naturally within this formalism, the result of stable discount rate spectra.

The *time consistency* and *preference reversal* behavior of generalized discounting has been explored. It was argued that micro-level time consistency does not generally imply macro-level time consistency. However, macro-time consistency does imply micro-time consistency.

The *integrability* of generalized discount functions has been studied, and a theorem proved concerning the structure of discount rate spectra necessary and sufficient to make present value integrals finite.

The *inverse problem*—of determining a discount rate spectrum from observed discounting behavior—has been investigated. The manifold effects of spectrum *non-stationarity* have been elucidated.

The present theory is an explanation of discounting behavior in terms of an underlying spectrum of exponential discount rates. It is closely related to *multiple selves* interpretations of non-exponential discounting. It remains to work out just what spectra govern human behavior in specific experimental and real-world situations. Doing so will lead to empirical progress in modeling economic behavior.

Appendix
A table of discount rate spectra and associated discount functions

Here we provide a short table of discount rate spectra and the corresponding generalized discount functions, developed from a variety of sources, including Feller [1971] and LePage [1980]. Some of these spectra are quite simple, some rather complicated. Table *A* includes several well-known *pdf*s and some less common ones.

$S(r)$	mean	$d(t)$	L^1?
uniform on $[p, q]$	$\dfrac{p+q}{2}$	$\dfrac{\exp(-pt) - \exp(-qt)}{(q-p)t}$	yes if $p > 0$
triangular on $[0,2p]$	p	$\dfrac{1 - \left[\exp(-pt)\right]^2}{p^2 t^2}$	yes
exponential: $p\exp(-pr)$	$\dfrac{1}{p}$	$\dfrac{p}{p+t}$	no
Γ-Erlang: $\dfrac{p^q r^{q-1}}{\Gamma(q)}\exp(-pr)$	$\dfrac{q}{p}$	$\dfrac{p^q}{(p+t)^q}$	yes if $q > 1$
1/2 normal: $\sqrt{\dfrac{2}{\pi}}\dfrac{1}{p}\exp\left(-\dfrac{r^2}{2p^2}\right)$	$p\sqrt{\dfrac{2}{\pi}}$	$\exp\left(\dfrac{p^2 t^2}{2}\right)\mathrm{erfc}\left(\dfrac{pt}{\sqrt{2}}\right)$	no
Rayleigh: $\dfrac{r}{p^2}\exp\left(-\dfrac{r^2}{2p^2}\right)$	$p\sqrt{\dfrac{\pi}{2}}$	$1 - \sqrt{2\pi}\,\dfrac{pt}{2}\exp\left(\dfrac{p^2 t^2}{2}\right)\mathrm{erfc}\left(\dfrac{pt}{2}\right)$	yes
1/2 Cauchy: $\dfrac{2p}{\pi(1+p^2 r^2)}$	does not exist	$\dfrac{2}{\pi}Ci\left(\dfrac{t}{b}\right)\sin\left(\dfrac{t}{b}\right) + \cos\left(\dfrac{t}{b}\right)\left[1 - \dfrac{2}{\pi}Si\left(\dfrac{t}{b}\right)\right]$	no
Lévy: $\dfrac{p}{2\sqrt{\pi r^3}}\exp\left(-\dfrac{p^2}{4r}\right)$	does not exist	$\exp(-p\sqrt{t})$	yes
Bessel: $\dfrac{p}{r}\exp(-r)I_p(r)$	does not exist	$\left(t+1 - \sqrt{(t+1)^2 - 1}\right)^p$	yes if $p > 1$
Dirac: $\delta(r-p)$	p	$\exp(-pt)$	yes if $p > 0$

Table *A*: Discount rate spectra and associated discount functions[21]

Example A: Use of table *A* to find a generalized discount function

For a uniform distribution of *r* on [0.03, 0.08], from the first line of the table *d(t)* = [exp(-0.03*t*) – exp(-0.08*t*)]/0.05*t*.. Figure A is a plot of this function, compared to two exponential discount functions, one each from the limits of the range on *r*, i.e., with *r* = 0.03 and *r* = 0.08; the generalized discount function is the interior one.

---

[21] In this table $r \in \boldsymbol{R}_+$, *p* and $q \in \boldsymbol{R}_{++}$, *erfc*(·) is the complementary Gaussian error function, I_p(·) is the modified Bessel function of order *p*, *Ci*(·) and *Si*(·) are the cosine integral and sine integral, respectively, and δ(·) is the Dirac delta function.
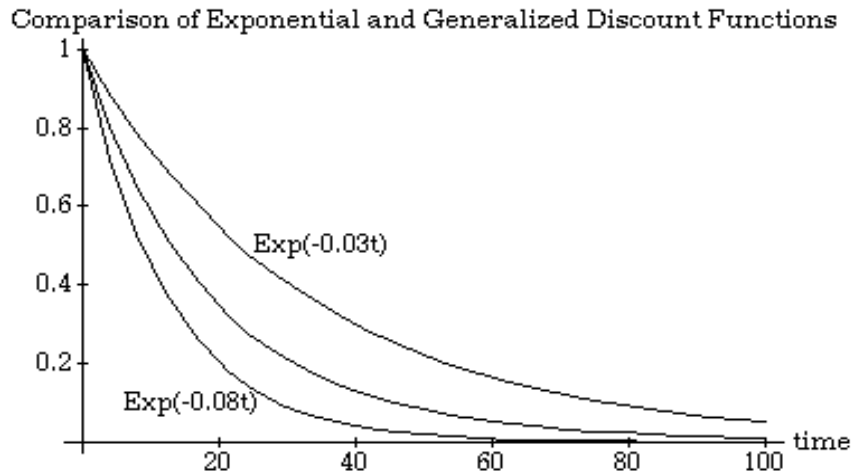
Figure A: Generalized discount function resulting from a uniform discount rate distribution on [0.03, 0.08]

Note that very early on the generalized discount function closely resembles the discount function with the larger rate, while for long time it approaches the function having the smaller rate.

Applying this discount function will produce different present values than that obtained from using the average rate, as is common practice [cf. Lave 1981]. The generalized discount function accurately 'weights' the average to account for all values of *r* in the distribution.

References

Abate, J., G. L. Choudhury and W. Whitt (1994). "Waiting-Time Tail Probabilities in Queues with Long-Tail Service-Time Distributions." <u>Queueing Systems</u> **16**: 311-338.

Abate, J. and W. Whitt (1995). "Numerical Inversion of Laplace Transforms on Probability Distributions." <u>ORSA Journal of Computing</u> **7**: 36-43.

Abate, J. and W. Whitt (1999). "Computing Laplace Transforms for Numerical Inversion via Continued Fractions." <u>INFORMS Journal on Computing</u> **11**(4): 394-405.

Abramowitz, M. and I. A. Stegun (1964). <u>Handbook of Mathematical Functions</u>. Washington, D.C., National Bureau of Standards.

Ainslie, G. W. (1992). <u>Picoeconomics</u>. New York, N.Y., Cambridge University Press.

Ainslie, G. W. and N. Haslam (1992). Hyperbolic Discounting. <u>Choice Over Time</u>. G. Loewenstein and J. Elster. New York, Russell Sage**:** 57-92.

Ainslie, G. W. and R. J. Herrnstein (1981). "Preference Reversal and Delayed Reinforcement." <u>Animal Learning Behavior</u> **9**(4): 476-482.

Aris, R. (1991). Multiple Indices, Simple Lumps, and Duplicitous Kinetics. <u>Chemical Reactions in Complex Mixtures</u>. A. V. Sapre and F. J. Krambeck. New York, N.Y., Van Nostrand Reinhold.

Axtell, R. L. (1992). Theory of Model Aggregation for Dynamical Systems with Application to Problems of Global Change. <u>Ph.D. Dissertation</u>. Pittsburgh, Penn., Carnegie Mellon University**:** xxxvii, 435 p. :.

Ayres, R. U. and R. L. Axtell (1996). "Foresight as a Survival Characteristic : When (if Ever) Does the Long View Pay?" <u>Technological Forecasting and Social Change</u> **51**: 209-235.

Ayres, R. U. and s. Mori (1989). "Time Preference and the Life Cycle: The Logic of Long-Term High Risk vs. Short-Term Low Risk." <u>European Journal of Operational Research</u> **38**: 329-349.

Barro, R. J. (1999). "Ramsey Meets Laibson in the Neoclassical Growth Model." Quarterly Journal of Economics **114**(4): 1125-1152.

Bernstein, S. N. (1928). "Sur les fonctions absolument monotones." Acta Mathematica **52**: 1-66.

Billingsley, P. (1986). Probability and Measure. New York, N.Y., John Wiley & Sons.

Blanchard, O. J. and S. Fischer (1989). Lectures on Macroeconomics. Cambridge, Mass., MIT Press.

Cadzow, J. A. (1973). Discrete-Time Systems: An Introduction with Interdisciplinary Applicatins. Englewood Cliffs, N.J., Prentice-Hall, Inc.

Cairns, J. A. and M. M. van der Pol (1999). "Do People Value Their Own Future Health Differently Than Others' Future Health?" Med. Decision Making **19**(4): 466-472.

Chung, S.-H. and R. J. Herrnstein (1961). "Relative and Absolute Strengths of Response as a Function of Frequency of Reinforcement." Journal of the Experimental Analysis of Animal Behavior **4**: 267-272.

Cropper, M. L., S. K. Aydede and P. R. Portney (1992). "Rates of Time Preference for Saving Lives." American Economic Review **82**(2): 469-472.

Cropper, M. L., S. K. Aydede and P. R. Portney (1994). "Preferences for Life Saving Programs: How the Public Discounts Time and Age." Journal of Risk and Uncertainty **8**(243-265).

Doetsch, G. (1961). Guide to the Applications of Laplace Transforms. Princeton, N.J., D. Van Nostrand Company, Ltd.

Doetsch, G. (1970). Introduction to the Theory and Application of the Laplace Transformation. New York, N.Y., Springer-Verlag.

Epstein, L. G. and J. A. Hynes (1983). "The Rate of Time Preference and Dynamic Economic Analysis." Journal of Political Economy **91**(4): 611-635.

Feller, W. (1971). An Introduction to Probability Theory and Its Applications, Volume II. New York, N.Y., John Wiley & Sons.

Frederick, S., G. Loewenstein and T. O'Donoghue (2002). "Time Discounting and Time Preference: A Critical Review." Journal of Economic Literature **40**(2): 351-401.

Glimcher, P. W. (2003). Decisions, Uncertainty and the Brain: The Science of Neuroeconomics. Cambridge, Mass., MIT Press.

Green, L., N. Fristoe and J. Myerson (1994). "Temporal Discounting and Preference Reversals in Choice Between Delayed Outcomes." Psychonomic Bull. Rev. **1**(3): 383-389.

Green, L., J. Myerson and E. McFadden (1997). "Rate of Temporal Discounting Decreases with Amount of Reward." Memory & Cognition **25**(5): 715-723.

Harris, C. M. (1968). "The Pareto distribution as a queue service discipline." Operations Research **16**: 307-313.

Harrison, G. W., M. I. Lau and M. B. Williams (2002). "Estimating Individual Discount Rates for Denmark." American Economic Review **92**(5): 1606-1617.

Harvey, C. M. (1986). "Value Functions for Infinite-Period Planning." Management Science **32**(9): 1123-1139.

Harvey, C. M. (1995). "Proportional Discounting of Future Cost and Benefits." Mathematics of Operations Research **20**(2): 381-399.

Hausman, J. A. (1979). "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables." Bell Journal of Economics **10**(1): 33-54.

Herrnstein, R. J. and J. E. Mazur (1987). Making Up Our Minds: A New Model for Economic Behavior. The Sciences.

Holden, S. T., B. Shiferaw and M. Wik (1998). "Poverty, market imperfections and time preferences: of relevance for environmental policy?" Environment and Development Economics **3**: 105-130.

Horowitz, J. K. (1991). Discounting Money Payoffs: An Experimental Analysis. Handbook of Behavioral Economics. S. Kaisch and B. Gilad. Greenwich, Conn., JAI Press. **II**.

Horowitz, J. K. (1992). "A Test of Intertemporal Consistency." <u>Journal of Economic Behavior and Organization</u> **17**: 171-182.

Horowitz, J. K. and R. T. Carson (1990). "Discounting Statistical Lives." <u>Journal of Risk and Uncertainty</u> **3**: 403-413.

Howe, C. W. (1990). "The Social Discount Rate." <u>Journal of Environmental Economics and Management</u> **18**.

Jamison, D. T. and J. S. Jamison (2003). Discounting. <u>DCPP Working Paper, no. 4</u>. Washington, D.C., NIH**:** 45 pages.

Kenley, C. R. and D. C. Armstead (2003). "Discounting Models For Long-Term Decision Making." <u>Systems Engineering</u> **7**(1): 13-24.

Kirby, K. N. and R. J. Herrnstein (1995). "Preference Reversals due to Myopic Discounting of Delayed Reward." <u>Psych. Sci.</u> **6**(2): 83-89.

Kreider, D. L., R. G. Kuller, D. R. Ostberg and F. W. Perkins (1966). <u>An Introduction to Linear Analysis</u>. Reading, Mass., Addison-Wesley Publishing Company.

Laibson, D. I. (1997). "Golden Eggs and Hyperbolic Discounting." <u>Quarterly Journal of Economics</u> **112**: 443-477.

Lave, L. B. (1981). <u>The Strategy of Social Regulation: Decision Frameworks for Policy</u>. Washington, D.C., Brookings Institution Press.

LePage, W. R. (1980). <u>Complex Variables and the Laplace Transform for Engineers</u>. New York, N.Y., Dover Publications, Inc.

Loewenstein, G. (1987). "Anticipation and the Valuation of Delayed Consumption." <u>Economic Journal</u> **97**(387): 666-684.

Loewenstein, G. and D. Prelec (1991). "Negative Time Preference." <u>American Economic Review</u> **81**(2): 347-352.

Loewenstein, G. and D. Prelec (1992). "Anomalies in Intertemporal Choice: Evidence and Interpretation." <u>Quarterly Journal of Economics</u> **107**(2): 573-597.

Loewenstein, G., D. Read and R. Baumeister, Eds. (2003). <u>Time and Decision</u>. New York, N.Y., Russell Sage Foundation.

Loewenstein, G. and R. H. Thaler (1989). "Intertemporal Choice." <u>Journal of Economic Perspectives</u> **3**(4): 181-193.

Luenberger, D. G. (1979). <u>An Introduction to Dynamical Systems: Theory, Models and Applications</u>. New York, N.Y., John Wiley & Sons.

Lyons, R. (1990). Discount Rate Policy: Theoretical and Practical Challenges. Washington, D.C., Resources for the Future.

Maguire, B. A., E. S. Pearson and H. A. Wynn (1952). "The time intervals between industrial accidents." <u>Biometrika</u> **39**: 168-180.

Marschak, J. (1964). "Actual versus Consistent Behavior." <u>Behavioral Science</u> **9**: 103-110.

Mischel, W. (1984). "Convergences and Challenges in the Search for Consistency." <u>American Psychologist</u> **39**: 351-364.

Moore, M. J. and W. K. Viscusi (1990). "Discounting Environmental Health Risks: New Evidence and Policy Implications." <u>Journal of Environmental Economics and Management</u> **18**: S51-S62.

Ok, E. A. and Y. Masatlioglu (forthcoming). "A Theory of (Relative) Discounting." <u>Journal of Economic Theory</u>.

Page, T. (1977). <u>Conservation and Economic Efficienty</u>. Baltimore, Maryland, Resources for the Future, Johns Hopkins University Press.

Phelps, E. S. and R. A. Pollak (1968). "On Second-Best National Saving and Game-Equilibrium Growth." <u>Review of Economic Studies</u> **35**: 185-199.

Pollak, R. (1968). "Consistent Planning." <u>Review of Economic Studies</u> **35**: 201-208.

Read, D. (2001). "Is Time-Discounting Hyperbolic or Subadditive?" <u>Journal of Risk and Uncertainty</u> **23**(1): 5-32.

Samorodnitsky, G. and M. S. Taqqu (1994). <u>Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance</u>. Washington, D.C., Chapman & Hall.

Samuelson, P. A. (1937). "A Note on Measurement of Utility." <u>Review of Economic Studies</u> **IV**(2): 155-61.

Samwick, A. A. (1998). "Discount Rate Heterogeneity and Social Security Reform." <u>Journal of Development Economics</u> **57**(1): 117-46.

Sargent, T. J. (1987). <u>Macroeconomic Theory</u>. San Diego, Academic Press.

Schelling, T. C. (1984). "Self-Command in Practice, in Policy, and in a Theory of Rational Choice." <u>American Economic Review</u> **74**(2): 1-11.

Scholten, M. and D. Read (2006). "Discounting by Intervals: A Generalized Model of Intertemporal Choice." <u>Management Science</u> **52**(9): 1424-1436.

Sozou, P. D. (1998). "On hyperbolic discounting and uncertain hazard Rrtes." <u>Proceedings of the Royal Society of London B</u> **265**: 2015-2020.

Strotz, R. H. (1956). "Myopia and Inconsistency in Dynamic Utility Maximization." <u>Review of Economic Studies</u> **XXIII**: 165-180.

Thaler, R. H. (1981). "Some Empirical Evidence on Dynamic Inconsistency." <u>Economics Letters</u> **8**: 201-207.

van der Pol, M. M. and J. A. Cairns (1999). "Individual Time Preferences for Own Health: Application of a Dichotomous Choice Question with Follow Up." <u>Applied Economics Letters</u> **6**(10): 649-654.

van der Pol, M. M. and J. A. Cairns (2001). "Estimating Time Preferences for Health Using Discrete Choice Experiments." <u>Social Sci. Med.</u> **52**: 1459-1470.

Weitzman, M. L. (2001). "Gamma Discounting." <u>American Economic Review</u> **91**(1): 260-271.

Wolfram, S. (1991). <u>Mathematica: A System for Doing Mathematics by Computer</u>. New York, N.Y., Addison-Wesley.