Discounting

DEAN T. JAMISON
JULIAN S. JAMISON

DISEASE CONTROL PRIORITIES PROJECT

DCPP Working Paper No.4

MARCH 2003

World Bank • World Health Organization • Fogarty International Center, National Institutes of Health, U.S. Department of Health and Human Services

DISEASE CONTROL PRIORITIES PROJECT

BACKGROUND

In the late 1980s, the World Bank initiated work to inform priorities for control of specific diseases and to generate comparative cost-effectiveness estimates for interventions addressing the full range of conditions important in developing countries. The purpose of the comparative cost-effectiveness work was to provide one input into decision-making within the health sectors of highly resource-constrained countries. This process resulted in the 1993 publication of *Disease Control Priorities in Developing Countries**. A decade after publication of the first edition, the World Bank, the World Health Organization, and the Fogarty International Center (FIC) of the U.S. National Institutes of Health (NIH) have initiated a "Disease Control Priorities Project" (DCPP) that will, among other outcomes, result in a second edition of *Disease Control Priorities* in Developing Countries (DCP*2*). The DCPP is financed in part by a grant from the Bill & Melinda Gates Foundation. DCP*2* is intended both to update DCP*1* and to go beyond it in a number of important ways, e.g. in documentation of success stories, in discussion of institutional and implementation issues, and in explicit discussion of research and development priorities. Publication of DCP*2* is intended for mid-2005.

*This volume was edited by Dean T. Jamison, W. Henry Mosley, Anthony R. Measham and Jose Luis Bobadilla and published by Oxford University Press in 1993.

EDITORS, DISEASE CONTROL PRIORITIES IN DEVELOPING COUNTRIES, 2ND EDITION

Dean T. Jamison, University of California, Los Angeles and Fogarty International Center (NIH)
George Alleyne, Pan-American Health Organization, retired
Joel G. Breman, Fogarty International Center (NIH)
Mariam Claeson, World Bank
David B. Evans, World Health Organization
Prabhat Jha, University of Toronto
Anthony R. Measham, World Bank, retired
Anne Mills, London School of Hygiene and Tropical Medicine

THE WORKING PAPERS SERIES

The Working Papers Series makes available DCPP background papers and chapter drafts for early dissemination and critical reaction. Most entries in the series are intended for more formal publication later. Queries and observations concerning a paper should be addressed to the author indicated on the title page.

DISCLAIMER

Table of Contents

Disease Control Priorities Project
Working Paper No. 4

DISCOUNTING

March 2003

Dean T. Jamison
University of California, Los Angeles
U.S. National Institutes of Health

Julian S. Jamison
Kellogg School of Management
Northwestern University

Corresponding author:
Julian S. Jamison
Northwestern University
Kellogg School of Management
Department of Managerial Economics and Decision Sciences
2001 Sheridan Road
Evanston, IL 60208
j-jamison@kellogg.northwestern.edu

Comments on this paper may be posted or read
at the Disease Control Priories Project Web site
www.fic.nih.gov/dcpp

Suggested citation:
Discounting. By Dean T. Jamison and Julian S. Jamison. To be presented at the 2003 Congress of the European Economic Association, Stockholm, August 2003. Working Paper No. 4, Disease Control Priorities Project. Bethesda, Maryland: Fogarty International Center, National Institutes of Health. March 2003.

Abstract

This paper provides theoretical underpinnings for nonconstant rate discounting procedures. We begin by distinguishing the concepts of amount and speed of discounting. Exponential discounting – discounting at a constant rate – sequesters both concepts into a single parameter that needs to be disaggregated in order to adequately characterize nonconstant rate procedures. The present value accorded to a unit stream of net outcomes – a stream of flow value 1 from time zero to infinity – corresponds intuitively to the total weight that a discounting procedure assigns to the future. The inverse of the weight provides a natural measure of the *amount* that the procedure discounts away the future. Time-paths for the accumulation of present value can differ for procedures that discount the future by the same amount. We propose a measure for the *speed* of discounting that reflects these differences. With speed defined we are able to show exponential discounting to be the fastest procedure possible for nonincreasing discount rates and the slowest procedure for nondecreasing discount rates. Another key result links a discounting procedure's speed and amount with the time it takes to accumulate present value (its *time horizon*). The paper then characterizes a number of existing and new discounting procedures and provides closed form expressions for their amounts, speeds, and time horizons. Hyperbolic discounting, for example, can be parameterized in terms of speed and amount but fails to allow for the fast discounting implicit in the rising yield curves observed in today's bond markets. The hyperbolic family can, however, be shown to include procedures that are, in a well-defined sense, infinitely slow but that nontheless converge to a finite present value. Currently proposed procedures for aggregation of individual discount functions into a representative one – by averaging of discount rates or of discount functions – suffer important shortcomings. We propose an approach to aggregation, averaging of normalized discount functions (ANDF), that avoids these problems and that we show to result from simple averaging of the amounts and time horizons of the discount functions being aggregated. We further prove that under generally applicable assumptions ANDF results in slow discounting (relative to the speeds of the discount functions being aggregated).

Contents

This draft: March 26, 2003

DISCOUNTING[1]

by

Dean T. Jamison and Julian S. Jamison

Discounting at a constant rate has the virtues of familiarity, analytic tractability, time-consistency of preferences, and a well-understood axiomatic foundation. That said, perceived shortcomings of exponential discounting procedures (i.e. those that have a constant discount rate) have led economists in diverse fields increasingly to employ discounting procedures in which the discount rate varies over time. Examples occur in contexts involving the full range of time horizons.

In the case of long horizons – for example, hepatitis B immunization – benefits can occur half a century or so after the intervention. Returns to R & D in health (and elsewhere) can continue over even longer periods. Systematic efforts to assign priority to health interventions must include mechanisms for comparing interventions whose outcomes span much time with those whose benefits accrue as they are implemented, such as treatment for diabetes or renal dialysis. Environmental and education policies likewise have outcomes in the distant future – often, in the case of the environment, a substantially more distant future than for health. Using exponential discounting to generate a present value for outcomes in the far future can generate, with a sufficiently low discount rate, as modest a discounting of far-future outcomes as might be desired. Giving noticeable weight to the far future with exponential discounting comes at the cost, however, of entailing virtually no discounting in the short- to medium-term. Hence the literature on economic evaluation of environmental and health projects has, increasingly, contained proposals

[1] National Institutes of Health (NIH) Contract No. YI-10005702 provided partial support for preparation of this paper. Jia Wang provided helpful inputs, and the authors received valuable comments on an early draft at a seminar at Resources for the Future. Johannes Hörner and Andrew Metrick provided useful guidance, and Nancy Hancock provided valuable research assistance. Parts of Section 1 were presented to a plenary session of the 2nd Congress of the International Health Economics Association, Rotterdam, June 1999. Conclusions and opinions expressed are those of the authors and not of the U.S. government.

for use of discount rates that decline with time, or 'slow discounting', although actual practice for the most part continues to use constant rates.[2,3]

A more recent strand of work in economics imports findings from experimental psychology into the understanding of economic behavior (George Loewenstein, 1992; Frederick Shane, Loewenstein and Ted O'Donoghue, 2002). Psychologists – and the behavioral economists using their findings and methods – focus much of their interest on behavior over relatively short periods, behavior that often seems consistent with initially high but rapidly falling discount rates. This generates interest in slow discounting from the perspective of short time horizons in addition to the far longer horizons of those with concerns for health and the environment.

The time period of interest for financial markets lies between that of the behavioral and environmental economists – typically six months to thirty years. The rising yield curves that characterize today's bond markets imply fast discounting (in a sense that we will make clear). Financial economics, then, includes a third strand of analysis that routinely employs non-constant rate discounting.

The multiplicity of origins of interest in non-constant rate discounting poses the challenge of generating discounting procedures that are flexible enough to handle short, medium and long time horizons and that are capable of generating discounting that is both slow and fast. Our purpose in this paper is to provide theoretical underpinnings for the increasingly diverse literature using non-constant rate discounting by drawing implications of distinguishing the *speed* of discounting from the total *amount* by which the future is discounted. Once these concepts are separated, the existing heterogeneous collection of proposals about how to discount can be quantitatively characterized and alternative proposals for aggregation of multiple individual procedures into a

[2] This paper has been prepared as background to the Disease Control Priorities Project (DCPP), an effort to assess the cost-effectiveness of a broad range of health interventions. The DCPP is cosponsored by the World Bank, the National Institutes of Health (NIH), the World Health Organization and the Bill & Melinda Gates Foundation. Antecedents to the DCPP (Dean T. Jamison et. al., 1993; World Bank, 1993) used constant rate discounting (at 3%), which is consistent with current guidelines for economic evaluation of health projects (Joseph Lipscomb, Milton C. Weinstein and George W. Torrance, 1996). The specific practical issue we seek to illuminate in this paper is whether – and, if so, how – the DCPP should depart from current best practice in the health cost-effectiveness literature.

[3] Although almost all advocates of variable rate discounting for project evaluation propose slow discounting, Amartya Sen (1972) has noted circumstances where fast discounting might be appropriate.

social one can be better evaluated. Important work continues on the nature of the factors (e.g. time preference, the return to capital and the completeness of financial markets) that influence discounting's magnitude, and on the arguments for and against discounting at a constant rate. There is a natural division of labor between these topics and task of characterizing the basic structure of discounting. This paper addresses the latter, the theory of discounting itself.

Section 1 provides examples of slow and fast discounting and introduces our approach by providing definitions of the amount, the relative speed and the time horizon of discounting procedures. We next discuss the ways discounting procedures can be characterized – i.e. by discount rates, discount functions, discount factors, yield curves, and present value functions – then summarize (and expand) the list of discounting procedures currently in the literature. Section 3 formalizes our definitions and conveys our results relating the amount, the speed, and the time horizons of discounting procedures. Section 4 deals with aggregating multiple discounting procedures into a representative one, and Section 5 draws brief conclusions for policy. An Appendix contains derivations of results stated in the text.

1 Introduction: Basic Concepts

If $b(t)$ is a time stream of net benefits then the standard formula for exponential discounting,

$$\text{present value of } b(t) = \int_0^\infty b(t)e^{-rt}dt \tag{1}$$

defines the present value of $b(t)$ as a function of the discount rate r.[4] Focusing on the special case of a unit stream of benefits $[b(t) = 1, 0 \leq t < \infty]$ allows examination of characteristics of the discounting procedure itself. We use this approach throughout. Two of the ways of defining a discounting procedure are by a discount function, $d(t)$ – where $d(0) = 1$, $d(t) \geq 0$ and $d(t)$ nonincreasing for all t – or by a present value function, $pv(t)$, that gives the present value of a unit stream of benefits accumulated to time t :

$$pv(t) = \int_0^t d(x)dx. \tag{2}$$

[4]The literature in engineering and physics refers to the present value of $b(t)$, expressed as a function of r, as its Laplace transform, and extensive tables of Laplace transforms provide a convenient source of closed form expressions for present values. See, for a brief treatment, I. S. Skolnikoff and R. M. Redheffer (1958, Appendix B).

We define the 'present value of $d(t)$' to be $pv(\infty)$, and the initial subsection of this chapter discusses the speed and time horizon of a discounting procedure in terms of how $pv(t)$ rises to $pv(\infty)$.

When discounting at a constant rate r, $pv(\infty) = r^{-1}$, which increases as r declines and becomes infinite when $r = 0$. More generally if $\int_0^\infty d(t)dt$ fails to converge, $pv(\infty) = \infty$. The infinite present value of exponential discounting when $r = 0$ raises the issue of nonconvergence of other discounting procedures for which $\lim_{t\to\infty} r(t) = 0$. The literature has neglected this issue although a number of procedures that have been proposed for slow discounting do, in fact, fail to converge. The second subsection of this section discusses these issues and introduces concepts of strong and weak convergence that play important roles later in the paper. Weakly convergent $d(t)$s, in particular, provide alternatives to zero rate discounting for those who – like Frank P. Ramsey (1928) and 46 economists out of the 2,160 surveyed by Martin L. Weitzman (2001) – hold a preference for giving genuine weight, in a sense to be made precise, to outcomes in the extremely distant future.

1.1 Slow and Fast Discounting

Consider the following four discount functions:

(3) exponential: $d(t) = e^{-rt}$, where $r = 0.02$

(4) hyperbolic: $d(t) = \left[1 + (\sigma^2/\mu)t\right]^{-\mu^2/\sigma^2}$, where $\mu = .04$ and $\sigma = .029$

(5) quasi-hyperbolic: $d(t) = 1$ for $t = 0$ and $d(t) = b(1+r)^{-(t-1)}$ for $1 \leq t < \infty$ and t an integer, where $b = 0.6$ and $r = .0121$.[5]

(6) time-transformed exponential: $d(t) = e^{-rt^{1/s}}$, where $r = .000314$ and $s = 0.5$.

Equation (3) is the discount function for an exponential with a discount rate of 0.02 and, hence, a present value $[= pv(\infty)]$ of $.02^{-1} = 50$. Equation (4) represents the particular form of the hyperbolic family that Weitzman (2001) labeled 'gamma discounting', with one parameter modified slightly to reduce present value from the 54.8 his parameters imply to 50. Equation (5) is the 'quasi-hyperbolic' function used by David Laibson (1997), again with parameters modified to reduce present value from his implied 60.4 to 50. Daniel Read (2001) has suggested the formulation in equation (6), which is an exponential with the value for time transformed – in this case by squaring it – before being exponentiated. We have again chosen parameters so that $pv(\infty) = 50$.

We will in due course further discuss each of these functions, but our purpose now is simply to observe that although each has a present value of 50 they differ markedly in how rapidly they acquire that present value. Figure 1 illustrates this point by plotting, for each of the discount functions, how its associated present value function rises with time to its asymptote of 50. Note several points: both the hyperbolic and quasi-hyperbolic functions rise more slowly than the exponential in the sense that, for both of them, $pv(t)$ is strictly less than it is for the exponential for all $t > 0$. They can thus, in some sense, be viewed as slower than the exponential with the same present value. Second, what we have labelled the time-transformed exponential is, with the indicated parameters, faster than the exponential. Third, the differences among the procedures translate into major differences in the weight given the far future: while the time transformed exponential has acquired essentially all of its present value within 150 years, and the exponential within 250 years, the hyperbolic has over 6% of its total present value still to be acquired after 500 years. Finally, since the present value functions for the hyperbolic and the quasi-hyperbolic cross, neither can be considered strictly slower than the other. Crossing of Lorenz curves provides a close analogy. Just as the Gini coefficient provides one way to complete the inequality ordering on income distributions generated by Lorenz curves, so, too, will an area based measure allow completion of the ordering of the relative speed of discount functions.

Figure 2 provides a geometrical motivation for the definitions of time horizon and relative speed that we now introduce. The greater the area between the $pv(\infty)$ of a discounting procedure and its $pv(t)$ function the slower, intuitively, it appears to be. We have denoted this area for the exponential in Figure 2 as **A**, and the following expression gives its value:

$$\mathbf{A} = \int_0^\infty [r^{-1} - \int_0^t e^{-rx} dx] \, dt, \tag{7}$$

where the inner integral is, of course, the expression for $pv(t)$ for an exponential, and r^{-1} is $pv(\infty)$ for the exponential (illustrated by the horizontal line at 50). Evaluating the integrals in equation

Figure 1:
Four Discounting Procedures with pv(∞) = 50

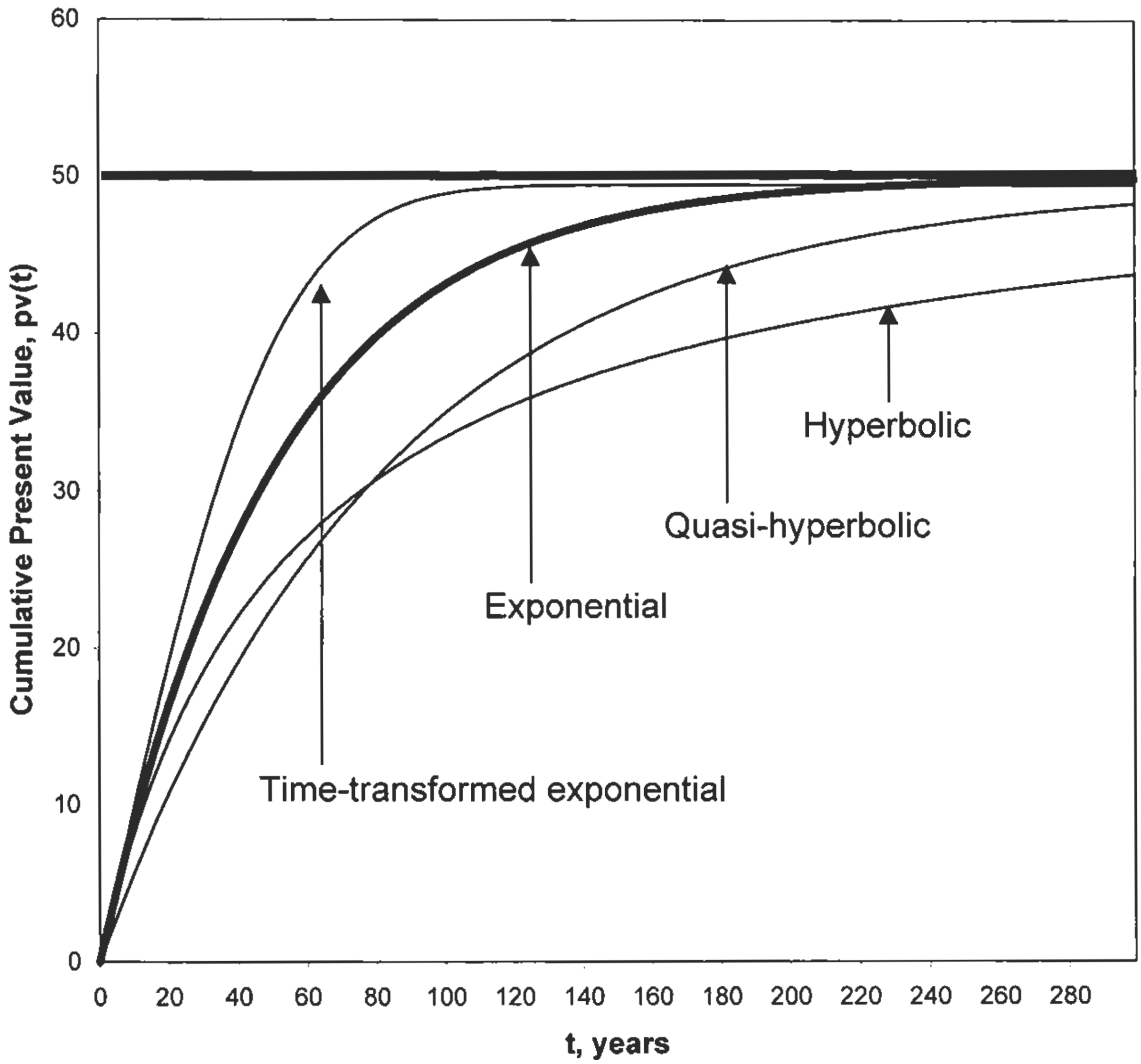
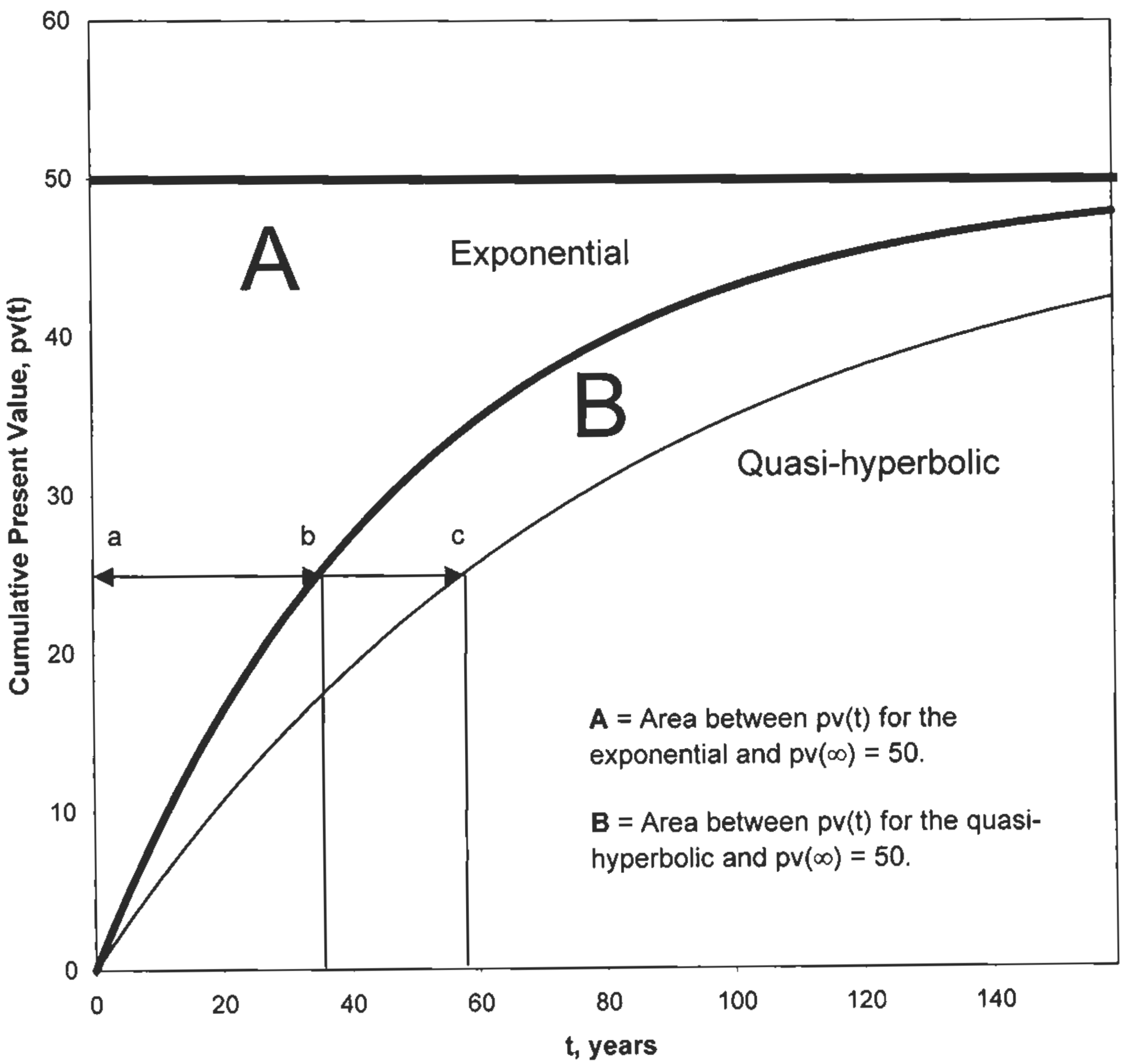


Figure 2:
The Time Horizon and Relative Speed of Discounting Procedures



Note: This figure displays pv(t) for the same exponential and quasi-hyperbolic discounting procedures shown in Figure 1.

(7) gives:

$$\mathbf{A} = r^{-2}. \tag{8}$$

Likewise $\mathbf{B}$ is the area between the quasi-hyperbolic and the line r^{-1} ($= 50$ in Figure 2). The ratio $\mathbf{A}/\mathbf{B}$ ($= r^{-2}\,\mathbf{B}^{-1}$) provides a natural index of the speed of this quasi-hyperbolic relative to that of an exponential with the same present value.

We denote the *relative speed* of a discounting procedure[6], D, to be $\rho(D)$ and the preceding discussion suggests the following definition:

(9) (9a) $\mathbf{A}(D)$ = area between $pv_D(\infty)$ and the exponential with the same present value as D, i.e. the exponential with an r of $pv_D(\infty)^{-1}$;

(9b) $\mathbf{B}(D)$ = area between $pv_D(\infty)$and $pv_D(t)$

$= \int_0^\infty [pv_D(\infty) - pv_D(t)]\; dt,$ if this integral converges; and

(9c) $\rho(D) = \mathbf{A}(D)/\mathbf{B}(D)$.

For an exponential discount function $\rho(D) = 1$. We define D to be *slow* if $\rho(D) < 1$ and *fast* if $\rho(D) > 1$.

We have defined the present value of a discounting procedure in terms of how much it discounts a unit stream of benefits. The more that one discounts the future (as a whole), the less the unit stream will be worth now. It is thus natural to define the *amount* of discounting for a procedure, $\alpha(D)$, to be the inverse of its present value:

$$\alpha(D) = \left[\int_0^\infty d_D(t)dt\right]^{-1}. \tag{10}$$

For example, if D is an exponential procedure with constant discount rate r, then $\alpha(D)$ is simply $r : \alpha(D) = \left[\int_0^\infty e^{-rt}dt\right]^{-1} = r$. This makes sense insofar as we think of higher discount rates as discounting away the future to a greater extent.

Another characteristic of potential interest for a discounting procedure concerns how long a time the procedure takes to build up to, say, half of its ultimate present value. We refer to the

[6]A discounting *procedure* can be defined in terms of its associated discount function, $d_D(t)$, its present value function, $pv_D(t)$, its discount rate function, $r_D(t)$, or in other ways, as discussed in the next section.

time required for accumulation of present value as the *time horizon* of a procedure and introduce several alternative mathematical definitions. The line segments (a,b) and (a,c) in Figure 2 show how many years it takes the exponential and quasi-hyperbolic procedures to accumulate 25 of their present values of 50. The times are, respectively, 35 and 58 years. We define 'median time to accumulation of present value' for a procedure to be the time required for it to accumulate 50% of its present value. We label this $\tau(D)$, which is given by the following expression:

(11) $$\tau(D) = t^* \text{ such that } \frac{\int_0^{t^*} d_D(t)\, dt}{pv(D)} = 0.5.$$

Just as there is a median time to accumulation of present value so, too, can one define a mean time, which can be thought of as how far from time zero, on average, the present value is being accumulated from. We label the mean time to accumulation as $\theta(D)$, given by:

(12) $$\theta(D) = \frac{\int_0^{\infty} t d_D(t)\, dt}{pv(D)}.$$

Although the concept of median time, τ, perhaps conveys more intuitive content than mean time, θ, the latter's analytic tractability more than repays its inclusion.

1.2 The Issue of Convergence

Initially proposed variants of hyperbolic discounting took the form:[7]

(13) $$d_h(t) = 1/(1 + at).$$

The following brief argument shows that $d_h(t)$ has infinite present value. Choose b to be greater than $a+1$. Then, for $t \geq 1$, $1/bt < 1/(1+at) = d_h(t)$. For the harmonic series $(1, \frac{1}{2}, \frac{1}{3}, ..., \frac{1}{t}, ...)$, however, $\lim_{i\to\infty} \sum_{t=1}^{i} 1/t = \infty$ and, hence, $\frac{1}{b} \lim_{i\to\infty} \sum_{t=1}^{i} 1/t = \lim_{i\to\infty} \sum_{t=1}^{i} 1/bt = \infty$. Since $d_h(t) = 1/(1+at)$ is everywhere greater than $1/bt$, it too diverges.[8] To put this slightly differently, an improved

[7]See, for examples, Charles M. Harvey's (1986, 1994) early important work on nonconstant rate discounting and work in the psychology literature of Richard J. Herrnstein (1981). George Loewenstein and Drazen Prelec (1992) discuss and provide an axiomatic foundation for a 'generalized hyperbolic' that places an exponent, k, on the denominator of equation (13). Many parameterizations are possible [e.g. equation (4)] and, in this paper, we refer to the generalized hyperbolic family simply as hyperbolic. The hyperbolic will have a finite present value if and only if $k > 1$.

[8]Basic convergence properties of infinite sums and improper integrals appear in standard texts such as those of Tom M. Apostol (1967) and Walter Rudin (1976).

outcome in each time period by a finite amount X, however large, up to time t^*, also arbitrarily large, would be more than counterbalanced in present value terms, using the discount function $d_h(t)$, by a decrement x, however small, to all outcomes after t^*. It is precisely this property of making outcome changes over any finite time horizon irrelevant compared to tiny but sustained changes in the extremely distant future that leads to objections to exponential discounting with a rate of 0.

The point of this example is simply to illustrate the importance of paying attention to the issue of convergence when selecting a discounting procedure for evaluation of long time horizon investments.[9] Even with a finite but long time horizon present values can be quite sensitive to the discount rate in outer years – hence the importance of explicit consideration of the total amount, $\alpha(D)$, by which a procedure discounts the future.

As the example of equation (13) makes clear, the condition that $\lim_{t\to\infty} d(t) = 0$ fails to guarantee convergence; this condition is, however, necessary. Put somewhat loosely, $d(t)$ must not only go to zero but go to zero sufficiently rapidly if $\sum_{t=0}^{\infty} d(t)$ is to be finite. Implications for $r(t)$ are the opposite: it must go to zero very slowly or not at all.[10] If

$$\lim_{t\to\infty} r(t) = r^*,\ \ r^* > 0, \tag{14}$$

then in the limit $d(t+1) \leq \frac{d(t)}{1+r^*}$, which guarantees convergence. Analogous results hold for continuous time. All exponential discount functions, then, or ones that are ultimately exponential – in the sense that equation (14) holds – will yield finite present values. On the other hand, if for all t greater than some t^*, $r(t) = 0$ [but $d(t) > 0$ still] then the present value will be infinite (e.g. Weitzman, 2001, Table 2).

Having a convergent discounting procedure, however, does not necessarily entail convergence of the integrals defining our concepts of the speed of discounting [equation (9b)] or of the mean time to accumulation of present value [equation (12)]. Equation (12) illustrates the question of

[9]The issue of convergence assumes less importance in modelling phenomena over shorter periods as in the psychological literature or some areas of engineering. Sokolnikoff and Redheffer (1958, p. 176) observe, for example, that "... divergent Fourier series often arise in practice, for example in the theory of Brownian motion, in problems of filtering and noise, or in analyzing the ground return to a radar system. Even when divergent the Fourier series represents the main features of $f(x)$ ...".

[10]To take an example: if $r(t) = r_o e^{-kt}$ then r will decline from an initial level of r_o to 0 too rapidly for the present value of the procedure to be finite.

convergence more clearly since its integrand, $td(t)$, is the product of a function going to zero and a function going to ∞. Stronger conditions on $d(t)$ must, obviously, obtain for this integral to converge than for the integral of $d(t)$ to converge.

We define a discounting procedure to be *strongly convergent* if the integrals both of its related discount function, $d(t)$, and of $td(t)$ converge. [The convergence of $td(t)$ will turn out to be equivalent to convergence of equation (9*b*).] We define a procedure to be *weakly convergent* if $d(t)$ converges but $td(t)$ fails to converge. A weakly convergent procedure has the potentially desirable property of giving infinite weight to the future – in the sense that the average time to accumulation of present value is infinite – while still having a finite present value. Equation (4) provides an example of a weakly convergent procedure.

We have in this Section introduced a number of concepts that we use to characterize discounting procedures and that we will relate to each other later in this paper. Table 1 draws those concepts and their definitions together in one place.

Table 1:

Definitions of Key Concepts

Concept	Defining Expression
1. the *present value* of a discounting procedure, $pv(D)$ or $pv_D(\infty)$	$pv(D) = \int_0^\infty d_D(t)\ dt$
2. the *amount* by which a procedure discounts the future, $\alpha(D)$	$\alpha(D) = pv(D)^{-1}$
3. the *relative speed* with which a procedure discounts the future, $\rho(D)$	$\rho(D) = \left[\alpha(D)^2 \int_0^\infty (pv(D) - pv_D(t))\, dt\right]^{-1}$; D is *slow* if $\rho(D) < 1$ and *fast* if $\rho(D) > 1$
4. the *median time to accumulation* of present value, $\tau(D)$	$\tau(D) = t^*$ such that $\int_0^{t^*} d_D(t)dt = 0.5\ pv(D)$
5. the *mean time to accumulation* of present value, $\theta(D)$	$\theta(D) = \alpha(D) \int_0^\infty td(t)\ dt$
6. *strong convergence* of a discounting procedure	D is strongly convergent if $pv(D)$ and $\theta(D)$, rows 1 and 5 above, are both finite.
7. *weak convergence* of a discounting procedure	D is weakly convergent if $pv(D)$ is finite but $\theta(D)$ is not, i.e. if the procedure has a finite present value but an infinite mean time to accumulation of that present value.

2 Discounting Procedures

A discounting procedure can be defined in multiple ways, each of which leads to a discount function, $d(t)$, that assigns to each future time t a coefficient by which the net outcome at that time can be converted to its present value or the time 0 'equivalent' of the time t outcome.[11] Recall that the requirements on a discount function d, defined for all $t \geq 0$, are that $d(0) = 1$ and that $d(t)$ be nonnegative and weakly decreasing everywhere. For instance, $d(0) = 1$ may be interpreted as the normalization of time 0 as the present. We now digress briefly from our development of results concerning characteristics of discounting procedures (and how they can be aggregated), first, to summarize how different ways of defining a discounting procedure relate to its discount function and, second, to selectively overview (and add to) the menu of discounting procedures currently in the literature.

2.1 Alternative Ways of Defining Discounting Procedures

Table 2 lists five functions that can, in addition to $d(t)$ itself, be used to define a discounting procedure and relates each of these to $d(t)$. For reference the relation to $d(t)$ is given for both discrete and continuous time formulations; as already indicated, however, we develop our results entirely in continuous time. Four of the potential defining functions – the discount rate, discount factor, yield curve, and present value – are familiar, although they often appear in different contexts. We introduce a fifth defining function in addition to these four, which we label the 'associated probability density function' (or *pdf*).

Table 2 states (without proving) the relation between each of the defining functions and $d(t)$, and thus provides a mechanism for transforming any one of them into another. Row 1, for example, shows the relation of $d(t)$ to the discount rate function $r(t)$ by giving $d(t)$ as a function of $r(t)$ and vice versa. Row 2 shows the relation of discount factors – the standard representation of discounting in game theory, for instance – to $d(t)$.

[11] We consider only deterministic, multiplicative discounting procedures that are additively separable over time. See conclusion for a brief discussion of various axiomatizations.

Table 2:

Relations Between the Discount Function and Other Defining Functions

Definitional function	Discrete time	Continuous time
1. discount rate, $r(t)$	$d(t) = \prod_{i=1}^{t}[1+r(i)]^{-1}$	$d(t) = e^{-\int_0^t r(x)dx}$
	$r(t) = \frac{d(t-1)}{d(t)} - 1$	$r(t) = \frac{-d'(t)}{d(t)}$
2. discount factor, $\delta(t)$	$d(t) = \prod_{i=1}^{t}\delta(i)$	$d(t) = e^{\int_0^t \ln \delta(x)dx}$
	$\delta(t) = \frac{d(t)}{d(t-1)}$	$\delta(t) = e^{d'(t)/d(t)}$
3. yield curve, $y(t)$	$d(t) = [1+y(t)]^{-t}$	$d(t) = e^{-y(t)t}$
	$y(t) = [d(t)]^{-1/t} - 1$	$y(t) = -\frac{\ln d(t)}{t}$
4. present value, $pv(t)$	$d(t) = pv(t) - pv(t-1)$	$d(t) = pv'(t)$
	$pv(t) = \sum_{i=0}^{t} d(i)$	$pv(t) = \int_0^t d(x)dx$
5. associated *pdf*, $f(t)$	$d(t) = \sum_{i=t}^{\infty} f(i)$	$d(t) = \int_t^{\infty} f(x)dx$
	$f(t) = d(t) - d(t+1)$	$f(t) = -d'(t)$

Note: our conventions here are the usual ones, namely that the empty product is equal to 1 and the unit stream begins to accrue immediately at time 0. In discrete time, $r(t)$ refers to the discount rate between time $t-1$ and time t, and similarly for $\delta(t)$. Also in discrete time, $pv(t)$ is the present value accumulated up through and inclusive of time t.

Row 3 relates the yield curve from the finance literature to $d(t)$. The relation here is a useful one for interpreting results in the empirical literature on assessment of discount rates since they are frequently reported (implicitly) as yield curves.[12] Row 4 relates $d(t)$ to the cumulative present value function. Finally, Row 5 results from the observation that the inverse cumulative[13] of any *pdf* defined on $[0, \infty)$ will, in fact, be a discount function with a present value equal to the expectation of a random variable with that *pdf*. Thus if f is a *pdf* on $[0, \infty)$, then

$$d_f(t) = 1 - \int_0^t f(x)dx = \int_t^\infty f(x)dx \tag{15}$$

will be a discount function.

The potential utility of the *pdf* representation is that equation (15) allows us to use knowledge of *pdf*s to generate potentially new classes of discount functions. If we start with the exponential *pdf*, $f(x) = re^{-rx}$ then this procedure leads back to the exponential discount function, so in this case equation (15) provides nothing new. Generalizing slightly, however, yields a new single parameter family of discount functions. Start with xe^{-rx} and normalize to get a *pdf*:

$$f(x) = \frac{xe^{-rx}}{\int_0^\infty xe^{-rx}dx} = r^2 xe^{-rx} . \tag{16}$$

The formula for $d(t)$ in row 5 of Table 2 gives, then, the following discount function:

$$d_f(t) = e^{-rt}(1 + rt). \tag{17}$$

Existing discount functions, other than the exponential, are either 2 or 3 parameters, and it may for some purposes be valuable to generalize the exponential but stay with a single parameter, although equation (17) generalizes the exponential only toward fast discounting. We find it more potentially useful in this case to generalize (17) to a two-parameter discounting procedure, which we have labelled the 'augmented exponential' procedure, and which leads directly to the next subsection.

[12] See Burton G. Malkiel (1998) or John H. Cochrane (2001, Chapter 19) for discussions of yield curves and the term structure of interest rates. Maureen L. Cropper, Sema K. Aydede and Paul R. Portney (1994), in an important early paper on discounting of lives saved, report a 16.8% discount rate over 5 years declining to 3.4% over 100 years, a 'yield curve' formulation.

[13] The inverse cumulative of a *pdf* is 1 minus the cumulative.

2.2 Discounting Procedures: Selected Options

The following paragraphs list a number of discounting procedures, some from the literature and some new. Brief background is provided for some. The exponential forms the basis for several procedures that result in a strictly positive asymptotic value of r, and we now turn to these.

The first is the just-mentioned augmented exponential, which takes the form

$$d(t) = e^{-rt}(1 + At). \tag{18}$$

The second class of functions derive from generalizations of the discrete time quasi-hyperbolic discount function introduced as equation (5). This function can be generalized in a variety of ways. First, and most obviously, the point of discontinuity can be made to be at any time, not just 1 (which was in any case already a parameter because of need for the choice of time scale). Second, an option that we do not pursue is to introduce splits at multiple time points [see, for example, Weitzman (2001)]. Third, as Figure 3 illustrates, the generalization to continuous time can be undertaken either by introducing a discontinuity in $r(t)$ – and, hence, in the derivative of $d(t)$ – or by introducing a discontinuity in $d(t)$. We label these the 'split rate quasi-hyperbolic' and the 'split function quasi-hyperbolic'. The latter allows $r(t)$ to remain constant [except for a spike to infinity at the time of the discontinuity in $d(t)$] and has been utilized by Christopher Harris and David Laibson (2000) and, in quite a different way, by Cline (1999). We find any of the discontinuities unaesthetic, but are more comfortable with splits in $r(t)$. In Proposition 3 and the Appendix we explore properties of both generalized quasi-hyperbolic procedures.

Finally, we wish to include procedures that have a limiting discount rate of 0 as time grows indefinitely. The only standard example of this is the hyperbolic that we have already mentioned. The other procedure that falls into this class (at least for some parameter choices) is what we have labelled the time-transformed exponential that was introduced by Read (2001) and used in equation (6):

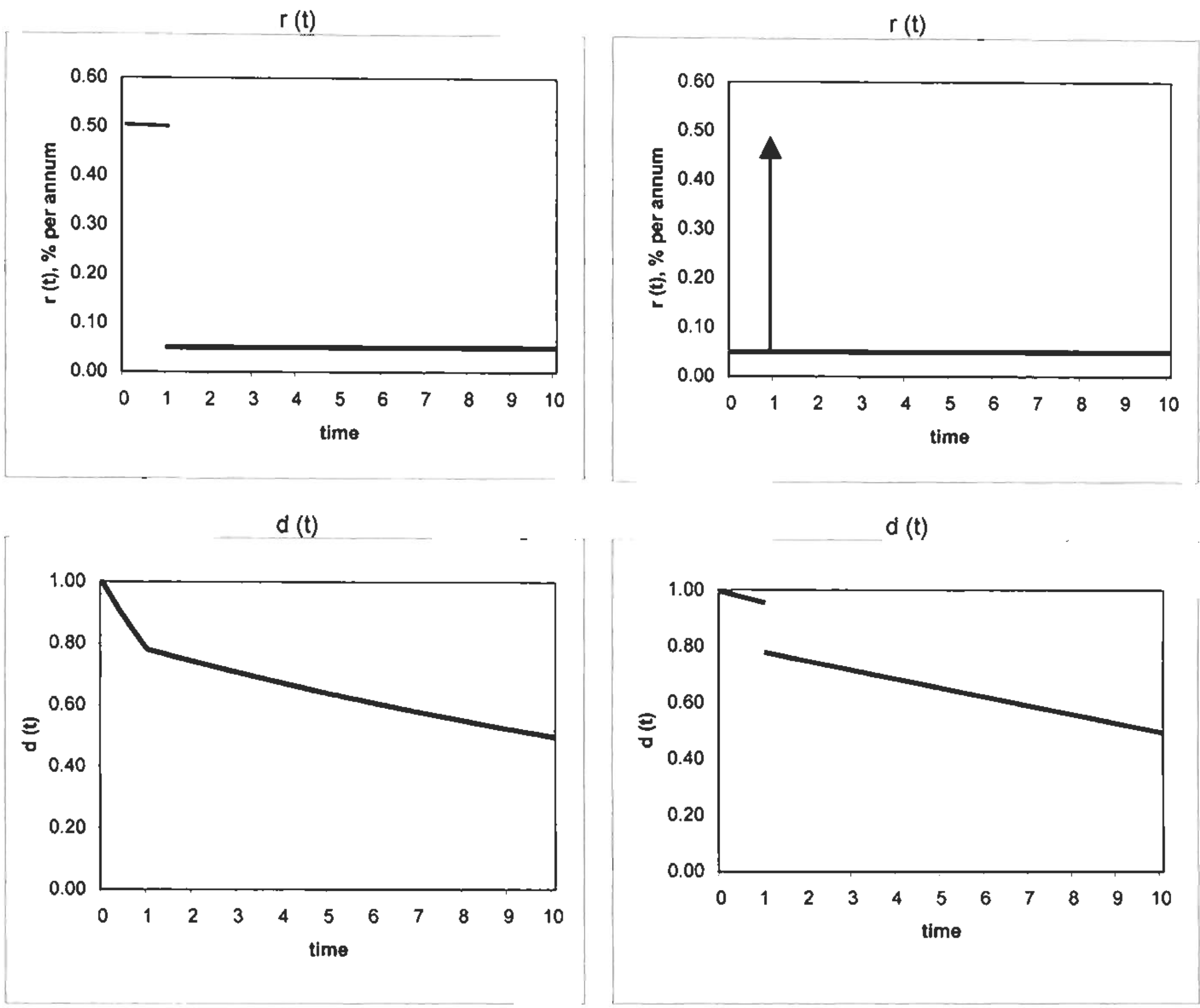
$$d(t) = e^{-rt^{1/s}}. \tag{19}$$

Here the parameter s either expands or contracts time relative to constant rate exponential discounting.[14] Both of these procedures are relatively tractable and intuitive methods of departing from exponential discounting, and both involve only two parameters.

[14] One can, in principle, transform time with a broad range of functions $g(t)$ to get $d(t) = e^{-rg(t)}$. For exam-

Figure 3:
Two Generalizations of the Quasi-hyperbolic Discounting Procedure

Panel A -- Split Rate Quasi-hyperbolic | **Panel B -- Split Function Quasi-hyperbolic**



Note: A quasi-hyperbolic is defined in discrete time, with values of the discount function given by (1, $\beta\delta$, $\beta\delta^2$, $\beta\delta^3$,...), where δ is a discount factor and β is a constant. This figure shows two alternative continuous time formulations that generalize a quasi-hyperbolic with $\beta = 0.78$ and $\delta = 0.95$.
In Panel A r(t) is discontinuous at t = 1, dropping from 0.25 to 0.05, leading to a change in the slope of d(t) at 1 where d(t) = 0.78. In Panel B r(t) = 0.05 for all t except t = 1 where it is infinite. If r(t) is an appropriately formulated Dirac delta function, then d(t) can be made to drop from 0.95 to 0.78 at t = 1. Choice of the unit of measurement for time determines the time of discontinuity which is assumed, without loss of generality, to occur at t = 1 in these examples.

3 The Amount, Speed and Time Horizon of Discounting

Concepts introduced in the preceding sections are not independent of one another, and the initial subsection that follows identifies and proves the central relationships among them. The next subsection then provides closed form expressions for key characteristics of a range of existing and newly proposed discounting procedures. Although many of these characteristics can be expressed simply (given an appropriate parameterization of the procedure), the proofs can be somewhat lengthy and have for the most part been relegated to the Appendix.

3.1 Relations Among Amount, Speed and Time Horizon

Recall that we define the amount of discounting to be

$$\alpha(D) = \left[\int_0^\infty d(t)dt\right]^{-1}, \tag{20}$$

where D is a discounting procedure with associated discount function $d(t)$. A large value for α (we drop the D when there can be no confusion) means that a given procedure discounts later times a lot, i.e. places relatively little weight on the future. As we saw, if D is an exponential procedure with constant discount rate $r > 0$, then $\alpha = \left[\int_0^\infty e^{-rt}dt\right]^{-1} = [1/r]^{-1} = r$.[15] We also defined the present value function $pv_D(t) = \int_0^t d(x)dx$, and from there the relative speed

$$\rho(D) = \frac{\int_0^\infty [pv_{DE}(\infty) - pv_{DE}(t)]\, dt}{\int_0^\infty [pv_D(\infty) - pv_D(t)]\, dt}, \tag{21}$$

assuming the integrals converge, where DE denotes the equivalent constant-rate exponential for any discounting procedure D, i.e. $e^{-\alpha(D)t}$. This ratio is then exactly the ratio of areas illustrated in Figure 2. If $\rho(D) < 1$, then D has a larger area above its present value function than does DE, which is what we have referred to as slow discounting. If $\rho(D) > 1$, then D has a smaller area (pv_D climbs more steeply) and we call it fast.[16]

ple, Roelofsma (1996) discusses Weber's Law from the psychology literature, for which $g(t) = \ln t$. In this case, interestingly, the resulting discounting procedure can be shown to be a member of the hyperbolic family.

[15] Of course, if $r = 0$, then $\int_0^\infty d(t)dt = \infty$ and $\alpha = \frac{1}{\infty} = 0 = r$ still holds.

[16] The diagrammatic representation suggests a link between our notion of relative speed and the standard definitions of stochastic dominance. Indeed, either first- or second-order stochastic dominance implies a slower procedure. One distinction is that our notion, being real-valued, induces a complete order, unlike stochastic dominance. The more fundamental difference is that in the uncertainty setting dominance does not concern preferences (utilities) but rather states of nature, whereas in our setting the dominance comparison is actually between individual discount functions.

Noting that $pv_D(\infty) = \alpha^{-1}$ and evaluating the numerator as $\int_0^\infty \left[pv_D(\infty) - \int_0^t e^{-\alpha x}dx\right] dt = \int_0^\infty \left[\alpha^{-1} - \alpha^{-1}(1 - e^{-\alpha t})\right] dt = \alpha^{-1} \int_0^\infty e^{-\alpha t}dt = \alpha^{-2}$, we get

(22)
$$\rho(D) = \left[\alpha^2 \int_0^\infty \left(\alpha^{-1} - pv_D(t)\right) dt\right]^{-1}.$$

Thus any exponential procedure trivially has a relative speed of 1.

Given a discounting procedure D with discount function $d(t)$, we can define the *normalized procedure* αD by

(23)
$$\alpha d(t) = \alpha(D)d(t)$$

(so the notation is not ambiguous). We refer to this as normalized because $\int_0^\infty \alpha d(t)dt = \alpha(D) \int_0^\infty d(t)dt = \alpha(D)[\alpha(D)]^{-1} = 1$. We will return to the normalized procedure in Section 4 when we discuss aggregation, but for now we point out that it can be interpreted as a probability density function on $[0, \infty)$ and this allows us to define the median and mean time horizons for accumulation of present value. Specifically, let $\tau(D)$ and $\theta(D)$ be the median and mean of this density:

(24)
$$\tau(D) \text{ solves } \int_0^{\tau(D)} \alpha d(t)dt = 0.5 \text{ and}$$

(25)
$$\theta(D) = \int_0^\infty t\alpha d(t)dt = \alpha(D) \int_0^\infty td(t)dt.$$

For the exponential with $d(t) = e^{-rt}$, we have $\alpha = r$ and thus $1 - e^{-r\tau} = 0.5$, so $\tau = (\ln 2)/r$. Similarly, $\theta = r \int_0^\infty te^{-rt}dt = r(1/r^2) = 1/r$.[17]

We may interpret the mean time $\theta(D)$ as the average point in the future at which present value is accumulated, and it should thus be related to the speed of discounting[18]. A natural question to ask is what happens if relative speed is defined as the ratio of mean times (with $\theta(D)$ in the

[17] If $r = 0$, this leads to $\tau = \infty$ and $\theta = \infty$, which is intuitively plausible since it is impossible to accumulate an infinite present value in a finite (expected) amount of time. In general, if $\alpha(\overline{D}) = 0$, we simply define $\tau(\overline{D})$ and $\theta(\overline{D})$ to be ∞.

[18] The observant reader has noticed that we defined a measure of relative speed but not of absolute speed. The most natural measure of absolute speed comes from the area above the normalized present value function, rather than the ratio of areas. Since higher values should represent faster procedures (smaller areas), the absolute speed σ is then the inverse of this area: $\sigma(D) = \left[\alpha(D) \int_0^\infty (pv_D(\infty) - pv_D(t))dt\right]^{-1}$. It can be shown (similarly to Proposition 1) that $\sigma(D)$ is then simply $[\theta(D)]^{-1}$, or equivalently $\alpha(D)\rho(D)$, for any discounting procedure D. For instance,

denominator, so that a large θ corresponds to slow discounting) rather than as a ratio of areas. Our first result says the two notions are equivalent, which will prove useful in calculating $\rho(D)$.

Proposition 1: The relative speed of any discounting procedure D is the ratio of the mean time horizon for the equivalent exponential (DE) to the mean time for D; i.e. $\rho(D) = \theta(DE)/\theta(D)$.

Proof: We see that, starting from the definition, $\rho(D) = \left[\alpha^2 \int_0^\infty \left(\alpha^{-1} - pv_D(t)\right) dt\right]^{-1} = \left[\alpha^2 \int_0^\infty \left(\int_0^\infty d(x)dx - \int_0^t d(x)dx\right) dt\right]^{-1} = \left[\alpha^2 \int_0^\infty \int_t^\infty d(x)dxdt\right]^{-1}$. Changing the order of integration, this is $\left[\alpha^2 \int_0^\infty \int_0^x d(x)dtdx\right]^{-1} = \alpha^{-1}\left[\alpha \int_0^\infty d(x)\left(\int_0^x dt\right) dx\right]^{-1} = \alpha^{-1}\left[\alpha \int_0^\infty xd(x)dx\right]^{-1} = \theta(DE)\left[\theta(D)\right]^{-1}$. ∎

Finally, we wish to relate the relative speed of a discounting procedure to its discount rate. In particular, intuition suggests that a decreasing discount rate yields a procedure that is in some sense slow. It is obvious that such a procedure is slower (formally, in the sense defined in footnote 18) than the constant-rate procedure that starts at the same discount rate and stays there, but on the other hand these two will not have the same total present value. Our second proposition states that, even compared to an exponential *with the same amount of discounting*, any decreasing-rate procedure is indeed slow.

Proposition 2: If the discount rate $r(t)$ is weakly decreasing (resp. increasing), then the corresponding discounting procedure is relatively slow (resp. fast), i.e. $\rho \leq 1$ (resp. $\rho \geq 1$). Furthermore, this result is tight in the sense that if $r(t)$ is weakly decreasing everywhere and strictly decreasing somewhere, then the inequality is strict.

Proof: Let $d(t)$ be the discount function with decreasing rate $r(t)$, and let $\mu = pv_d(\infty) = \int_0^\infty d(t)dt$. We can think of d as the survivor function for a failure density[19], in which case $r(t)$

the exponential with $d(t) = e^{-rt}$ has an absolute speed of r, so in some sense it has the same amount and speed of discounting; this is exactly why the two notions are difficult to disentangle in a constant rate world. Although this measure of absolute speed has some utility and interpretive appeal, it too mixes the ideas of amount and speed of discounting (and of course is not independent of them), so we have not pushed it further.

[19]Failure analysis arises in the study of systems reliability. The primitive in these analyses is the probability of no failures before time t; this is the role played by $d(t)$ in our setting. In reliability studies this is referred to as the survivor or reliability function, and it is the inverse *cdf* of the failure density (i.e. the probability of failure at any given time). The failure rate as a function of time is $-d'/d$, which is thus exactly our discount rate $r(t)$.

decreasing is exactly the definition of a decreasing failure rate (and μ is the mean time to failure). Then for any strictly increasing function f on $[0, \infty)$, $\int_0^\infty f(t)d(t)dt \geq \int_0^\infty f(t)e^{-t/\mu}dt$ by Theorem 4.8 (p. 32) of Barlow and Proschan (1965), with equality only if $d(t) = e^{-t/\mu}$ identically. So let $f(t) = t$: $\int_0^\infty td(t)dt \geq \int_0^\infty te^{-t/\mu}dt = \mu^2 = \left[\int_0^\infty d(t)dt\right]^2$, implying that

$$\rho = \frac{\left[\int_0^\infty d(t)dt\right]^2}{\int_0^\infty td(t)dt} \leq 1,$$

as desired. Equality implies that d must be exponential with amount $\alpha = 1/\mu$. Likewise, $r(t)$ increasing corresponds to an increasing failure rate, and all inequalities are reversed. ∎

We can alternately interpret this conclusion as saying that the exponential is the fastest discounting procedure within the family of those with weakly decreasing discount rates. As we have seen, the exponential discounting procedure $d(t) = e^{-rt}$ also has a total present value equal to its mean time horizon: $pv_d(\infty) = \theta = 1/r$. Proposition 2 allows us to answer the question of whether or not it is the unique discounting procedure with this property.

Corollary: If $d(t)$ is a discount function with monotone discount rate $r(t)$ and satisfying $pv_d(\infty) = \theta(D)$, then d is exponential: $d(t) = e^{-t/\theta}$.

Proof: From the definitions, $pv_d(\infty) = \theta$ means $\int_0^\infty d(t)dt = \frac{\int_0^\infty td(t)dt}{\int_0^\infty d(t)dt}$, i.e. $\left[\int_0^\infty d(t)dt\right]^2 = \int_0^\infty td(t)dt$. But this again corresponds to equality in the proof of Proposition 2, so d must identically equal the equivalent exponential. ∎

3.2 The Amount, Speed and Time Horizons of Key Discounting Procedures

Our main characterization theorem provides, for several important discounting procedures, actual values for each of the various concepts we have defined.

Proposition 3: The discount rate, amount, relative speed, and median and mean time horizons for the following procedures are as stated in Tables 3a and 3b: exponential, augmented exponential, quasi-hyperbolic (of both types), hyperbolic, and time-transformed exponential.

Proof: See Appendix.

Table 3a:

Characteristics of Selected Procedures: $r(t)$ Asymptotically Positive

Characteristic	EXP	AEX	QHR	QHF
1. discount function, $d(t)$	e^{-rt} $(r > 0)$	$e^{-rst}(1 + rs(s-1)t)$ $(r > 0;\ 1 \leq s \leq 2)$	e^{-rt} *if* $t \leq t^*$; $e^{-rt^*-s(t-t^*)}$ *if* $t > t^*$ $(r, t^* \geq 0;\ s > 0)$	e^{-rt} *if* $t \leq t^*$; λe^{-rt} *if* $t > t^*$ $(r, t^* > 0;\ \lambda \in [0,1])$
2. discount rate, $r(t)$	r	$rs\left(1 - \frac{s-1}{1+rs(s-1)t}\right)$	r *if* $t < t^*$; s *if* $t > t^*$	r *if* $t \neq t^*$; ∞ *if* $t = t^*$
3. amount of discounting, α	r	r	$\frac{rs}{\beta r+(1-\beta)s}$; $\beta = e^{-rt^*}$	$\frac{r}{1-(1-\lambda)\beta}$; $\beta = e^{-rt^*}$
4. relative speed, ρ	1	$1 + \frac{(s-1)^2}{2s-1}$ $(1 \leq \rho \leq 4/3)$	$\frac{[\beta r+(1-\beta)s]^2}{\beta r^2+(1-\beta)s^2+\beta t^* rs(r-s)}$ $(0 < \rho < 2)$	$\frac{[1-(1-\lambda)\beta]^2}{1-(1-\lambda)\beta(1+rt^*)}$ $(0 < \rho < 2)$
5. median time, τ	$r^{-1}\ln 2$	$\tau = \frac{\ln(2+2r(s-1)\tau)}{rs}$	$t^* + s^{-1}\ln\frac{2\beta r}{\beta r+(1-\beta)s}$; or $r^{-1}\ln\frac{2s}{s-\beta(r-s)}$ *if* $rt^* > \ln\frac{r+s}{s}$	$r^{-1}\ln\frac{2\lambda}{1-(1-\lambda)\beta}$; or $r^{-1}\ln\frac{2}{1+(1-\lambda)\beta}$ *if* $rt^* > \ln(1+\lambda)$
6. mean time, θ	r^{-1}	$\frac{2s-1}{s^2}r^{-1}$	$\frac{\beta r/s+(1-\beta)s/r+\beta t^*(r-s)}{\beta r+(1-\beta)s}$	$\frac{1-(1-\lambda)\beta(1+rt^*)}{1-(1-\lambda)\beta}r^{-1}$

Note: EXP refers to an exponential discounting procedure; AEX refers to an augmented exponential; QHR refers to the split rate quasi-hyperbolic; and QHF refers to the split function quasi-hyperbolic.

Table 3b:

Characteristics of Selected Procedures: $r(t)$ Potentially Asymptotically 0

Characteristic	HYP	TTE
1. discount function, $d(t)$	$[1+r(1-s)t]^{-(1+\frac{1}{1-s})}$	$\exp(-rt^{1/s})$
	$(r>0;\ s<1)$	$(r,s>0)$
	$(0<s<1$ strongly convergent;	
	$s\leq 0$ weakly convergent)	
2. discount rate, $r(t)$	$r\frac{2-s}{1+r(1-s)t}$	$\frac{r}{s}t^{(1-s)/s}$
3. amount, α	r	$r^s/\Gamma(s+1)$
4. relative speed, ρ	s	$\Gamma(s)\Gamma(s+1)/\Gamma(2s)$
	$(\rho<1)$	$(0<\rho<2)$
5. median time, τ	$r^{-1}(2^{1-s}-1)/(1-s)$	*no closed form*
6. mean time, θ	$(rs)^{-1}$	$r^{-s}\Gamma(2s)/\Gamma(s)$

Note: HYP refers to a hyperbolic discounting procedure; TTE refers to a time-transformed exponential.

Note that several of the procedures listed in Table 3 reduce to a standard constant-rate exponential under certain parameter values, namely the augmented exponential for $s = 1$; the split rate quasi-hyperbolic for $t^* = 0$ or $r = s$; the split function quasi-hyperbolic for $\lambda = 1$; and the time-transformed exponential for $s = 1$. Although we do not allow $s = 1$ explicitly in the hyperbolic procedure, if we take the limit as s approaches 1 (from below) we get an exponential with discount rate r, as expected. This can be shown directly, but it also follows immediately from the discount rate function for the hyperbolic.

We point out that $s > 0$ is not required for the hyperbolic procedure. In fact, $s \leq 0$ is perfectly legitimate; this corresponds to weak convergence in our terminology. Technically, the relative speed as a ratio of areas will be identically 0 for any weakly convergent procedure, but it is useful to allow $s < 0$ in the definition and to continue to refer to this as a speed.

The formulas for the time-transformed exponential make use of the gamma function $\Gamma(y) = \int_0^\infty e^{-x} x^{y-1} dx$. We can calculate, for instance, that if $s = 2$ (which corresponds to compressed time), then $\alpha = r^2/2$, $\rho = 1/3 < 1$, and $\theta = 6/r^2$; this is an example of slow discounting (and more generally $\rho < 1$ exactly when $s > 1$). Likewise, if $s = 1/2$ (expanded time), then $\alpha = \sqrt{4r/\pi}$, $\rho = \pi/2 > 1$, and $\theta = 1/\sqrt{r\pi}$ (fast discounting). The limiting behavior here is also fairly simple (derivations are in the Appendix). On the slow side, as $s \to \infty$, both α and ρ go to 0; on the fast side, as $s \to 0$, $\alpha \to 1$ and $\rho \to 2$ (both τ and θ go to 1/2). Although the median time τ for the time-transformed exponential is not available in closed form, one could easily write down the integral equation that it solves.

Finally, we note that one can interpret the amount, α, for the split rate quasi-hyperbolic as a β-weighted harmonic mean of the two underlying rates r and s, where β is defined in terms of r and the time, t^*, of the discontinuity in $r(t) : \beta = e^{-rt^*}$. For this same procedure, the relative speed ρ will be greater than (less than) 1 exactly as s is greater than (less than) r, which makes intuitive sense.

4 Aggregation of Discounting Procedures

Suppose that we start with a population of individuals each of whom uses some discounting procedure. The individual procedures may differ both in their parameter values and in their actual functional forms. We wish to aggregate these procedures to achieve a social discounting procedure that reflects the preferences of all members of society. Weitzman (2001) pointed to the importance of this question and the nonobvious nature of the response.[20] Our work builds on his, although our conclusion differs.

4.1 Aggregation in General

One obvious aggregation option is to average discount rate functions across individuals. For example, if person A uses a standard constant-rate procedure with $d(t) = e^{-rt}$ (i.e. a rate of r), and person B uses $d(t) = e^{-st}$, then this method would yield a social discounting procedure characterized by a constant rate equal to $\frac{1}{2}(r+s)$. One advantage of this process is that even if one of the individuals uses nonconvergent discounting (e.g. $s = 0$ above), the aggregate social procedure may well be convergent (a notion of robustness). As Weitzman emphasized, averaging rates has a down side: when society decides how to trade off between two given time points in the future (effectively measured by the discount rate), it counts everyone's opinion on that question equally, even those who do not care much about the future. Thus while a procedure that averages rates is able to successfully aggregate *amounts* of discounting, it does less well on the shape over time that discounting should take.

Another natural option is simply to average the discount functions of individuals. In the example with two exponential discounting individuals, this would lead to a social discount function of $\frac{1}{2}(e^{-rt} + e^{-st})$. This aggregate procedure has a discount rate that starts at $\frac{1}{2}(r+s)$ for $t = 0$, and then declines over time to the minimum of r and s. To redress the inadequacy of aggregation by averaging rates, Weitzman (2001) proposed averaging discount functions and interprets this process as valuing a dollar at time t according to any particular discounting procedure weighted by the "probability of correctness" of that procedure. In our example, this probability is 0.5 for each of the two individual procedures in the domain and is 0 for all others.

One immediate concern about such an aggregating process is that if any individual discounting procedure is nonconvergent, then the social procedure will also be, no matter how large the society

[20] Amartya Sen (1982) overviews an extensive earlier literature that addressed the circumstances under which an exponential social discounting procedure should have a rate less than the average of the rates of the individual procedures (also exponential), from which it is being aggregated. More recently, Harris and Laibson (2001) have considered aggregation across different values of the time of discontinuity in the quasi-hyperbolic discounting procedure shown in Figure 3, Panel B.

(assuming it is finite). This can be avoided if we rule out nonconvergent procedures in the first place, but it speaks to a larger concern. Basically, under this process, individuals are weighted equally (in terms of either discount functions or discount rates) at the beginning of time, after which those who discount less are increasingly favored. This process yields social discounting procedures whose discount rate functions are both declining over time (as patient individuals are increasingly favored) *and* that are low overall (since they start at the social average and go down). So the shape of the discount function faithfully reflects a weighted average of the individual shapes (weighted in a reasonable manner by their own sense of relative time preference), but it skews the amount of discounting toward those who discount less (in the extreme case, yielding an aggregate α of 0 if any individual has an α of 0).[21]

We seek an aggregating process that avoids the problems associated with simple averaging of either discount rates or functions. That is, we feel an aggregation process should satisfy two criteria:

(i) the aggregate procedure should discount the future by an amount that is the average of the individual amounts (as averaging the discount rates does)[22]; and

(ii) the aggregate procedure's discount rates in the future should place greater weight on individuals who value the future more highly (as averaging the discount functions does).

Both criteria can be met by averaging the normalized discount functions for each individual (recall that the normalized function is $\alpha d(t)$ and has a total present value of 1). We then divide by the value at 0 of this average function in order to un-normalize and recover a valid discount function (with $d(0) = 1$, as is necessary). We label this the average normalized discount function (ANDF) aggregation process. The ANDF process results in a shape equal to the average shape, and we will prove that it has an amount exactly equal to the average amount $\overline{\alpha}$ (and a mean time horizon equal to the average mean time $\overline{\theta}$). In our running example, the normalized individual

[21] Joseph Lipscomb (1989) and Dorte Gyrd-Hansen and Jes Søgaard (1998) have proposed an ingenious 2-step process in which each individual i discounts back (at i's personal rate) to the time $t_i > 0$ that an initial event (e.g. an illness or a treatment) occurs, and that a social rate is then applied to bring each individual's present value at the time of the event back to a single present value at time 0. As currently formulated, however, this process would (like the averaging of discount functions) favor individuals with low discount rates.

[22] Note that this criterion refers to aggregation of individual procedures into a social one. It has less relevance for generating an expected value when there is underlying uncertainty in interest rates, e.g. the situation considered in Weitzman (1998) and in Richard Newell and William Pizer (2000).

discount functions are re^{-rt} and se^{-st}, respectively, so the aggregate social procedure is

$$d_{\overline{D}}(t) = \frac{re^{-rt} + se^{-st}}{r+s}. \tag{26}$$

Here $\alpha(\overline{D}) = (r+s)/2$ and $\theta(\overline{D}) = (r^{-1}+s^{-1})/2$. Figure 4 illustrates the $d(t)$ and $r(t)$ resulting from each of the three ways of aggregating two exponentials, in this case with $r = 0.02$ and $s = 0.20$. For the ANDF process applied to these two exponentials, $\alpha(\overline{D}) = 0.11$, $\theta(\overline{D}) = 27.5$, and $\rho(\overline{D}) = 0.33$ (so it is slow).

One interpretation is that the ANDF gives each individual a total weight of 1 to spread across the future however he or she likes (with the understanding that this will be used to perform cost-benefit analyses, for instance, so that they will in fact use their true individual discount function to determine how to distribute this weight). We average those shapes and then un-normalize, which works out in just such a way as to retrieve the average amount of discounting. This separation of amount and shape ensures that there is no problem with convergence and that more patient members of society are favored at later times relative to earlier times, but not overall.

Formally, assume that we have a collection of individuals parameterized by $x \in X$ (possibly multivariate), with a probability density across parameters of $f(x)$, so that in particular $\int_X f(x)dx = 1$. Then if individual x uses a discount function $d(t;x)$ with associated amount $\alpha(x)$, we define the ANDF aggregate procedure $\overline{D}$ by its discount function as follows:[23]

$$d_{\overline{D}}(t) = \frac{\int_X \alpha(x)d(t;x)f(x)dx}{\int_X \alpha(x)f(x)dx}. \tag{27}$$

Because $\int_X \alpha(x)d(0;x)f(x)dx = \int_X \alpha(x)(1)f(x)dx = \int_X \alpha(x)f(x)dx$, $d_{\overline{D}}(0) = 1$ as required.

To each individual discount function $d(t;x)$ there corresponds a discount rate function $r(t;x)$ satisfying $\dot{d}(t;x) = -r(t;x)d(t;x)$ (as in Table 2), where the superscript dot denotes a time derivative. For the aggregate procedure $\overline{D}$, since the $d(t;x)$ in the numerator is the only term involving time,

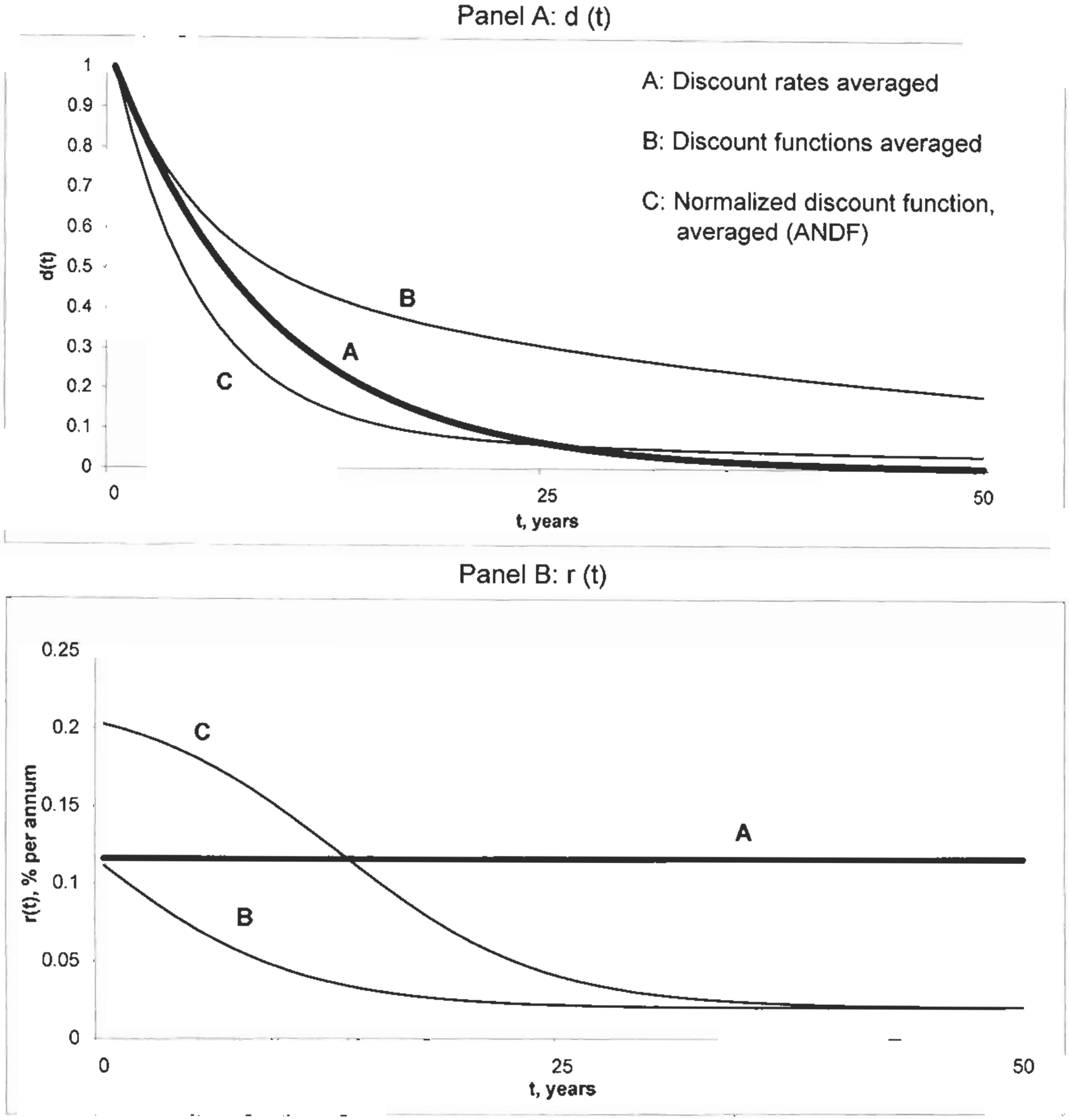
$$\dot{d}_{\overline{D}}(t) = \frac{-\int_X \alpha(x)r(t;x)d(t;x)f(x)dx}{\int_X \alpha(x)f(x)dx}, \tag{28}$$

and therefore

$$r_{\overline{D}}(t) = \frac{-\dot{d}_{\overline{D}}(t)}{d_{\overline{D}}(t)} = \frac{\int_X \alpha(x)r(t;x)d(t;x)f(x)dx}{\int_X \alpha(x)d(t;x)f(x)dx}. \tag{29}$$

[23] We require only that $\int_X \alpha(x)f(x)dx > 0$, i.e. that at least some nonzero fraction of the population uses convergent discounting procedures. If $\int_X \alpha(x)f(x)dx = 0$, then there is no need to normalize (essentially all individuals already have the same α, namely $\alpha = 0$), so we define $d_{\overline{D}}(t)$ as simply $\int_X d(t;x)f(x)dx$.

Figure 4:
Three Procedures for Aggregating Individual Discount Functions



Notes: This figure shows the functions d(t) and r(t) that result from different ways of aggregating two discount functions: $d_1(t) = e^{-.2t}$ and $d_2(t) = e^{-.02t}$.

We now prove three core results that relate the characteristics of the ANDF aggregation procedure to the corresponding characteristics of the individual procedures that were aggregated. Let $Y = \{x \in X | \alpha(x) = 0\}$, i.e. Y is the set of parameters that correspond to nonconvergent discounting procedures.

Proposition 4: If $\int_Y f(x)dx = 0$, then, for the ANDF process defined above, the amount $\alpha(\overline{D})$ and mean time horizon $\theta(\overline{D})$ of the aggregate procedure $\overline{D}$ are the average amount $\overline{\alpha}$ and the average mean time $\overline{\theta}$ respectively.[24]

Proof: Since $\alpha(x) = 0$ for all $x \in Y$, anytime the integrand involves $\alpha(x)$ we can switch the domain of integration between X and $X \backslash Y = X - Y$ as we wish. We first verify that $\int_0^\infty \int_X \alpha(x)d(t;x)f(x)dxdt = \int_0^\infty \int_{X\backslash Y} \alpha(x)d(t;x)f(x)dxdt = \int_{X\backslash Y} \alpha(x)f(x)\left[\int_0^\infty d(t;x)dt\right]dx = \int_{X\backslash Y} \alpha(x)f(x)\frac{1}{\alpha(x)}dx$ ($\alpha(x) > 0$ on $X\backslash Y$ so $\frac{1}{\alpha(x)}$ is well-behaved) $= \int_{X\backslash Y} f(x)dx = \int_X f(x)dx - \int_Y f(x)dx = 1 - 0 = 1$, which makes sense since this was an average normalized function. Now note that the denominator of the aggregate function $d_{\overline{D}}$ is constant in t, and thus, using the calculation we just made, the aggregate amount is

$$\alpha(\overline{D}) = \left[\int_0^\infty d_{\overline{D}}(t)dt\right]^{-1} = \left[\frac{1}{\int_X \alpha(x)f(x)dx}\right]^{-1} = \int_X \alpha(x)f(x)dx,$$

which is precisely the formula for the average amount $\overline{\alpha}$ of discounting across the population, as desired.

For the mean time horizon, we compute θ for the aggregate procedure as

$$\begin{aligned}
\theta(\overline{D}) &= \alpha(\overline{D}) \int_0^\infty t d_{\overline{D}}(t)dt \\
&= \alpha(\overline{D}) \int_0^\infty t \left[\frac{\int_X \alpha(x)d(t;x)f(x)dx}{\alpha(\overline{D})}\right] dt \\
&= \int_0^\infty \int_X t\alpha(x)d(t;x)f(x)dxdt \\
&= \int_X \left[\alpha(x) \int_0^\infty td(t;x)dt\right] f(x)dx \\
&= \int_X \theta(x)f(x)dx \\
&= \overline{\theta}. \quad \blacksquare
\end{aligned}$$

[24] Thus the bar above a particular characteristic denotes expectation with respect to the density f. Exactly because of the proposition, this will not prove to be confusing terminology.

In general, if $\int_Y f(x)dx > 0$, $\alpha(\overline{D})$ will simply be the average over all strictly positive α in the population: formally, $\alpha(\overline{D}) = \int_{X \setminus Y} \alpha(x) f(x) dx$ (with an analogous outcome for θ). The intuitive interpretation is that any individual who chooses $\alpha = 0$ (i.e. a nonconvergent procedure), effectively suggesting an infinite present value, ends up spreading his/her normalized weight so thinly over time that it has no effect at all on the aggregate[25]. We turn next to the relative speed of the aggregate procedure.

Proposition 5: If the amount $\alpha(x)$ and the mean time $\theta(x)$ are negatively covariant within a population X, then the relative speed resulting from the ANDF process is lower than the average relative speed, i.e. $\rho(\overline{D}) \leq \overline{\rho}$, with equality only if all individuals have the same amount of discounting α and all individuals have the same relative speed ρ.

Proof: By Propositions 1 and 4, $\rho(\overline{D}) = \theta(\overline{D}E)/\theta(\overline{D}) = \left(\overline{\alpha}\overline{\theta}\right)^{-1}$, while $\overline{\rho} = \int_X \rho(x) f(x) dx = \int_X \left(\alpha(x)\theta(x)\right)^{-1} f(x) dx = \overline{(\alpha\theta)^{-1}}$, where the bar continues to denote expectation with respect to f. But since $(\cdot)^{-1}$ (i.e. taking inverses) is a convex function on $\mathbb{R}$, Jensen's inequality implies that the inverse of the average is weakly less than the average of the inverses, i.e. $\left(\overline{\alpha\theta}\right)^{-1} \leq \overline{(\alpha\theta)^{-1}}$. Now $cov(\alpha, \theta) = \overline{\alpha\theta} - \overline{\alpha}\overline{\theta}$ by definition, and this is negative by assumption. Hence $\overline{\alpha\theta} \leq \overline{\alpha}\overline{\theta}$, so $\rho(\overline{D}) = \left(\overline{\alpha}\overline{\theta}\right)^{-1} \leq \left(\overline{\alpha\theta}\right)^{-1} \leq \overline{(\alpha\theta)^{-1}} = \overline{\rho}$, and we're done. Furthermore, since $(\cdot)^{-1}$ is in fact strictly convex, the inequality is strict unless both $cov(\alpha, \theta) = 0$ *and* $(\alpha\theta)^{-1}$ is constant across the population. But these cannot hold simultaneously unless α and θ are themselves constant, which is equivalent to α and ρ being constant. ∎

In general, we do expect α and θ to have negative covariance, since placing less total present value on the future often implies that that present value is reached more quickly. This tendency is confirmed by looking at the expressions in Table 3, but it need not hold for every possible choice of distributions f over the underlying parameters. Thus we expect an aggregate procedure to be slower overall than its components, but it is possible for it to be faster under certain circumstances.[26] In the special case of constant relative speed, however, we can rule out this possibility.

[25] It is possible to construct an aggregate discounting procedure whose amount is *always* equal to the average amount, inclusive of x such that $\alpha(x) = 0$. In this case the shape of the aggregate function is identical to the aggregate as defined, so the larger present value means that $pv(t)$ does not converge to $pv(\infty)$, and thus of course $\theta(\overline{D}) = \infty$ and $\rho(\overline{D}) = 0$. Details are available upon request.

[26] Hara and Kuzmics (2002) prove a conceptually similar result in the context of risk aversion. Specifically, they

Corollary: If $\rho(x)$ is constant across x, then $\rho(\overline{D}) \leq \overline{\rho}$, with equality only if $\alpha(x)$ is also constant.

Proof: As always (Proposition 1), $\rho = (\alpha\theta)^{-1}$, so $\rho(x)$ constant implies $\alpha(x)\theta(x) = C$ for all x. Then we can compute $cov(\alpha, \theta) = cov(\alpha, C/\alpha) = \overline{\alpha(C/\alpha)} - \overline{\alpha}\overline{(C/\alpha)} = C - \overline{\alpha}C\overline{\alpha^{-1}} = C\left(1 - \overline{\alpha}\overline{\alpha^{-1}}\right)$. But Jensen's inequality once again implies that $(\overline{\alpha})^{-1} \leq \overline{\alpha^{-1}}$, or $1 \leq \overline{\alpha}\overline{\alpha^{-1}}$. Hence $cov(\alpha, \theta) \leq 0$ and Proposition 5 applies. The inequality is clearly strict unless α is constant across the population. ■

We turn now to the relationship between the aggregate rate $r_{\overline{D}}(t)$ and the individual discount rates $r(t;x)$. In particular, one variable of interest is the limiting discount rate $r^*(x) = \lim_{t\to\infty} r(t;x)$, if it exists, which gives the asymptotic discount rate for individual x. For any potential social limiting rate $r \geq 0$, let $A(r) = \{x \in X \text{ s.t. } \alpha(x) > 0,\ r^*(x) \text{ exists, and } r^*(x) \leq r\}$; this is the set of individuals who use convergent procedures and whose limit is no larger than r. Finally, we define $r^*_{\min} = \inf\left\langle r \,\middle|\, \int_{A(r)} f(x)dx > 0 \right\rangle$. This is the lowest rate r such that at least some nonzero fraction of the population has a limiting rate no higher than r. Weitzman (1998) showed that if the social discount function is constructed by simple averaging of a finite number of individual discount functions, then $r^*_{\min}$ is exactly the asymptotic discount rate for the aggregate. Our final proposition states that the same is true, without assuming a finite number of individuals, for the ANDF process, i.e. that in the limit the social discount rate is in some sense the smallest of any across the population (where, as always for the ANDF process, anyone using a nonconvergent procedure does not affect the outcome).

Proposition 6: The asymptotic social discount rate for the ANDF aggregation process is given by: $\lim_{t\to\infty} r_{\overline{D}}(t) = r^*_{\min}$.[27]

Proof: See appendix.

4.2 Aggregation of Exponential Procedures

The aggregation processes defined above can be carried out no matter what the underlying discounting procedures are. However, if we hope for simple closed-form outcomes, we will need to

show that the representative consumer exhibits strictly decreasing relative risk aversion, ranging from that of the most risk averse individual to that of the least risk averse as the aggregate consumption level increases.

[27] We follow Weitzman (1998) in our notation.

make some specific assumptions. For the remainder of this section, we assume that each individual i uses a constant-rate procedure given by $d(t) = e^{-r_i t}$.[28] This not only simplifies the analysis, but also allows us to compute empirical results from the data collected by Weitzman (2001), who surveyed 2160 economists and asked for a single discount rate from each.

We first observe that when aggregating exponentials, all of which have $\rho(x) = 1$, the corollary to Proposition 5 applies, so the relative speed of the ANDF aggregate is [weakly] less than 1. The aggregate procedure therefore exhibits slow discounting. Furthermore, the aggregate speed is strictly less than 1 unless all individuals use the same constant discount rate, i.e. are identical.

We begin by considering discrete aggregation of exponentials, i.e. the case when there is a finite population. Let the number of individuals be n, with respective constant rates $r_1, ..., r_n$ (possibly with multiplicity, of course). We assume $r_i > 0$ for all i (or equivalently that there are n individuals with $r_i > 0$ and the rest can be ignored; see comment after Proposition 4). In this case the ANDF aggregate procedure has

$$\overline{\alpha} = \frac{1}{n}\sum_{i=1}^{n} r_i, \tag{30}$$

$$\overline{\theta} = \frac{1}{n}\sum_{i=1}^{n} r_i^{-1}, \text{ and} \tag{31}$$

$$\rho(\overline{D}) = \frac{n^2}{\left(\sum_{i=1}^{n} r_i\right)\left(\sum_{i=1}^{n} r_i^{-1}\right)}. \tag{32}$$

These results follow directly from Proposition 4, given that individual i's amount α_i is r_i and i's mean time horizon θ_i is r_i^{-1}. Then, using Proposition 1, $\rho(\overline{D})$ can be computed as $(\overline{\alpha}\overline{\theta})^{-1} = \left[\frac{1}{n^2}\left(\sum_{i=1}^{n} r_i\right)\left(\sum_{i=1}^{n} r_i^{-1}\right)\right]^{-1}$, as stated.

We next study continuous populations by considering two different densities f on the underlying distribution of discount rates. First is the gamma distribution, for $x > 0$:

$$f(x) = \frac{a^b}{\Gamma(b)} e^{-ax} x^{b-1}, \tag{33}$$

with $a, b > 0$. This has mean $\mu = b/a$ and variance $\sigma^2 = b/a^2$. Such a distribution fits the data well (although it has the disadvantage of putting no weight at 0, whereas almost 2.5% of

[28] We have also found specific closed-form solutions when aggregating the augmented exponential procedure, but the calculations are not particularly illuminating. Details are available from the authors upon request.

the respondents chose discount rates at or below 0). Weitzman, aggregating by averaging of the $d(t)$s, showed that the gamma distribution on parameters leads to an aggregate procedure in the hyperbolic family:

$$d_\gamma(t) = \left[\frac{1}{1+t/a}\right]^b. \tag{34}$$

The ANDF aggregation process, with the same gamma distribution for individual discount rates, leads to a social discount function (see Appendix for proof) given by

$$d_{\overline{D}}(t) = \left[\frac{1}{1+t/a}\right]^{1+b}. \tag{35}$$

So it is also hyperbolic, but with an exponent that is greater by 1 than for $d_\gamma(t)$, which helps with many of the convergence issues. Using the formulas in Table 3, we can compute $\overline{\alpha} = b/a$ and $\rho(\overline{D}) = 1 - 1/b$. In terms of the mean μ and variance σ^2 of the density f, we get

$$\overline{\alpha} = \mu \text{ and } \rho(\overline{D}) = 1 - \frac{\sigma^2}{\mu^2} \tag{36}$$

whereas Weitzman's aggregate – given in his parameterization by equation 4 – has

$$\alpha_\gamma = \mu - \frac{\sigma^2}{\mu} \text{ and } \rho_\gamma = 1 - \frac{\sigma^2}{\mu^2 - \sigma^2}. \tag{37}$$

These values imply, among other things, that even after ruling out zero discount rates, the aggregate function $d_\gamma(t)$ fails to converge when $\sigma \geq \mu$. Fortunately, this is not the case for Weitzman's sample (where σ is roughly 3% and μ is roughly 4%), but of course there is no inherent reason why it should turn out one way or the other. If the ANDF aggregation process (which is always convergent) is used when $\sigma \geq \mu$, then the aggregate discounting procedure $\overline{D}$ is convergent but only weakly so; in comparison, $d_\gamma(t)$ is weakly convergent exactly when $\sigma < \mu$ but $\sqrt{2}\sigma \geq \mu$, as occurs in the data.

We wish to also consider an underlying distribution on parameters that is strictly positive at 0, so we adopt the form of the augmented exponential (discussed above as a possible discounting procedure). Note that here it is playing a very different role, and in fact reasonable parameter values in this context would cause it to be increasing at first and would thus be impermissible for a discount function. Recall that the formulation (normalized here to be a probability measure) was

$$f(x) = \frac{a^2}{a+b} e^{-ax}(1+bx). \tag{38}$$

In this case the resulting aggregate procedure is

$$d_{\overline{D}}(t) = \frac{a}{a+2b}\left[\frac{1}{1+t/a}\right]^2 + \frac{2b}{a+2b}\left[\frac{1}{1+t/a}\right]^3, \tag{39}$$

which may be interpreted as a weighted average of hyperbolic procedures. Here we find that $\overline{\alpha} = \frac{a+2b}{a(a+b)}$ and $\rho(\overline{D}) = 0$ (again, see Appendix for all computations). So this is another example of a weakly convergent procedure, in this case due to the members of society who do not discount the future at all (i.e. due to $f(0) > 0$ in the probability density).

5 Conclusion

This paper proceeds from the observation that economists in a range of specialized fields use non-exponential discounting functions. Within the exponential framework, alternative discounting procedures align along the single dimension of discount rate. This simplicity, along with the intuitive appeal of the axiomatic formulation of constant rate discounting and the analytic tractability of the procedure, have ensured its dominance until recently.[29] This paper develops a conceptual framework for nonconstant rate discounting that arrays procedures along the two dimensions of amount and speed of discounting, thereby facilitating systematic comparison of procedures and enhancing their tractability.

A question that is antecedent to discounting in the context of assessing long-term investments, addressed by Kenneth J. Arrow (1999), concerns the extent to which decisions made today will have their influence attenuated (or eliminated) by compensatory decisions of subsequent generations. This issue is very real, and arises also in the design of other sequential decision procedures, such as foreign aid programs and transfers across levels of government. It has been not our purpose in this paper to address that issue but, more simply, to improve the tools available in

[29] The axiomatic foundation for constant rate discounting goes back to Tjalling C. Koopmans (1960) who introduced a 'stationarity' axiom concerning preferences over time streams of outcomes and established plausible conditions that require discounting to be exponential. A number of authors have explored weaker axiom systems that allow representation of intertemporal preferences by a discounting procedure that is not necessarily constant rate - see Dean T. Jamison (1969), Peter C. Fishburn and Ariel Rubinstein (1982), Han Bleichrodt and Amiram Gafni (1996) and Han Bleichrodt and Magnus Johannesson (2001). The simple existence of an axiomatic foundation, therefore, is no argument in favor of constant- over variable-rate discounting; the question is one of assessing the descriptive, normative, and tractability consequences of adding the strong stationarity axiom.

circumstances where discounting is being used. Similarly, although we acknowledge the existence and practical relevance of time-inconsistency issues with regard to variable rate discounting – initially addressed by Robert H. Strotz (1956) – our focus has been on a different facet of the problem.

Key results of the paper include:

(i) precise formulations for concepts of 'amount', 'speed', and 'time horizon' of discounting procedures;

(ii) proofs of key relations among the concepts of amount, speed, and time horizon;

(iii) descriptions of a range of existing and new discounting procedures and provision of closed form characterizations relating amount, speed, and time horizon to their underlying parameters;

(iv) identification of inadequacies in existing approaches to aggregating individual discounting procedures into a socially representative one and formulation of an alternative process – averaging of normalized discount functions or ANDF – that overcomes these inadequacies; and

(v) proof that under reasonable assumptions the ANDF aggregate procedure will be slower than the average of the speeds of the individual procedures.

Frederick, Loewenstein and O'Donoghue (2002) and Gretchen B. Chapman (2003) provide valuable and up-to-date compilations of the empirical literature on time preference and discounting. Newell and Pizer (2000) propose an altogether different empirical approach. Transforming the empirical literature into discounting procedures for policy application will require two additional steps. First, to the extent practical, data underlying the reported literature will need to be characterized in terms of estimates of the amount and speed of individual discounting. Second, the ANDF aggregation algorithm can be used (through our Proposition 4) to generate candidate social discounting procedures. We feel that the theoretical approach to discounting that we propose both undermines many of the practical objections to expanded use of nonconstant rate procedures and provides a needed framework for integrating and comparing results in the existing literature.

6 Appendix

This appendix conveys, first, the proof of Proposition 3 from Section 3.2; second, the proof of Proposition 6 from Section 4.1; and, third, derivations concerning the aggregation of exponentials in Section 4.2.

6.1 Proof of Proposition 3

Proposition 3 provides formulas for discount rates, amounts, relative speeds, median time horizons and mean time horizons for six important discounting procedures. Tables 3a and 3b convey these results, which are proved below.

6.1.1 Exponential

See the text of Section 3.1.

6.1.2 Augmented exponential

For $d(t) = e^{-rst}(1 + rs(s-1)t)$, we compute

$$r(t) = -\frac{d'(t)}{d(t)} = \frac{-e^{-rst}[rs(s-1) - rs(1 + rs(s-1)t)]}{e^{-rst}(1 + rs(s-1)t)} = rs\left(1 - \frac{s-1}{1 + rs(s-1)t}\right)$$

and

$$\int_0^\infty d(t)dt = \int_0^\infty e^{-rst}dt + rs(s-1)\int_0^\infty te^{-rst}dt = \frac{1}{rs} + \frac{rs(s-1)}{(rs)^2} = \frac{1}{r},$$

from which $\alpha = r$ as claimed. Meanwhile,

$$\int_0^\infty td(t)dt = \frac{1}{(rs)^2} + \frac{2rs(s-1)}{(rs)^3} = \frac{2s-1}{(rs)^2}$$

and θ follows directly from the definition. Similarly,

$$\rho = \left[\alpha^2 \int_0^\infty td(t)dt\right]^{-1} = \frac{s^2}{2s-1} = 1 + \frac{(s-1)^2}{2s-1}.$$

Finally,

$$pv(t) = \int_0^t d(x)dx = \frac{1}{r}\left(1 - e^{-rst} - r(s-1)te^{-rst}\right)$$

and thus $pv(t) = pv(\infty)/2 = 1/2r$ when $e^{-rst}(1 + r(s-1)t) = 1/2$. Taking logarithms and inverting, we get that τ must satisfy $rs\tau - \ln(1 + r(s-1)\tau) = \ln 2$, and the stated result follows. We cannot solve explicitly for τ, but it is easy to see that there is a unique value of τ that satisfies the equation (for instance, the LHS is 0 at 0 and has a boundedly positive derivative).

6.1.3 Split rate quasi-hyperbolic

If $d(t) = e^{-rt}$ for $t \le t^*$ and $d(t) = \beta e^{-s(t-t^*)}$ for $t > t^*$ (where $\beta = e^{-rt^*}$), then the discount rate is immediate (it is in fact the defining characteristic for this procedure), though it is undefined at $t = t^*$ where d is not differentiable[30]. For the amount,

$$\int_0^\infty d(t)dt = \int_0^{t^*} e^{-rt}dt + \beta \int_{t^*}^\infty e^{-s(t-t^*)}dt = (1-\beta)\frac{1}{r} + \beta\frac{1}{s},$$

from which it is clear that the inverse is indeed a [weighted] harmonic mean:

$$\alpha = \frac{rs}{\beta r + (1-\beta)s}.$$

For the speed, we calculate

$$\begin{aligned}
\int_0^\infty td(t)dt &= \int_0^{t^*} te^{-rt}dt + \beta \int_{t^*}^\infty te^{-s(t-t^*)}dt \\
&= \left[-\frac{1}{r}te^{-rt} - \frac{1}{r^2}e^{-rt}\right]\Bigg|_0^{t^*} + \beta \int_0^\infty (t+t^*)e^{-st}dt \\
&= \frac{1}{r}(-t^*\beta - \frac{\beta}{r} + 0 + \frac{1}{r}) + \beta \int_0^\infty te^{-st}dt + \beta t^* \int_0^\infty e^{-st}dt \\
&= \frac{1}{r}((1-\beta)\frac{1}{r} - t^*\beta) + \beta\frac{1}{s^2} + \beta t^*\frac{1}{s} \\
&= \frac{1}{rs}\left[(1-\beta)\frac{s}{r} - t^*\beta s + \beta\frac{r}{s} + \beta t^* r\right]
\end{aligned}$$

so that the relative speed is the inverse of

$$\alpha^2 \int_0^\infty td(t)dt = \frac{\beta r^2 + (1-\beta)s^2 + \beta t^* rs(r-s)}{[\beta r + (1-\beta)s]^2},$$

as claimed. The formula for θ also follows from this computation. It is easy to calculate that $\rho \gtrless 1$ as $s \gtrless r$. To find bounds for ρ, we first consider $s >> r$: in this case, ρ is approximately $f(\beta) \equiv (1-\beta)^2/(1-\beta+\beta\ln\beta)$. The function f is monotonically increasing in β, with $\lim_{\beta\to 0} f(\beta) = 1$ and $\lim_{\beta\to 1} f(\beta) = 2$ (use L'Hôpital's rule twice), so ρ can get arbitrarily close to 2. On the other side, for $r >> s$, ρ is approximately $\beta/(1+st^*)$, which is obviously minimal for β near 0 (i.e. rt^* very large), and hence ρ near 0.

For the median time, we need to distinguish two cases: either τ is smaller than t^* (in which case τ occurs while the discount rate is still r and we can use the appropriate expression for $pv(t)$) or it is larger than t^* (where the rate is s and the expression for $pv(t)$ is different). The boundary case will

[30] If $s > r$, say, we can think of $r(t)$ as having something like a Dirac delta function spike at t^*.

be if $\tau = t^*$, which occurs exactly when $pv(t^*) = pv(\infty)/2$, i.e. if $1/r - \beta/r = [\beta r + (1-\beta)s]/2rs$. This is true if $\beta r = (1-\beta)s$ (which is intuitively reasonable), i.e. for $\beta = e^{-rt^*} = \frac{s}{r+s}$. Thus the cutoff value for t^* is $r^{-1}\ln\frac{r+s}{s}$. If t^* is larger than this (so that τ is in the initial range), $pv(\tau) = 1/r - e^{-r\tau}/r$ and we need to set this equal to $pv(\infty)/2 = [\beta r + (1-\beta)s]/2rs$. Solving,

$$\begin{aligned} 1 - e^{-r\tau} &= \frac{1}{2}\frac{\beta r + (1-\beta)s}{s} \implies \\ e^{-r\tau} &= \frac{s - \beta(r-s)}{2s} \implies \\ \tau &= r^{-1}\ln\frac{2s}{s - \beta(r-s)}. \end{aligned}$$

If t^* is smaller than the cutoff value (so that $\tau > t^*$), then $pv(\tau) = \int_0^{t^*} e^{-rt}dt + \beta\int_{t^*}^{\tau} e^{-s(t-t^*)}dt = (1-\beta)/r + \beta\left(1 - e^{-s(\tau - t^*)}\right)/s$ and we solve

$$\begin{aligned} \frac{1-\beta}{r} + \frac{\beta\left(1 - e^{-s(\tau-t^*)}\right)}{s} &= \frac{1}{2}\frac{\beta r + (1-\beta)s}{rs} \implies \\ \beta r e^{-s(\tau-t^*)} &= \frac{1}{2}[\beta r + (1-\beta)s] \implies \\ e^{-s(\tau-t^*)} &= \frac{\beta r + (1-\beta)s}{2\beta r} \implies \\ \tau &= t^* + s^{-1}\ln\frac{2\beta r}{\beta r + (1-\beta)s}. \end{aligned}$$

6.1.4 Split function quasi-hyperbolic

We now consider the procedure that splits the discount function rather than the discount rate:

$$d(t) = \begin{cases} e^{-rt} & \text{if } t \le t^* \\ \lambda e^{-rt} & \text{if } t > t^* \end{cases}$$

where $r > 0$ and $\lambda < 1$. The discount rate for this procedure is constant at r except at $t = t^*$, where it is infinite. If we again denote e^{-rt^*} by β, we see that

$$\int_0^\infty d(t)dt = \int_0^{t^*} e^{-rt}dt + \lambda\int_{t^*}^\infty e^{-rt}dt = (1-\beta)\frac{1}{r} + \lambda\frac{1}{r}\beta = \frac{1-(1-\lambda)\beta}{r},$$

and the amount of discounting is the inverse of this:

$$\alpha = \frac{r}{1-(1-\lambda)\beta}.$$

For the speed,

$$\begin{aligned}
\int_0^\infty t d(t) dt &= \int_0^{t^*} te^{-rt}dt + \lambda \int_{t^*}^\infty te^{-rt}dt \\
&= \frac{1}{r}((1-\beta)\frac{1}{r} - t^*\beta) + \lambda \frac{1}{r}(t^*\beta + \frac{1}{r}\beta) \\
&= \frac{1}{r^2}[1-(1-\lambda)\beta(1+rt^*)]
\end{aligned}$$

and so

$$\rho = \frac{[1-(1-\lambda)\beta]^2}{1-(1-\lambda)\beta(1+rt^*)}.$$

From here, θ follows as always. To find bounds on ρ for this procedure, we first consider the extreme case $\lambda = 0$: here $\rho = f(\beta) = (1-\beta)^2/(1-\beta+\beta\ln\beta)$ (same as in the last subsection). So once more the maximum value for ρ is 2. Note that if $\lambda > 0$, then ρ is maximized at an interior choice of β, but the arg max goes to 1 (and the maximum value goes to 2) as λ goes to 0. On the other hand, fixing λ and letting β go to 1, ρ is approximately λ – so we can make it as small as we want (greater than 0). Thus the order of limits makes a big difference! Finally, we point out that for any fixed λ, ρ goes to 1 as β goes to 0.

For the median time, we again distinguish two cases: either τ is smaller than t^* (and τ occurs before the λ jump) or it is larger than t^* (where the expression for $pv(t)$ is different due to the λ factor). The boundary case will be if $\tau = t^*$, which occurs exactly when $pv(t^*) = pv(\infty)/2$, i.e. if $1/r - \beta/r = [1-(1-\lambda)\beta]/2r$. This is true when $\beta = 1/(1+\lambda)$, i.e. for $t^* = r^{-1}\ln(1+\lambda)$. If t^* is larger than this (so that τ is in the initial range), $pv(\tau) = 1/r - e^{-r\tau}/r$ and we need to set this equal to $pv(\infty)/2 = [1-(1-\lambda)\beta]/2r$. Solving,

$$\begin{aligned}
1 - e^{-r\tau} &= \frac{1-(1-\lambda)\beta}{2} \implies \\
e^{-r\tau} &= \frac{1+(1-\lambda)\beta}{2} \implies \\
\tau &= r^{-1}\ln\frac{2}{1+(1-\lambda)\beta}.
\end{aligned}$$

If t^* is smaller than the cutoff value (so that $\tau > t^*$), then $pv(\tau) = \int_0^{t^*} e^{-rt}dt + \beta\int_{t^*}^{\tau} \lambda e^{-r(t-t^*)}dt =$

$(1-\beta)/r+\lambda\left(\beta-e^{-r\tau}\right)/r$ and we solve

$$\begin{aligned}
\frac{1-\beta}{r}+\frac{\lambda\left(\beta-e^{-r\tau}\right)}{r} &= \frac{1-(1-\lambda)\beta}{2r} \implies \\
\lambda\left(\beta-e^{-r\tau}\right) &= \frac{1}{2}\left[\lambda\beta-(1-\beta)\right] \implies \\
e^{-r\tau} &= \frac{\lambda\beta+(1-\beta)}{2\lambda} \implies \\
\tau &= r^{-1}\ln\frac{2\lambda}{1-(1-\lambda)\beta}.
\end{aligned}$$

6.1.5 Hyperbolic

If $d(t)=[1+r(1-s)t]^{-\frac{1}{1-s}-1}$ then

$$r(t)=-(-\frac{1}{1-s}-1)r(1-s)\frac{[1+r(1-s)t]^{-\frac{1}{1-s}-2}}{[1+r(1-s)t]^{-\frac{1}{1-s}-1}}=\frac{r(2-s)}{1+r(1-s)t}$$

and

$$\int_0^\infty d(t)dt=\frac{1}{r(1-s)}\frac{1}{-1/(1-s)}\left[1+r(1-s)t\right]^{-\frac{1}{1-s}}\Big|_0^\infty=\frac{1}{r}$$

and so $\alpha=r$ as claimed. For the speed, we begin by observing that

$$\int_0^\infty td(t)dt=\int_0^\infty\frac{t}{[1+r(1-s)t]^{1+\frac{1}{1-s}}}dt=\frac{1}{r^2(1-s)^2}\int_0^\infty\frac{x}{[1+x]^{1+\frac{1}{1-s}}}dx,$$

with the change of variable $x=r(1-s)t$. The beta function $B(y,z)$ can be written

$$B(y,z)=\int_0^\infty\frac{x^{y-1}}{[1+x]^{y+z}}dx,$$

so inserting $y=2$ and $z=\frac{1}{1-s}-1=\frac{s}{1-s}$ yields

$$\int_0^\infty td(t)dt=\frac{1}{r^2(1-s)^2}B(2,\frac{s}{1-s}).$$

But the beta function can be expressed in terms of the gamma function (Gradshteyn and Ryzhik 1980, section 8.38) as

$$B(y,z)=\frac{\Gamma(y)\Gamma(z)}{\Gamma(y+z)},$$

which itself has the property that $\Gamma(y+1)=y\Gamma(y)$. Plugging in (and recalling that $\Gamma(n)=(n-1)!$ for integer n), we find that

$$B(2,\frac{s}{1-s})=\frac{\Gamma(2)\Gamma(\frac{s}{1-s})}{\Gamma(\frac{s}{1-s}+2)}=\frac{1\cdot\Gamma(\frac{s}{1-s})}{(\frac{s}{1-s}+1)\Gamma(\frac{s}{1-s}+1)}=\frac{\Gamma(\frac{s}{1-s})}{(\frac{1}{1-s})(\frac{s}{1-s})\Gamma(\frac{s}{r-s})}=\frac{(1-s)^2}{s}.$$

Finally,

$$\rho = \left[\alpha^2 \int_0^\infty td(t)dt\right]^{-1} = \left[r^2 \frac{1}{r^2(1-s)^2} \frac{(1-s)^2}{s}\right]^{-1} = s$$

as desired. And of course θ is then $(rs)^{-1}$. For the median time, we set $pv(\tau) = pv(\infty)/2$:

$$\begin{aligned}
\frac{1-[1+r(1-s)\tau]^{-\frac{1}{1-s}}}{r} &= \frac{1}{2r} \implies \\
1+r(1-s)\tau &= 2^{1-s} \implies \\
\tau &= \frac{2^{1-s}-1}{r(1-s)}.
\end{aligned}$$

Note that, as expected, $\lim_{s\to 1}\tau = r^{-1}\ln 2$ (use L'Hôpital's rule), which is the value for the standard exponential.

6.1.6 Time-transformed exponential

For $d(t) = \exp(-rt^{1/s})$, we can immediately observe that $r(t) = (r/s)t^{1/s-1}$. To calculate amount and speed, we will use the following formula (Gradshteyn and Ryzhik 1980, section 3.4781):

$$\int_0^\infty x^{y-1}\exp(-rx^z)dx = \frac{1}{z}r^{-y/z}\Gamma(y/z)$$

with $z = 1/s$. First substituting $y = 1$, we find

$$\int_0^\infty d(t)dt = \frac{1}{1/s}r^{-s}\Gamma(s) = \frac{s\Gamma(s)}{r^s} = \frac{\Gamma(s+1)}{r^s}$$

using the same identity $\Gamma(y+1) = y\Gamma(y)$ as before. The inverse of this is then α, as claimed. To calculate the speed, we instead substitute $y = 2$, implying that

$$\int_0^\infty td(t)dt = \frac{1}{1/s}r^{-2s}\Gamma(2s) = \frac{s\Gamma(2s)}{r^{2s}}$$

so that

$$\rho = \left[\alpha^2 \int_0^\infty td(t)dt\right]^{-1} = \frac{s^2[\Gamma(s)]^2}{r^{2s}} \frac{r^{2s}}{s\Gamma(2s)} = s\frac{[\Gamma(s)]^2}{\Gamma(2s)} = \frac{\Gamma(s)\Gamma(s+1)}{\Gamma(2s)},$$

also as claimed. The mean time follows immediately:

$$\theta = \alpha \int_0^\infty td(t)dt = \frac{\Gamma(2s)}{r^s\Gamma(s)}.$$

Unfortunately, there is no closed form integral for $d(t)$ and thus no closed form expression for the median time horizon.

We can further investigate several properties of this less well-known discounting procedure. The binomial coefficient $\binom{n}{m}$ is by definition

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{\Gamma(n+1)}{\Gamma(m+1)\Gamma(n-m+1)},$$

and the formulation in terms of gamma functions can be used to define a continuous binomial coefficient for non-integer n, m. In that case, the relative speed is the inverse of

$$\frac{\Gamma(2s)}{\Gamma(s)\Gamma(s+1)} = \frac{\Gamma((2s-1)+1)}{\Gamma((2s-1)-s+1)\Gamma(s+1)} = \binom{2s-1}{s},$$

which is strictly increasing in s for $s > 0$ (since it is an extension of the binomial coefficient). Thus the relative speed is strictly decreasing in s, and of course it equals $\binom{2-1}{1} = 1$ when $s = 1$. This validates our claim that the transformed time discounting procedure is slow (i.e. $\rho < 1$) exactly when $s > 1$; and it is fast ($\rho > 1$) when $s < 1$. It also provides a simple way to calculate ρ when s is an integer. The claim is also a direct consequence of Theorem 2, once we note that the derivative of the discount rate has the same sign as $1 - s$:

$$r'(t) = (1-s)\frac{r}{s^2}t^{(1/s)-2}.$$

So when $s > 1$, the discount rate is declining over time (slow discounting), and vice-versa.

Finally, we examine the limit cases for extreme values of s. As $s \to \infty$, $d(t)$ starts to look almost constant at a value of e^{-r}, so the procedure is less and less convergent (though of course for any finite s there is no actual problem). This means that $\alpha \to 0$. Since $\Gamma(y)$ grows "factorially" in y, we find that ρ goes to 0 as well (and θ grows without bound), which does not automatically follow from the result for α. That is, the limit discounting procedure is slow even relative to the equivalent exponential (which has, of course, a zero discount rate in the limit). In the other extreme, as $s \to 0$, $d(t)$ starts to appear constant at 1 until $t = 1$, where it precipitously drops to 0. This suggests that α should be near 1, and indeed

$$\lim_{s\to 0} \alpha = \lim_{s \to 0} \frac{r^s}{\Gamma(s+1)} = \frac{r^0}{\Gamma(1)} = 1$$

since both numerator and denominator are continuous at $s = 0$. For the speed, we make use of the doubling formula for the gamma function:

$$\Gamma(2y) = \frac{2^{2y-1}}{\sqrt{\pi}}\Gamma(y)\Gamma(y+\frac{1}{2}),$$

from which we obtain

$$\rho = \frac{\Gamma(s)\Gamma(s+1)}{\Gamma(2s)} = \frac{\sqrt{\pi}}{2^{2s-1}}\frac{\Gamma(s+1)}{\Gamma(s+1/2)}$$

and thus (using the known value $\Gamma(1/2) = \sqrt{\pi}$)

$$\lim_{s\to 0}\rho = \lim_{s\to 0}\frac{\sqrt{\pi}}{2^{2s-1}}\frac{\Gamma(s+1)}{\Gamma(s+1/2)} = \frac{\sqrt{\pi}}{2^{0-1}}\frac{1}{\sqrt{\pi}} = 2,$$

again by continuity at $s = 0$ of all functions involved. This is the "fastest" that we can get with this type of discounting procedure. Therefore, θ approaches $1/2$, and it is clear in this case that τ approaches $1/2$ as well.

6.2 Proof of Proposition 6

Proposition 6 relates the limiting value of the social discount rate function to the minimum of the limiting values of the individual discount rate functions. Recall that

$$r_{\overline{D}}(t) = \frac{\int_X \alpha(x)r(t;x)d(t;x)f(x)dx}{\int_X \alpha(x)d(t;x)f(x)dx},$$

which we can rewrite as a weighted average

$$r_{\overline{D}}(t) = \int_X \beta(t;x)r(t;x)dx$$

with time-dependent weights $\beta(t;x)$ given by

$$\beta(t;x) = \frac{\alpha(x)d(t;x)f(x)}{\int_X \alpha(x)d(t;x)f(x)dx}.$$

Obviously, $\int_X \beta(t;x)dx = 1$ for any t.

For any $\varepsilon > 0$, let us partition X into A_1, $A_2(\varepsilon)$, and $A_3(\varepsilon)$ as follows: $A_1 = \{x \in X \text{ s.t. } r^*(x) < r^*_{\min}\}$; $A_2(\varepsilon) = \{x \in X \text{ s.t. } r^*_{\min} \le r^*(x) \le r^*_{\min} + \varepsilon\}$; and $A_3(\varepsilon) = \{x \in X \text{ s.t. } r^*_{\min} + \varepsilon < r^*(x)\}$. Then we can write $\lim_{t\to\infty} r_{\overline{D}}(t) = \lim_{t\to\infty}\int_X \beta r dx = \lim_{t\to\infty}\int_{A_1}\beta r dx + \lim_{t\to\infty}\int_{A_2(\varepsilon)}\beta r dx + \lim_{t\to\infty}\int_{A_3(\varepsilon)}\beta r dx$.

For any x with $\alpha(x) = 0$, $\beta(t;x) = 0$ identically. So ignoring any such x and noting that $\int_{A(r^*_{\min})} f(x)dx = 0$ by definition of $r^*_{\min}$, we see that $\int_{A_1}\beta(t;x)dx = 0$ as well (for all t), since $\alpha(x)$ and $d(t;x)$ are finite (and α, d, f are all weakly positive). But then $\lim_{t\to\infty}\int_{A_1}\beta(t;x)r(t;x)dx = 0$, so these values of x have no effect on $\lim_{t\to\infty} r_{\overline{D}}(t)$. This establishes that $\lim_{t\to\infty} r_{\overline{D}}(t) \ge r^*_{\min}$; it remains to show that $\lim_{t\to\infty} r_{\overline{D}}(t) \le r^*_{\min}$.

Also by definition of $r^*_{\min}$, $\int_{A(r^*_{\min}+\varepsilon)} f(x)dx > 0$ for all $\varepsilon > 0$. But, as we just saw, $\int_{A_1} f(x)dx = 0$. Hence, for any ε, we can pick $x_2 \in A_2(\varepsilon)$ with $\alpha(x_2) > 0$ and $f(x_2) > 0$; now take any $x_3 \in A_3(\varepsilon)$. Then

$$\frac{\beta(t;x_3)}{\beta(t;x_2)} = \frac{\alpha(x_3)d(t;x_3)f(x_3)}{\alpha(x_2)d(t;x_2)f(x_2)}$$

Rewriting $d(t;x)$ as $\exp\left(-\int_0^t r(\tau;x)d\tau\right)$, we get

$$\frac{\beta(t;x_3)}{\beta(t;x_2)} = \frac{\alpha(x_3)f(x_3)}{\alpha(x_2)f(x_2)} \frac{\exp\left(-\int_0^t r(\tau;x_3)d\tau\right)}{\exp\left(-\int_0^t r(\tau;x_2)d\tau\right)} = M\exp\left(-\int_0^t \left[r(\tau;x_3) - r(\tau;x_2)\right]d\tau\right),$$

where $M = \frac{\alpha(x_3)f(x_3)}{\alpha(x_2)f(x_2)}$ is constant in t (and is finite by the choice of x_2). Therefore

$$\lim_{t\to\infty} \frac{\beta(t;x_3)}{\beta(t;x_2)} = M \lim_{t\to\infty} \exp\left(-\int_0^t \left[r(\tau;x_3) - r(\tau;x_2)\right]d\tau\right).$$

But of course $x_2 \in A_2(\varepsilon)$ and $x_3 \in A_3(\varepsilon)$ imply that $\lim_{t\to\infty} r(t;x_3) > r^*_{\min} + \varepsilon \geq \lim_{t\to\infty} r(t;x_2)$, from which it is clear that $\lim_{t\to\infty} \int_0^t \left[r(\tau;x_3) - r(\tau;x_2)\right]d\tau = \infty$ and thus

$$\lim_{t\to\infty} \frac{\beta(t;x_3)}{\beta(t;x_2)} = 0.$$

This implies that, for any $x_3 \in A_3(\varepsilon)$, we must have $\lim_{t\to\infty} \beta(t;x_3) = 0$. This in turn yields $\lim_{t\to\infty} \int_{A_3(\varepsilon)} \beta(t;x)dx = 0$ (so that $\lim_{t\to\infty} \int_{A_2(\varepsilon)} \beta(t;x)dx = 1$), but it also implies the stronger conclusion that $\lim_{t\to\infty} \int_{A_3(\varepsilon)} \beta(t;x)r(t;x)dx = 0$. Therefore only $x \in A_2(\varepsilon)$ influence $\lim_{t\to\infty} r_{\overline{D}}(t)$, and in particular $\lim_{t\to\infty} r_{\overline{D}}(t) \leq r^*_{\min} + \varepsilon$ for any $\varepsilon > 0$. But that means exactly that $\lim_{t\to\infty} r_{\overline{D}}(t) \leq r^*_{\min}$, as needed. ■

6.3 Aggregation of Exponentials with Specific Parameter Distributions

We first start with an underlying gamma distribution for the parameter of individual exponential discount functions:

$$f(x) = \frac{a^b}{\Gamma(b)} e^{-ax} x^{b-1},$$

where $x \in (0,\infty)$ denotes the discount rate parameter in a standard exponential, and $a, b > 0$. Thus $\alpha(x) = x$, $d(t;x) = e^{-xt}$, and $f(x)$ is as above. Then

$$\int_0^\infty \alpha(x)d(t;x)f(x)dx = \frac{a^b}{\Gamma(b)} \int_0^\infty x^b e^{-(a+t)x}dx = \frac{a^b}{\Gamma(b)} \frac{\Gamma(b+1)}{(a+t)^{b+1}} = \frac{b}{a}\left[\frac{a}{a+t}\right]^{b+1}.$$

But

$$\overline{\alpha} = \int_0^\infty \alpha(x) f(x) dx = \frac{a^b}{\Gamma(b)} \int_0^\infty x^b e^{-ax} dx = \frac{a^b}{\Gamma(b)} \frac{\Gamma(b+1)}{a^{b+1}} = \frac{b}{a},$$

so $d_{\overline{D}}(t) = [1 + t/a]^{-(1+b)}$ as claimed. The relative speed follows from Theorem 3 with $r = b/a$ and $s = 1 - 1/b$.

Next, in order to allow for a positive weight at 0, we consider an augmented exponential as the underlying distribution on the parameter:

$$f(x) = \frac{a^2}{a+b} e^{-ax}(1 + bx).$$

We have $\alpha(x) = x$ and $d(t;x) = e^{-xt}$ as before, so

$$\begin{aligned}
\int_0^\infty \alpha(x) d(t;x) f(x) dx &= \frac{a^2}{a+b} \int_0^\infty x e^{-(a+t)x}(1+bx) dx \\
&= \frac{a^2}{a+b} \left[\frac{1}{(a+t)^2} + \frac{2b}{(a+t)^3} \right] \\
&= \frac{a^2}{a+b} \left[\frac{a+2b+t}{(a+t)^3} \right]
\end{aligned}$$

and

$$\overline{\alpha} = \int_0^\infty \alpha(x) f(x) dx = \frac{a^2}{a+b} \int_0^\infty x e^{-ax}(1+bx) dx = \frac{a^2}{a+b} \left[\frac{1}{a^2} + \frac{2b}{a^3} \right] = \frac{a+2b}{a(a+b)}.$$

Thus

$$\begin{aligned}
d_{\overline{D}}(t) &= \frac{a^3}{a+2b} \left[\frac{a+2b+t}{(a+t)^3} \right] \\
&= \left(1 + \frac{t}{a+2b} \right) \left[\frac{a}{a+t} \right]^3 \\
&= \left[\frac{a}{a+2b}(1+t/a) + \left(1 - \frac{a}{a+2b} \right) \right] \left[\frac{1}{1+t/a} \right]^3 \\
&= \frac{a}{a+2b} \left[\frac{1}{1+t/a} \right]^2 + \frac{2b}{a+2b} \left[\frac{1}{1+t/a} \right]^3,
\end{aligned}$$

as claimed. Finally, the first of these weighted hyperbolics is [just barely] weakly convergent, with $r = 1/a$ and $s = 0$, so the overall procedure will have $\theta = \infty$ and $\rho = 0$.

References

[1] Apostol, Tom M. *Calculus, volume I*, Second Edition. New York: John Wiley and Sons, 1967.

[2] Arrow, Kenneth J. "Discounting, Morality and Gaming," in Portney, Paul R. and John P. Weyant, eds., *Discounting and intergenerational equity.* Washington D.C.: Resources for the Future, 1999. Pp.13-22.

[3] Barlow, Richard E. and Proschan, Frank. *Mathematical theory of reliability.* New York: John Wiley and Sons, 1965. (Reprinted in 1996 by the Society for Industrial and Applied Mathematics, Philadelphia.)

[4] Bleichrodt, Han and Gafni, Amiram. "Time Preference, the Discounted Utility Model and Health." *Journal of Health Economics*, 1996, *15*, pp. 49-66.

[5] Bleichrodt, Han and Johannesson, Magnus. "Time Preference and Health." *Journal of Mathematical Psychology*, 2001, *45*, pp. 265-282.

[6] Chapman, Gretchen B. "Time Discounting of Health Outcomes," in Loewenstein, George, Read, Daniel, and Roy F. Baumeister, eds., *Time and decision.* New York: Russell Sage Foundation, 2003. Pp. 395-418.

[7] Cline, William R. "Discounting for the Very Long Term," in Portney, Paul R. and John P.Weyant, eds., *Discounting and intergenerational equity.* Washington, D.C.: Resources for the Future, 1999. Pp. 131-140.

[8] Cochrane, John H. *Asset pricing.* Princeton and Oxford: Princeton University Press, 2001.

[9] Cropper, Maureen L. Aydede, S.L. and Portney, Paul R. "Preferences for Life Saving Programs: How the Public Discounts Time and Age." *Journal of Risk and Uncertainty*, 1994, *8*, pp. 243-265.

[10] Fishburn, Peter C. and Rubinstein, Ariel. "Time Preference." *International Economic Review*, 1982, *23*, pp. 677-694.

[11] Frederick, Shane, Loewenstein, George, and O'Donoghue, Ted. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, June 2002, *XL*, pp. 351-401.

[12] Gradshteyn, I.S. and Ryzhik, I.M. *Tables of Integrals, Series, and Products.* San Diego: Academic Press, 1980.

[13] Gyrd-Hansen, Dorte and Søgaard, Jes. "Discounting Life-Years: Whither Time Preference?" *Health Economics,* 1998, *7*, pp. 121-127.

[14] Hara, Chiaki and Kuzmics, Christoph. "Representative Consumer's Risk Aversion and Efficient Risk-Sharing Rules." Unpublished paper, October 2002.

[15] Harris, Christopher and Laibson, David. "Instantaneous Gratification." Unpublished paper, May 2001.

[16] Harvey, Charles M. "Value Functions for Infinite Period Planning." *Management Science*, 1986, *32*, pp. 1123-39.

[17] Harvey, Charles M. "The Reasonableness of Non-Constant Discounting." *Journal of Public Economics*, 1994, *53*, pp. 31-51.

[18] Herrnstein, Richard J. "Self-Control as Response Strength," in Bradshaw, C.M., E. Szabadi and C.F. Lowe, eds., *Quantification of steady-state operant behavior.* Amsterdam: Elsevier/North-Holland, 1981.

[19] Jamison, Dean T. *Conjoint measurement of time preference and utility.* Rand Report RM6029-PR. Santa Monica, California: Rand Corporation, 1969.

[20] Jamison, Dean T., Mosley, W. Henry, Measham, Anthony R., and Bobodilla, José-Luis, eds., *Disease control priorities in developing countries.* Oxford: Oxford University Press for the World Bank, 1993.

[21] Koopmans, Tjalling C. "Stationary Ordinal Utility and Impatience." *Econometrica*, April 1960, *28(2)*, pp. 287-309.

[22] Laibson, David. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, May 1997, *112(2)*, pp. 443-477.

[23] Lipscomb, Joseph. "Time Preference for Health in Cost-Effectiveness Analysis." *Medical Care*, March 1989, *27(3)*, pp. 233-253.

[24] Lipscomb, Joseph, Weinstein, Milton C. and Torrance, George W. "Time Preference," in Gold, Marthe R., Russell, Louise B., Siegal, Joanna A. and Milton C. Weinstein, eds., *Cost-effectiveness in health and medicine.* New York and Oxford: Oxford University Press, 1996. Pp. 214-246.

[25] Loewenstein, George. "The Fall and Rise of Psychological Explanations in the Economics of Intertemporal Choice," in Loewenstein, George and Jon Elster, eds., *Choice over time.* New York: Russell Sage Foundation, 1992. Pp. 3-34.

[26] Loewenstein, George and Prelec, Drazen. "Anomalies in Intertemporal Choice: Evidence and an Interpretation." *Quarterly Journal of Economics*, May 1992, *107(2)*, pp. 573-597.

[27] Malkiel, Burton G. "Term Structure of Interest Rates," in Eatwell, J., Milgate, M., and P. Newman, eds., *The new palgrave dictionary of economics, Vol. 4.* London: Macmillan Reference, Ltd., 1998. Pp. 629-631.

[28] Newell, Richard and Pizer, William. "Discounting the Distant Future: How Much Do Uncertain Rates Increase Valuations?" Discussion Paper 00-45, October 2000. Washington, D.C.: Resources for the Future.

[29] Ramsey, Frank P. "A Mathematical Theory of Saving." *The Economic Journal*, December 1928, *XXXVIII(152)*, pp. 543-559.

[30] Read, Daniel. "Is Time-Discounting Hyperbolic or Subadditive?" *Journal of Risk and Uncertainty*, 2001, *23(1)*, pp. 5-32.

[31] Roelofsma, Peter H.M.P. "Modelling Intertemporal Choices." *Acta Psychologica*, 1996, *93*, pp. 5-22.

[32] Rudin, Walter. *Principles of mathematical analysis*, Third Edition. New York: McGraw-Hill, 1976.

[33] Sen, Amartya. "Approaches to the Choice of Discount Rates for Social Benefit-Cost Analysis," in Lind, Robert, ed., *Discounting for time and risk in energy policy.* Washington, D.C.: Resources for the Future, 1982. Pp. 325-353.

[34] Sen, Amartya. "Choice Involving Time," in Sen, Amartya, *Choice of techniques.* London: Blackwell, 1972, Pp. 72-80.

[35] Sokolnikoff, I.S. and Redheffer, Raymond A. *Mathematics of physics and modern engineering.* New York: McGraw-Hill, 1958.

[36] Strotz, Robert H. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies*, 1956, *23(3)*, pp. 165-180.

[37] Weitzman, Martin L. "Gamma Discounting." *American Economic Review*, March 2001, *91(1)*, pp. 260-271.

[38] Weitzman, Martin L. "Why the Far-Distant Future Should Be Discounted at Its Lowest Possible Rate." *Journal of Environmental Economics and Management*, Novemeber 1998, *36(3)*, pp. 201-208.

[39] World Bank. *Investing in health: world development report 1993.* Oxford and New York: Oxford Univesity Press for the World Bank, 1993.

DCPP Working Papers

1. At Least One-Third Of Poor Countries' Disease Burden Is Due To Malnutrition. By John B. Mason, Philip Musgrove and Jean-Pierre Habicht. March 2003.
2. Progress In The Development Of A Recombinant Vaccine For Human Hookworm Disease: The Human Hookworm Vaccine Initiative. By Peter J. Hotez, Bin Zhan, Jeffrey M. Bethony, Alex Loukas, Angela Williamson, Gaddam Narsa Goud, John M. Hawdon, Azra Dobardzic, Reshad Zook, Yan Wang, Sen Liu, Idong Essiet-Gibson, Sophia Chung-Debose, Shushua Xiao, David Knox, Michael Meagherf, Mehmet Inan, Rodrigo Correa-Oliveira, Paul Vilk, Herman R. Shephard, Walter Brandt, and Philip K. Russell. March 2003.
3. Soil Transmitted Helminth Infections: The Nature, Causes And Burden Of The Condition. By Peter J. Hotez, Nilanthi de Silva, Simon Brooker and Jeffrey Bethony. March 2003.
4. Discounting. By Dean T. Jamison and Julian S. Jamison. March 2003.
5. Economics of Malaria Resistance and the Optimal Use of Artemisinin-Based Combination Treatments (ACTs). Ramanan Laxminarayan. July 2003.
6. Do Malaria Control Interventions Reach the Poor? A View Through the Equity Lens. Lawrence Barat, Suprotik Basu, Eve Worrall, Kara Hanson, Anne Mills. July 2003.
7. New Perspectives on the Causes and Potential Costs of Malaria: The Growth and Development of Children. What Should We Be Measuring and How Should We Be Measuring It? Penny Holding, P.K. Kitsao-Wekulo. July 2003.
8. Pediatric Anemia and Mortality in Africa: Plasmodium falciparum Malaria as a Cause or Risk? Robert W.Snow, Eline L. Korenromp, Chris Drakeley, Eleanor Gouws. July 2003.
9. Unit Prices of Health Care Inputs in Low and Middle Income Regions: An outline and discussion of data for use in DCPP chapters. Mulligan J, Fox-Rushby J, Mills A. August 2003.
10. Health's Contribution to Economic Growth in an Environment of Partially Endogenous Technical Progress. Dean T. Jamison, Lawrence J. Lau, Jia Wang. July 2003
11. The Public Health of Plasmodium falciparum Malaria in Africa: Deriving the Numbers. By Robert W. Snow, Marlies H. Craig, Charles R.J.C. Newton, and Richard W. Steketee. Working Paper No. 11, Disease Control Priorities Project. Bethesda, Maryland: Fogerty International Center, National Institutes of Health. September 2003.
12. Soil-Transmitted Helminth Infections: Updating the Global Picture. By Nilanthi de Silva, Simon Brooker, Peter Hotez, Antonio Montresor, Dirk Engels, and Lorenzo Savioli. Working Paper No. 12, Disease Control Priorities Project. Bethesda, Maryland: Fogerty International Center, National Institutes of Health. September 2003.

Secretariat Address:
Fogarty International Center
National Institutes of Health
16 Center Drive, MSC 6705
Bethesda, MD 20892-6705
USA
• Tel: 1-301-496-2091
• Fax: 1-301-496-8496
• Web: www.fic.nih.gov/dcpp
• Email : dcppwps@nih.gov