"FORESIGHT AS A SURVIVAL CHARACTERISTIC: WHEN (IF EVER) DOES THE LONG VIEW PAY?"

by

R.U. AYRES*
and
R. AXTELL**
93/83/EPS

* Previously Professor at Carnegie-Mellon University, Pittsburgh, PA, 15213, now Professor of Environmental Economics, Sandoz Chair in Management and the Environment, at INSEAD, Boulevard de Constance, Fontainebleau 77305 Cedex, France.

** Previously Professor at Carnegie-Mellon University, Pittsburgh, PA, 15213 now Professor at Brookings Institution, Washington, USA.

A working paper in the INSEAD Working Paper Series is intended as a means whereby a faculty researcher's thoughts and findings may be communicated to interested readers. The paper should be considered preliminary in nature and may require revision.

Printed at INSEAD, Fontainebleau, France
FORESIGHT AS A SURVIVAL CHARACTERISTIC: WHEN (IF EVER) DOES THE LONG VIEW PAY?

R.U. Ayres and R. Axtell
Department of Engineering & Public Policy
Carnegie-Mellon University
Pittsburgh, PA, 15213

Lecture prepared for
Hydro-Quebec Seminar on Technological Change
University of Quebec at Montreal,
Montreal, Canada
May 2, 1991

Abstract

Long range R&D and capital investment projects are normally evaluated by means of a procedure (benefit/cost analysis) that involves a choice of time preference functions. Benefit-cost analysts and many economists typically assume that time preference is a behavioral fact of life, and that the time preference function is essentially equivalent to a compound interest law. In practice, they tend to choose discount rates in the range of 3%-8% p.a. in real terms. Variability in projected benefit/cost ratios resulting from this uncertainty is commonly dealt with ad hoc, e.g. by simply presenting results for several different discount rates and letting the ‘decision-maker’ select among them. It is argued in this paper that the basic discounting methodology is fundamentally flawed and can lead to significantly inferior social choices, i.e. choices that would be rejected by virtually any rational actor to whom the choice were fairly presented. A more general methodology is needed. Examples from several realms, including energy policy and environmental protection are discussed to illustrate the thesis.
Introduction

The problem of forecasting in economics has not been given the attention it deserves. On the one hand, much of the basic theory of micro-economics is still built on the basic model of static utility-maximization with perfect information. Each actor in the market place is assumed to know all about the others’ utility preferences and demand functions, as well as supply schedules and actual market prices. Walrasian and post-Walrasian theory of competitive equilibrium is predicated on a static or quasi-static situation, in which structural change is ruled out by assumption. (The notion of equilibrium, as it is understood in the physical sciences, is not even well-defined in a truly dynamic economy characterized by structural change).

Time has no secure place in the classical micro-economic theory of utility maximization. It is implicitly assumed in most of the micro-economics literature that the "Pareto optimum" is a state of the economic system that exists, and can be determined, once and for all time. In this framework, it was not necessary to distinguish between the short term and the long term. Yet, such a distinction between short-term (or immediate) and long-term satisfaction is clearly required in even the simplest theoretical treatment of economic growth. Growth requires savings and investment. Both are types of economic behavior that cannot be explained without introducing both the notion of duration, and some theory of time preference. (In more common jargon, the latter is known as discounting). In other words, it is necessary to explain why, and on what basis, people choose between consumption in the present and consumption in the future.

The economics literature, for the most part, tends to brush this issue under the rug by asserting (with only slight evidence) a kind of "compound interest" law of time preference. This law postulates a fixed market-determined discount rate valid at all times for the typical individual or firm "at the margin" [e.g. Pigou 20; see also Arrow 76]. This is the rate at which the decision-maker is indifferent to consuming now or saving/investing for the future. If the expected return on an investment is sufficiently higher than the discount rate, the marginal decision will be to invest. If not, the marginal decision will be to consume. Private firms are generally assumed to discount on the basis of their capital costs or average internal return on investment. Government agencies are assumed to discount at a rate based on either the cost of borrowing or the GNP growth rate.

Nevertheless, it is also a fact of life that decision-makers in different circumstances make different choices; often these choices are inconsistent with standard benefit-cost assumptions. For instance, it is well known that the implicit discount rate of consumers depends very much on income level, with low-income consumers exhibiting implicit discount rates as high as 80% p.a. [e.g. Hausman 79]. Discounting behavior, based on propensity to save, also depends on age and future income expectations. There is also some evidence that discount rates depend on the time horizon involved. Specifically, it appears that individuals may tend to discount at higher rates for short times than for longer times. In mathematical terms, such behavior is simulated by so-called hyperbolic functions. Some interesting results are presented in Appendix I.
In the circumstances it is less than obvious that, for discount rates, a "single size fits all", and understandably wide disagreement over what the "correct" discount rate should be, or how it should be justified [e.g. Page 77; Hanke & Anwyll 80]. There is, however, fairly general agreement that the choice involves ethical considerations (e.g. how much of a "vote" should be given to future generations or to other species, etc), but little agreement beyond that. This is because there is virtually no general theory to explain why discount rates vary for different decision-makers, at different times. To put it another way, there is little or no theory to explain how discount rates or time preferences are selected, or how they depend on exogenous conditions. We will return to this topic later.

Uncertainty, Forecasting & Far-sightedness

In response to the obvious unrealism of several of the standard micro-economic assumptions, there has been a major effort in recent decades to modify and extend the fundamental theoretical constructs by explicitly allowing for uncertainty. For instance, Herbert Simon’s idea of "bounded rationality" — which effectively jettison’s one of the fundamental props of the standard theory — is justified, in part, on the physical impossibility of processing the massive amounts of information that would be needed to implement a maximization criterion in any moderately complex situation [Simon 55, 59].

In practice, uncertainty has been treated largely in terms of probability theory and distribution functions. The underlying conceptual model is analogous to a rifle shooting at a target. The "truth" is the bulls-eye. If one knows the radial and angular distribution of the shots, one can compute a probability of hitting within a given radius of the bulls-eye as a function of the number of attempts, the distance, etc. Sophisticated statistical tools, such as multiple-correlation analysis have been developed to manipulate these concepts and quantify degrees of uncertainty. These tools can also be used to tease out the relationships among "hidden variables", if such relationships exist.

The probabilistic model above is not sufficiently comprehensive, however. One useful taxonomy that brings some of the fundamental difficulties out into the open is the following:

1. **Risk**: the odds are known and calculable (e.g. by actuaries). This obviously applies to the calculation of insurance premiums, for instance.

2. **Uncertainty**: the odds aren’t known, but the variables are. In this case one cannot compute insurance premiums, but it may be possible to make useful inferences about how the premiums should depend on conditions. "Scenarios" of the future can be constructed, parametrizing the quantitative uncertainties.

3. **Ignorance**: the variables aren’t known, at least completely. But even with relatively little knowledge, it may be possible to make some inferences about stability, for instance.
Forecasting is possible only to the extent that the areas of essential ignorance, at least, are qualitatively understood. In this case, scenarios have limited value.

4. **Indeterminacy**: causal chains are inherently unstable or chaotic. Virtually nothing can be inferred about the future of the system, except that it is unpredictable. Forecasting in the normal is not possible, but it may be extremely important to recognize the likelihood of indeterminacy.

In short, forecasting is an activity intended to minimize risk and reduce uncertainty and ignorance about the future, i.e. to increase far-sightedness. Forecasting in the real world is limited mainly by ignorance and indeterminacy. Ignorance can be reduced, in principle, by research. Indeterminacy cannot. Forecasting must never be confused with prediction, which implies either a testable theory or a crystal ball. In the real world true prediction is rarely possible, and when it is possible it is likely to be trivial. A useful analogy for forecasting is the use of radar as a navigational aid. The accuracy of a forecast clearly decreases with range, but its value to decision-makers may actually *increase* with range, at least up to a point. There is no value in seeing the iceberg ahead if it is too late to turn the ship.

One important use of forecasting is to foresee possible undesirable outcomes (catastrophes) that can be extrapolated from present actions or trajectories. For instance, the captain of the Titanic could have avoided the loss of his ship and many lives if he had been able to see the iceberg in his path and take evasive action. Of course it is easy to cite such examples in retrospect. However, there are many cases where certain types of catastrophes — e.g. earthquakes, tornados and floods — occur on a probabilistically predictable basis. There are many other examples of a similar kind that could be cited [e.g. Ayres & Sandilya 87].

In the above context, it should be noted that there is often a direct and quantifiable relationship between forecasting ability and the cost of damage avoidance and/or damage control. For instance, consider the case of primitive agriculturalists, such as the Anazasi Indians, living on marginal land where severe droughts occur from time to time. Their only possible protection from such occurrences is to store surplus food during times of good harvests, against possible times of need. How big should the stockpile be? The answer depends on the length of the longest possible drought, based on the historical record. (For the Biblical tribes of Israel the traditional answer was seven years). But this is a very wasteful strategy if there is no drought. A great deal of stored grain will rot or be eaten by rodents. Suppose, however, the tribes were able to develop a complete theory to predict the onset and length of droughts. Based on such knowledge the Indians could build up their stockpiles only when a drought was forecast, and use the surplus more productively — perhaps to develop water storage capability or other assets.

For purposes of this paper, a somewhat more immediate example might be the choice of a strategy to respond to the threat of greenhouse warming. This problem exhibits all of the forms of uncertainty, ignorance and indeterminacy cited above. The climate models and impact models involve considerable scientific uncertainty and some degree of scientific ignorance (e.g. of some
biological control mechanisms). Moreover, it is not unlikely that the entire climate-geosphere-biosphere system is so non-linear that it behaves chaotically. This means the system "moves" (in phase-space) inherently unpredictably within some domain (defined by a "strange attractor") the size of which is, itself, unpredictable. The appropriate response to the threat of climate warming is clearly complicated by these factors. An aspect of the problem is that we do not know and probably cannot exactly how much it is "worth", in strict economic terms, to eliminate a ton of carbon dioxide emissions. This bears, of course, on the selection of future energy technologies.

Consider now, in more detail, the problem of long term technology strategy. Suppose a firm, or a government, faces a declining resource. (For the sake of concreteness, take energy. Note that U.S. petroleum output has been declining since 1970; the world’s largest current producer, the U.S.S.R. probably saw peak production in 1987 or 1988). A list of future options has been compiled. Among them, some are available immediately — like nuclear power — while others still require R&D. On the other hand, the nuclear option is relatively mature: it is not likely to decline in cost with further deployment. Other options, such as solar photovoltaic cells, may become far cheaper — but only if a commitment to large-scale production and deployment is assured. If solar power is neglected in the short run, i.e. never subsidized, it may never become competitive. On the other hand, if higher costs are tolerated in the short run, long-run costs may even fall well below present levels.

To decide among the above options (not to mention others) it is necessary to forecast the improvements that will occur in nuclear and solar cell technology as a function of R&D investment and scale of production of each. In this forecast, the economies of scale will be critical. As will be seen later, a wrong choice now may be impossible to reverse later, since the winner of the contest is quite likely to "lock out" its unsuccessful rivals. Notice, however, that the contest is likely to be conducted by proxy, in the form of paper studies using benefit-cost methodology.

Details of the above case will be discussed later. However, the economic value of forecasting per se is not really at issue here. The question we want to address is, rather, the relative value of a short-range perspective vis à vis a long-range perspective, in various circumstances.

Time Preference Revisited

To clarify the more fundamental issues being addressed here, we return now to a point made earlier: there is no real theory to explain how discount rates or time preferences depend on circumstances. This is tantamount to saying that we do not thoroughly understand the evolutionary advantages of far-sightedness, including discounting behavior. It is natural for neo-
classical economists to suggest that discount rates are, in effect, interest rates. Based on this idea, it would follow that competitive capital markets "select" interest rates in the same way that markets determine prices, namely by matching the supply of lendable savings with the demand for investment. But, even if this explanation is appealing at first glance, it suffers from serious flaws.

In the first place, capital markets are different from labor and commodity markets in that they reflect future expectations rather than current reality. In other words, both the supply of savings and the demand for investment capital depend on economic forecasts — implicit or explicit. In a static or quasi-static framework, of course, the forecast itself is relatively trivial. Future per capita growth in such a framework depends only on the savings rate. The latter, in turn, would depend only on the market interest rate on savings, which would depend on the continued profitability of investments of the current type, also at the current rate of investment and growth. It is a closed system. Absent external constraints, such a growth path can continue indefinitely along a "turnpike".

Growth of the above kind is called homothetic by theorists. It resembles the growth of a bacterial colony or a coral reef. It is characterized by ever increasing output of the same goods and services that were produced at the start. This is not the sort of technologically driven economic growth we actually experience, especially since the industrial revolution. Real economic growth, by contrast, is not quasi-static or structure-conserving. On the contrary, as Joseph Schumpeter emphatically pointed out, it is built on a succession of radical technological and social innovations resulting, ultimately, in structural change [Schumpeter 34]. To be sure, these radical innovations are followed by numerous small improvements that may be quantitatively far more important in terms of performance and cost than the basic innovation [Rosenberg 72, 76]. However, while the process of continuous technological improvement might conceivably be stretched or squeezed into the procrustean framework of homothetic growth theory, Schumpeterian innovation cannot.

Schumpeter also postulated the economic mechanism that drives radical innovation [ibid]. It is the opportunity that an innovator gets (either by virtue of legal patent protection or simply by being first and staying ahead of the imitators) to achieve a temporary monopoly position in the market. This enables the monopolist to charge a premium price on the new product or service and thus obtain an extraordinary profit for himself. It is the extra profit that compensates the innovator for risk. Economists in recent decades have found this model quite helpful in explaining differences in innovation behavior between large and small firms, firms in "mature"
vs. "adolescent" industries, and so on. It is also helpful in explaining the evolution of industrial structure [e.g. Nelson & Winter 78, 82].

Returning to the question of time preference, the point we want to make is can be stated as follows: The classical notion that discount rates are like any other interest rate, — at least in the sense that the price of capital in a competitive marketplace (where each actor knows everything about all the others) — doesn’t hold up. This interpretation is only possible if there is no structural change i.e. no innovation in the economy. The reason is that the classical existence proofs of the Walrasian equilibrium — where supply and demand mutually adjust via some tâtonnement process — are predicated on the assumption that every economic actor has perfect information about the current state of the market. In the Walrasian paradigm there is, in effect, no future, or at any rate the future is predictable and ‘futures’ markets operate perfectly.

Real economies do not fit the Walrasian prescription, for many reasons. To understand dynamic economic systems characterized by waves of Schumpeterian radical innovations — dynamic economic behavior with a time dimension — it is important to take into account the related phenomena of short-sightedness (myopia) and far-sightedness (presbyopia). These phenomena depend upon, and reflect time preferences and discounting; they also reflect both ignorance and indeterminacy.

This brings us back to the relationship of discounting to forecasting per se. In the economics literature, the argument tends to revolve around the "right" choice, which most authors take to be somewhere between 3% and 8% in real terms. Arrow [Arrow 76] has argued that the discount rate has two components:

1. Pure utility time preference, which is equal to the (social) time preference for market goods and services, hence equal to the market rate of interest for savings.

2. Decreasing marginal utility due to (the expectation of) increasing wealth.

---

2 The literature is enormous, but the first general proof was due to Arrow & Debreu [Arrow & Debreu 54]. Most important developments up to 1970 or so have been summarized in [Arrow & Hahn 71]. Key papers in recent years include contributions by [Smale 76] and [Aubin 81]. However the first proof of convergence that does not assume that each actor possesses perfect information at least about other market transactions was offered recently by the authors [Ayres & Martinis 90, 90a].

3 This is the essence of so-called ‘rational expectations’ theory.

4 However a number of economists have argued that the discount rate should be zero, on ethical grounds, to insure that future generations have a ‘vote’ equal to that of the present generation [e.g. Ramsey 28; Page 77; Georgescu-Roegen 79; Daly 90].
The first of these components is seen by many economists as a kind of intrinsic behavioral attribute of humans — "a weakness in our telescopic faculty" [Pigou 20 pp. 24-25]. In effect, Pigou attributes the pure utility time preference to our inability to forecast the future! (Recall the comments made above regarding the psychological aspect of this problem). One source of Pigouvian "prevision impairment", as one might term it, is surely the possibility that the investor, or lender, will not get (some or all of) his money back. This possibility exists even if the money is "in the bank". Banks can fail, and many have — even recently. The market interest rate unquestionably reflects, in part, the (perceived) security of the financial system. The long decline in sterling interest rates from the 1660's until the first end of the 19th Century must be a measure of the increasing safety of British investments, as perceived by the financial markets. By the same token, market interest rates in countries with unstable governments or shaky economies — as, for instance, Latin America and Africa — tend to be very high. Thus, the market interest rate is surely a measure of apparent investment risk. On the other hand, the differences between interest rates in different countries surely do not reflect national differences in human behavior, but rather national differences in risk of loss.

Another aspect of Pigouvian "prevision impairment" is human mortality. Money set aside for future enjoyment in the "golden years" may never be enjoyed, if death or incapacity intervene. Meanwhile, savings "for a rainy day" may not be needed. It logically follows from this hypothesis that discount rates for individual humans depend on such variables as life expectancy, health, the amount of insurance protection available and on earning power. In particular, it would imply negative discount rates for individuals with high present incomes and good health but relatively long life expectancy with declining future income expectations, as suggested by Figure 1). This is the time to save money, even if the return is low (or negative), because it will be needed later when current income is lower. Again, the ability to forecast is obviously important: discounting behavior depends on the state of knowledge about the future. For instance, a person who knows he has a fatal disease (say AIDS) may not bother about saving for his retirement. (On the other hand, there is always the possibility of the discovery of a cure; this would change the calculation).

For individuals (at least) the second component of the discount rate, as identified by Arrow, is just another way of saying that the marginal utility of each additional increment of money (or any form of wealth) tends to decline as the quantity of money (wealth) increases. This is one of the most basic behavioral assumptions of economics, and one we do not propose to challenge here.

To reiterate: on deeper scrutiny, both components of the discount rate (as applied to individuals) apparently depend on forecasting: either on the limits to our "telescopic ability" (Pigou) or on actual, if implicit, forecasts of future life expectancy and future income or future wealth. Again,

---

5 Negative discount rates can be inferred, for example, if people put money into savings accounts that pay interest rates less than inflation. This occurred several times in the U.S. during the 1970's.
it is human perceptions of the future that really matter, rather than mechanistic trend extrapolations or expert opinions. (It follows, therefore, that the choice of time preference function belongs in the domain of psychology as much as in the domain of economics).

For managers making decisions in a corporate context, there are two cases to consider in light of the above. One is the manager acting as an individual, making decisions within the corporate context, but reflecting his own personal situation (age, life expectancy, income, wealth, obligations, other options, etc). In this case, the manager optimizes (or satisfices) within this set of constraints. Economists (except organization theorists, who know better) tend to assume without much justification that what is good for the manager is automatically good for the owners, or conversely. The other case is the (hypothetical) manager whose decision is impersonal and reflects only a fiduciary responsibility to optimize the financial interests of the owners of the firm. Such a manager would have to be programmed like a computer or a robot with artificial intelligence, not a human.
For the first case, the "human" manager, the analog of finite life expectancy is finite job tenure. Many managers remain in a particular job — especially at the higher, decision-making, levels — only a few years. Indeed, it has been suggested that short term orientation of U.S. firms may be attributable in part to this fact. If managers are rewarded or promoted on the basis of "bottom line" results, they would have relatively little incentive to invest in projects with a payoff several years in the future. This situation would tend to be associated with a high discount rate and a short-term outlook.

Needless to say, an immortal decision-maker whose interests reflected only those of the (immortal) owners, would *ipso facto*, be more inclined to take a long term view. In other words, for a corporation, the expectation of mortality would not be a factor. It is less clear whether or not a firm *qua* firm (i.e. a firm managed by an impersonal robot dedicated to the interests of the owners) would have a declining marginal utility of money. In the case of humans, declining marginal utility of consumption can easily be explained in terms of finite time for enjoyment, finite physical capacity for food and drink, and so on. (It is not at all clear that the marginal utility of wealth *per se* declines at all. In fact, the behavior of compulsive wealth accumulators in the Maxwell, Milken or Trump tradition suggests that enough is never enough). In any case, there is no direct evidence of declining marginal utility of money, in the case of real corporations.

Still, there is a direct analogy between the individual life cycle, depicted in *Figure 1*, and the 'industry life cycle' confronted by corporations. Some key features of the life cycle are summarized in *Table 1*. However, *Figure 2* tells the important part of the story. Even an immortal corporation (managed by immortals) is likely to face a declining market prospect, sooner or later. The implications of this prospect are also similar to the individual case. When this situation arises, current dollars are worth less (in utility terms) than a flow of dollars some time in the future. It is time to seek a new product or a new line of business. It is the time to bet on risky, long-term, high payoff ventures.

<table>
<thead>
<tr>
<th>Life Cycle Stages</th>
<th>Implicit Discount Rate</th>
<th>R&amp;D Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infancy</td>
<td>High &amp; Positive</td>
<td>Develop the Product</td>
</tr>
<tr>
<td>Childhood</td>
<td>Decreasing Positive</td>
<td>Improve the Product</td>
</tr>
<tr>
<td>Adolescence</td>
<td>Small Positive</td>
<td>Standardize the Product;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Develop the Process</td>
</tr>
<tr>
<td>Maturity</td>
<td>Zero</td>
<td>Improve the Process; Search</td>
</tr>
<tr>
<td></td>
<td></td>
<td>for New Products</td>
</tr>
<tr>
<td>Senescence</td>
<td>Negative</td>
<td>Develop New Product</td>
</tr>
</tbody>
</table>

*Source: [Adapted from Ayres 87 and also Ayres & Mori 89]*
Part of GAP filled:
1) Internally, through technological innovation and new product development and/or
2) Externally, through diversification from acquisitions and mergers, licensing, and/or spinoffs

Sales GAP diagram for a multi-product enterprise
Source: Blackman, 1973

Figure 2. Sales GAP Diagram for a Multi-Product Enterprise
Source: [Blackman 73]

We have argued, above, that the choice of a discount rate or a time-preference function is, in large part, a forecast of the future socio-economic environment. The argument can be extended to a wide range of firm or individual choices, but for the remainder of this paper we restrict the discussion to the special case of technology choice.
Discounting & Forecasting in Technology Choice

The problem we address can be conceptualized as the allocation of a fixed budget among competing investment opportunities. This requires a quantitative comparison of alternative ventures. Note that the question is not whether accurate socio-economic and/or technological forecasts are possible. To raise such a question is to confuse forecasting with prediction. It is clearly not possible to predict discoveries that have not yet been made [e.g. Heertje 83]. Nor can socio-economic behavior "in the large" be predicted accurately beyond the very short term, even in the absence of major innovations or structural changes, due to the non-linear 'chaotic' character of economic systems [Ayres 89]. The fact remains, however, that both savings and investment of all kinds require (and imply the existence of) underlying forecasts. There is no choice in the matter. The question at issue remains: when (if ever) does it pay to select a risky "high payoff" venture that would take a long time to mature, as against a less ambitious "low payoff" venture that would mature quickly?

There are two generic approaches to the comparison of alternative investments. The first consists of three stages (1) to estimate the future time-varying cost and income or profit streams, (2) to discount each future yearly contribution to an equivalent present value (PV) and (3) to add them up. From these numbers one can compute simple numerical measures, such as the benefit/cost ratio or the (expected) return on investment (ROI). The second generic approach is to project a simple annual percentage yield for the projected investment over its target lifetime (allowing for replacement of the capital) and compare it with the anticipated yield of the same amount of money invested in fixed income securities over the same period. Again, a simple ratio can be derived in this case also. Here discounting is not explicitly used, but it is only avoided by the rather crude averaging process which converts a time-varying yield (and cost) function into an "equivalent" annual average. Since this conversion is far from straightforward in cases where either the income stream or the cost stream varies significantly over time, the first method above is normally preferred.

Clearly, there is an explicit role for technological forecasting in project evaluation. What is less clear is how technological forecasting influences discounting behavior. Is there any linkage between them? We argue that the answer is yes. In the first place, we believe that the choice of discount rate does and should reflect perceptions of the external environment — except when discounting is explicit and carried out by "cookbook" formula. To make this notion more concrete, the general competitive environment faced by a firm depends on the stage of the "life cycle" of the product or industry (Table I). The normal situation for a growing and profitable firm is illustrated by Figure 3a.

However, for most firms in most industries (as with individuals) a time comes eventually when rising competition and declining marginal returns on the existing business signal to managers of the need for a new product or service to replace the ones currently providing the bulk of revenue and profits. This situation was shown in Figure 2. It corresponds to Figure 3b. (It is closely analogous to the situation faced by older individuals, depicted in Figure 1). There is plenty of
Figure 3a
Scenario I: Increasing prosperity and growth (Early in the life cycle).

Figure 3b
Scenario II: End of prosperity (Late in the life cycle).

Figure 3. Life Cycle Scenarios
Source: author
evidence that both individuals and firms faced with increasing (declining) expectations behave in a manner consistent with positive (negative) discounting, respectively. Yet, most economists (and accountants) are extremely reluctant to take such a notion seriously. Instead, managers in industry are inclined to argue that large firms "should" pursue a limited number of long term, high risk, R&D projects that are exempted from the rigors of conventional B/C analysis. However, this is an obvious subterfuge: the emperor is wearing no clothes.

Clearly, the economic performance of a firm is likely to reflect the choices made by its managers. In this regard, one must acknowledge the widely shared perception that the "owners" of corporations (e.g. the pension funds and other institutions that dominate the U.S. stock market) seem to demand high current profits and put relatively little value on long-term investments. This is consistent with the fund managers' high discount rates (job tenure is short, memories are short), but it is not necessarily appropriate for the firms whose stock is being traded. In a similar vein, it would seem that excessively mechanistic discounting behavior by U.S. firms — as compared, for instance, to German or Japanese firms — may be attributable to their institutional environments and to the fact that their managers are allowed, or even encouraged by existing incentives such as profit-sharing, to behave in their own personal interests at the expense of the corporate interest.

Having said all this, we want to limit further discussion to the somewhat artificial case of an immortal and impersonal decision-maker (i.e. a firm or a nation) that makes rational economic choices (with one exception, to be discussed in a moment) but that confronts a changing environment. This environment comprises both technological and institutional elements. We will consider, primarily, the former. The corporation (or nation) is assumed to have an "endowment" of business skills, resources or other competitive advantages of some sort. It exists in a market niche, in which it is able to compete effectively. However, just as a human faces finite life expectancy, a firm's 'niche' usually has finite life expectancy, depending on the nature of the product or service it sells, and the existing competition.

While it may not be strictly rational behavior in the traditional economic sense, we assume that the firm or nation (like an individual human) will try to preserve its identity and its integrity, even in cases where an independent and disinterested financial analyst (for instance) might recommend voluntary dissolution or merger. In other words, options that would compromise the integrity of the firm or the sovereignty of the nation are excluded from consideration, hereafter. Given that a firm's competitive ability may not be transferable to different products or services (as many sad experiences with conglomerate mergers have amply illustrated) the above restriction may not be too irrational.

---

6 There is now empirical evidence pointing in this direction. In particular, it has been observed that firms tend to decrease R&D/sales following the implementation of anti-takeover amendments. These results can be taken to imply that managers freed from the discipline of the corporate control market tend to become more myopic [Meulbroek et al 90].
Examples: Waste Disposal and Energy Technology

A plausible, if simplified, contest between two energy technologies was described very briefly above. We now consider the decision problem in more detail. We assume some non-decreasing demand for some service — say hazardous waste disposal or electric energy demand — which is increasingly difficult to meet with conventional technology (e.g. landfill or coal-fired power plants). It is straightforward to forecast that, in the near-term future, there will be a gap between demand for the service and the capacity to supply it by conventional means. The decision-maker must decide how to bridge this gap.

Two alternative technologies exist; call them A and B. Technology A is readily available on a commercial scale and can be put in place at reasonable cost in a relatively short time (call it incineration). However, it is an imperfect, partial and temporary solution to the problem. In the hazardous waste case, the temporary solution might be incineration to reduce the volume of the solid waste to be disposed of; but incineration does not completely eliminate the problem of toxicity, and leaves some risk of air pollution. Thus, it is an interim technology with a finite effective lifetime. In the electric energy case, the interim solution might be nuclear (fission) power. Again, this is not a perfect or permanent solution because (even if the safety problem were completely solved) disposal of radioactive wastes and decommissioning of obsolete plants can never be without risk.\(^7\)

By contrast, we assume that in each case there is a competing technology, B, that is much further from commercialization than A. In fact, in order to implement B on a large-scale it will be necessary to invest in it heavily over many years. Not only that, implementation of A will inhibit the development of B, since B depends on major cost reductions and the creation of viable commercial markets. (Alternatively, we could differentiate the two technologies by assuming technology A is "mature", in the sense of being close to its maximum potential, whereas technology B, while currently inferior, has a much greater potential in the long run). In the hazardous waste case, technology B might be recycling of waste materials (e.g. paper, plastic, glass, etc) on an unprecedented scale. In the electric energy case, technology B might be the large-scale use of photovoltaic "farms".\(^8\)

\(^7\) Moreover, a very large investment in nuclear power would make it very difficult not to shift, eventually, to plutonium "breeder" reactors, with the associated risk of diversion and misuse of fissionable materials for nuclear terror weapons.

\(^8\) To be more concrete, assume that electricity demand is forecast to rise continuously, but that the dominant existing supply technology (coal-fired thermal plants) is likely to be phased out for environmental reasons. This creates a competition between two electric power generating technologies. In this case technology A ("nuclear") is already available for deployment; its characteristics and costs are reasonably well-known. However, with this technology there is not much more to be gained, in terms of lower costs, either from R&D, or from economies of scale or cumulative experience. (In fact, for environmental reasons, costs may rise). Technology B ("solar photovoltaics"), is currently available at fairly high prices, and is much further from large-scale commercialization. The rate at which costs can be brought down depends on the level of R&D support and — more important — on experience, or "learning by doing" in actual production.
In either case, we can assume the decision-maker faces a tight budget constraint. This means that any investment in R&D on B at the present time comes out of the same pot that would otherwise be available for investment in A. Thus, at present prices, some demand for the service in question (hazardous waste disposal or electric power) would go unmet, so prices would have to rise to limit demand to available supply. The key difference between A and B, from a modelling point of view, is that B has a much greater potential for improvement, but that this cannot be realized without large-scale implementation. However, full scale development of and investment in B necessitates tolerating (or subsidizing) excess short term costs during the short and intermediate time periods.

The solution to the decision-maker's problem is a function of his/her time preferences. Such preferences manifest themselves in two ways. First, as described in detail above, the magnitude of the discount rate employed in benefit cost analysis explicitly expresses a time preference. Secondly, the time horizon of decision-making also implicitly parametrizes time preferences. The qualitative solution to the decision problem can be stated in terms of the time preferences employed. The basic result is easily stated: faced with a high (low) positive discount rate and a short (long) time horizon, the planner will select technology A (B) to meet the residual demand.

Exact definitions of 'high' ('low') and 'short' ('long') will depend on the parameters in the actual optimization problem solved. In the intermediate cases—high (low) discounting with a long (short) time horizon, as well as intermediate discount rates and planning periods—the solution will depend intimately on the problem parameters and will almost certainly involve both technologies in some (parameter-specific) combination. We have carried out model calculations of this type, comparing the benefits and costs of 'A' (e.g. "incineration" or "nuclear power") and 'B' (e.g. "recycling" or "solar power") futures.

But now consider the strong time preference (high discount) case: is the technology A "solution" really socially optimal? The choice of interim technology A has (by assumption) inhibited the development of alternative technology B. Once technology A reaches its limits and begins to decline society again faces a crisis. Worse, the alternative technology B would still be insufficiently developed/commercialized to be brought on-line on a large scale immediately. In fact, it will be little, if any, easier to introduce technology B after technology A reaches its limits than it would be to start now on B and leapfrog A.

Thus, based on the assumptions of our model — which we believe to be realistic — society will confront a crisis situation either sooner or later. The real choice is not whether, but when, to face up to it and pay the costs of implementing B. The advantage of waiting is that short term costs will be lower and the money "saved" (be deferring the development costs of B) could be used...

\[9\] For instance, for the recycling technology to be effective still requires the creation of efficient, large-scale collection and distribution systems and viable markets for secondary materials. In the case of photovoltaic power, low costs depend on accumulating manufacturing experience, which depends on large-scale implementation.
for other beneficial purposes. The disadvantage of waiting is that avoidable environmental damages and/or avoidance costs associated with the interim technology (for instance, mining and dissipative use of toxic metals or the transport and storage of radioactive wastes) are also accumulating all the while. Furthermore, the delay in adopting technology B also means that its long term benefits (lower costs) are deferred.

In the model examples (and others of a similar nature), the inferior technology A may be 'locked-in' indefinitely, if it is not self-limiting for some reason. For instance, given a high enough discount rate, the technology with initially lower costs (A) will tend to attract incremental investment, thus gaining economies of scale and increasing its cost advantage over technology B. There are also accumulated investments in infrastructure, training, and organization external to technology A per se. These are sometimes termed economies of adoption. In any case, the existence of such investments and linkages constitutes a barrier to change [see, for example, David 85; Arthur 88].

From a social planner's perspective, one way to avoid "lock in" of an inferior technology would be to constrain the optimization problem such that the social welfare is non-decreasing, not just during a fixed time horizon, but over the indefinite future. This notion has recently been employed as one possible formalization of the notion of economic sustainability [Pezzey 89, 89a]. In our model, such a constraint (assuming a solution exists) would require at least some investment in technology B independent of the discount rate. This strategy would avoid the possibility of locking-in an inferior, short-lived technology. However, in the absence of a central planning authority it is not clear how to implement an equivalent constraint legislatively — or otherwise.

Incidentally, there is some justification in the psychology literature for employing hyperbolic discounting, in contrast to the usual exponential function of time [e.g. Herrnstein & Mazur 87]. In brief, there is empirical evidence that humans tend to discount faster in the very short run than the conventional (compound interest) model indicates, but slower in the very long run. Hyperbolic functions can be parametrized to yield this kind of behavior. While the long run behavior of hyperbolic functions is to approach 0 asymptotically, they do so at a much slower rate than the exponential ones. As applied to our model decision problem, hyperbolic discounting values technology B, relative to A, slightly more than exponential discounting does. Thus, some continuing investment in B is more likely to be optimal in this scheme. See Appendix I.

The usual form of the so-called "experience curve" is expressed as a fixed percentage reduction in unit costs for each doubling of cumulative experience. Based on this model, the only way to

---

10 There are plausible cases where the costs of implementation of B can be expected to increase over time. For instance, it was relatively inexpensive to acquire land to build railways and tramways in the 19th century. Land acquisition for any new large-scale national ground transportation system — such as the high speed "MagLev system" that has been developed in West Germany (but not implemented) — would be prohibitively expensive in the U.S., Japan or Europe today.
Figure 4. Parameters of the Experience Curve for Various Industries
get the costs of technology B down to presently competitive levels is to accept temporarily higher costs and invest in actual production and deployment. The faster the buildup, the more rapidly costs will decline. Of course, there is no absolute guarantee that costs can be brought down at any given rate or by any given degree. One can only point to historical experience with other industries (Figure 4). Some representational computation results are shown in Appendix II.

Evolutionary Implications

It is well-known that the conventional (compound interest) method of discounting tends to emphasize short term costs or benefits in relation to longer range ones. Evidently this can easily 'tilt' project selection in favor of projects with lower but more immediate benefits, even if this selection results in "locking in" an inferior choice of technology. (To be clear, we mean a choice that will appear inferior to future generations).

Suppose, however, that two competing firms, or two competing nations A and B have different time-preference functions. Assume that firm or nation A consistently chooses the option with higher immediate benefits but lower long-term benefits. The other (B) chooses options with lower short term benefits but greater long-term benefits. Type A can be characterized as short-sighted or myopic, while type B can be characterized as far-sighted or presbyopic. Two outcomes are possible. It could happen, if competition is completely unrestrained, that firm A is able to drive firm B out of business before the long-term benefits of B's technological choices are realized.

One indicator of type A behavior is high current profits and high dividends. In a truly open, competitive marketplace, such firms should preferentially attract equity capital, resulting in lower capital costs than their type B competitors. This would permit expansion of manufacturing facilities, resulting in lower unit costs, increased demand (due to price-elasticity effects) and eventual market dominance. In this case, type A will be the long-run survivor.

On the other hand, if A does not, or cannot, use its short term advantage to remove B from the field, then B will eventually be the winner — and the long term survivor. For instance, if type A firms are inhibited by anti-trust regulations from exploiting their short-term advantages to the limit, the effect is to give type B firms a protected 'niche' in which some of them can thrive. Something like this occurred in the U.S.long distance telephone market, where AT&T was prohibited from monopolizing its critical discoveries in semi-conductor and laser technology. Later, AT&T was forced to allow competitors in the long-distance market to underprice it.

The US-Japanese competition, in general, follows a similar pattern. It is natural to categorize U.S. firms, in general terms, as 'type A' and Japanese firms as 'type B'. Japanese firms were (and still are) protected from foreign competition in their own market by a variety of devices, from tariffs to 'standards'. They were free from takeover threats, due to a web of interlocking ownership relations coordinated by large banks and trading companies. They were also given exclusive access to the Japanese capital market, at very favorable rates. They were (and still are) largely...
free of the pressure to show immediate "bottom line" results. Theoretically, this might have led to inefficiency. But, in practice, it has enabled Japanese firms to invest very heavily in building up long-term advantages in human resources, organization, and manufacturing capability — even basic science and technology — that will be increasingly hard for type A firms outside of Japan to overcome.

In the real world of anti-trust regulation and sovereign nations able and willing to protect their national firms, type A firms may not be in a position to drive out type B firms while they still have the technological edge. In this case, the short term strategy has no real chance of ultimate success in the evolutionary competition. To conclude, the far-sighted (presbyopic) strategy looks increasingly like the long-term evolutionary survivor.

The discussion of competitive dynamics (Type A vs. Type B) in the last few paragraphs implicitly assumes that the choice of time preference is determined once-for-all, e.g. by inheritance. On the other hand, our thesis in this paper has been that the choice of time preference function is at least partially determined by external circumstances. In fact, we argue that both Type A and Type B behaviors are, to some extent, self-denying expectations. The logic is quite straightforward: assume an economic unit starts with very high (and justified) expectations for future market growth and increasing wealth. This will lead to a very high discount rate (Type A), which will result in an investment strategy focussed exclusively on low risk, short term improvements to the existing product or technology. During this phase, the myopic strategy is appropriate.

However, such a strategy (at the firm level) will inevitably — sooner or later — encounter the problem of market saturation and/or declining marginal returns. Revenue growth will slow down, profitability will peak and decline. Eventually a superior product or technology will come along from the outside and undermine the firm’s formerly secure position. At this point the firm faces the sort of problem illustrated by Figure 2 and Figure 3b. Expectations of growth and increasing wealth falter. Discount rates fall. If the management is alert, it begins to look around for promising long-range projects with high potential payoff. If the search is efficient, and management is reasonably lucky, an innovative and profitable venture will be found to replace (or supplement) the former business. A new growth cycle starts.11

11 The alert reader will recognize that, at the aggregate national level, this sequence is consistent with the so-called Kondratieff cycle, of alternating periods of high and low growth for the economy as a whole. In the latter case, market interest rates will tend to rise and fall more or less in synchrony with average discount rates.
REFERENCES


Appendix

I. Hyperbolic Discounting

In order to account for the value of some economic quantity over time, it is orthodox in economics and related fields to discount future values relative to the present. Noting time by 't', $V(t)$ as the value stream to be discounted, and $d(t)$ as the discount function, the the usual definition of the present value of the value stream is

$$\int_{0}^{T} d(t) V(t) \, dt$$

A particular functional form for the discount function is nearly universally employed, the so called exponential form which we note as $d_e(t)$ and write as follows:

$$d_e(t) = \exp(-rt) = \beta^t$$

where the expression on the extreme right-hand side (RHS) is to be commonly found in modern writings on the subject.

Other forms for $d(t)$ are, however, imaginable. The only properties that we require of $d(t)$ are that:

1. $d(0) = 1$, that is, no discounting of the present;
2. the function is monotone non-increasing--the value of $\$1$ in the future can never be greater than $\$1$--or perhaps the slightly stronger condition of being monotone decreasing--the value of $\$1$ in the future must be less than $\$1$.

There exist, of course, an infinite number of non-exponential functions which satisfy these two modest properties. It would be useful if there was some theoretical or empirical basis in economic agents' actual behavior for selecting one functional form over all others. Practically, the reason why the exponential form is employed virtually always and everywhere is that (discrete) interest payments paid during a particular (discrete) time period are calculated as a fixed percentage of extant balances; and when this notion is formalized mathematically and transformed from discrete to continuous time the exponential function results.

However, that fact that cash balances and (relatively) short-term loans are subject to exponential discounting does not seem to imply that all items of economic value should be discounted in this way. While it is clear that if the value of a particular commodity is readily obtainable in cash at a

---

1 All calculations described herein have been executed using the Mathematica ® software package; they are available from the authors in the form of a Mathematica Notebook.

2 Note that these two requirements do not preclude the possibility of having $d(t) = 1$, i.e. no discounting at all.
market-determined price then the value of the commodity—insofar as the cash so obtained may be invested at a positive interest rate—is properly dealt with over time through exponential discounting, it is far from clear that all pecuniarily-relevant items should be so discounted. Environmental resources are significant examples of economically important quantities which, since no markets exist for them (or those markets which do exist are notoriously 'thin'), should perhaps not be exponentially discounted. Or, stated somewhat differently, there is at least no theoretically unambiguous reason for discounting such resources exponentially. Other functional forms may be more appropriate in such cases but until recently there has been very little discussion in the literature on theoretical reasons for employing non-exponential forms. Thus, even such non-market resources as recreation services in National Parks in the United States are currently valued over time with exponential discounting, and the active research question becomes one of eliciting the 'appropriate'—in some more or less well-defined sense—discount rate.³

Recently, there has appeared a first attempt to justify the use of an alternative discount function [Herrnstein and Mazur, 1987]. These authors suggest that the psychological process of discounting is such that a hyperbolic form is more reasonable. They argue that while discounting in the short term can be quite significant, i.e. over the course of weeks or months, individuals do not continue to maintain high discount rates into the long-off future (years, decades). In comparison to the exponential discount function, the hyperbolic form has precisely these properties: larger discounting in the short run, reduced long run discounting. This functional form can be written in terms of one adjustable parameter, p, as

\[ d_h(t) = \frac{p}{p+t} \]

Clearly, this form satisfies the two conditions described above.

To compare the shape of \( d_h(t) \) with \( d_e(t) \), and thus relate the different ways in which these functions effect discounted values, it is necessary to obtain some sort of equivalence between the two parameters in the functions, p and r. That is, given a particular discount rate, call it \( r^* \), what is the value of p, say \( p^* \) such that the hyperbolic form behaves in some similar way to the exponential discount function? One possible definition of equivalence might involve the areas under the functions. However, insofar as the integral of \( d_h(t) \) involves the \( \ln(\cdot) \) function, which is not bounded as t approaches infinity, while the area under \( d_e(t) \) is finite, such a comparison cannot be executed generally. Instead of this, we shall define equivalence in terms of the two functions actually being equal at some particular time, T, the crossover time, thus implicitly defining p as a function of r (and T) as \( d_h(T; p) = d_e(T; r) \). In fact, it is possible to obtain p explicitly as

³As exemplified by the contingent valuation (CV) paradigm currently so popular in the applied environmental law and economics communities.
\[ p = \frac{T}{\exp(rT) - 1} \]

This functional relationship is plotted below with \( p \) along the (primarily) vertical axis, \( r \) noted as \( 1/x \) with \( x \) varying from 2 to 40 (that is, \( r \) varying from 2.5% to 50%) along the (principally) horizontal axis, and \( T \) varying from 1 to 25 along the dimension into the page:

\( p \) as a function of \( r \) and \( T \)

Figure A.I.1

This graph is useful for quickly approximating \( p \) from prescribed \( r \) and \( T \).

As a direct comparison between \( d_h(t) \) and \( d_e(t) \), the two functions are shown graphically below for \( r = 0.10 \) and \( T = 10 \), with \( p \) calculated according to the expression above.
The hyperbolic function is less than the exponential one for all time less than 10, while it lies above for all greater time. This suggests, as has already been described, that the hyperbolic function gives relatively greater discounting for short times and less as time gets large. Note that the difference in the areas under the curves before the crossover time—i.e. the area between the curves—is, in this case, just about equal to the difference in areas under the curves after the crossover time up through about \( t = 20 \). That is, the areas under the two respective curves is about equal at 20 time periods, with \( d_e(t) \) having greater area before twenty, but \( d_h(t) \) being larger thereafter. If the time units are years, as we shall always consider them to be in what follows, then for this choice of parameters \( (r = 0.10, T = 10) \) we can expect that investment decisions having lifetimes on the order of 20 years may not be very sensitive to which discount function is employed.\(^4\) However, for other sets of parameters the choice of discount functions may make a great deal of difference in the final results.

For example, for \( r = 0.10 \) as before and \( T = 5 \), we have, as before, \( d_h(t) \) being below \( d_e(t) \) early on, but then lying above it for all time greater than 5. This is shown graphically in the following figure:

\(^4\)Actually, if the investment stream, \( V(t) \), being discounted changes its shape significantly around the crossover time (10 years in the present case) then the particular form of \( d(t) \) employed may indeed change the results.
In this case, for relatively long-lived investment projects, $d_h(t)$ would give very different discounted values from $d_e(t)$ since the former is so far above the latter for large $t$. In particular, $d_h(t)$ puts much greater weight than does $d_e(t)$ on that part of the value stream which occurs late in the life of the project.

This example points to an important structural difference in the properties of $d_e(t)$ and $d_h(t)$. As $t$ approaches infinity, $d_e(t)$ approaches 0 at such a rate that, as alluded to earlier, the area under the curve is finite. However, this is not the case for $d_h(t)$, which has a first integral involving $\ln(t)$ and thus the area under the function is unbounded as $t$ increases. The implications of this for questions of valuation over time is as follows: with exponential discounting there always exists some time beyond which future value flow contributes only negligibly to total discounted value, while this is not the case for hyperbolic discounting.

We have considered only one hyperbolic functional form above, although other hyperbolic functions are, of course, feasible discount functions. We define a hyperbolic form of order $m$ as

$$d_{hm}(t) = \frac{p}{p+t}^m$$

and, of course, $d_{h1}(t) = d_h(t)$ as previously defined. In general, the shape of any particular hyperbolic form vis-a-vis $d_e(t)$ is similar to the relation of $d_h(t)$ to $d_e(t)$, i.e. there always exists a crossover time, $T$, such that $d_{hm}(t)$ is below $d_e(t)$ for all time less than $T$ and above it for all larger times. The general effect of larger $m$ is to give the hyperbolic form more curvature, thus as $m$ increases the area between the exponential and hyperbolic functions increases.
An Analytical Basis for Hyperbolic Discounting: When Agents/Institutions Possess a Spectrum of Exponential Discount Rates

We have suggested a psychological basis for the use of hyperbolic discounting, but this is a qualitative motivation and cannot resolve, for example, the use of one hyperbolic form over another. In this section we present a quantitative basis for the use of hyperbolic discounting.

Imagine that an economic agent, attempting to account for the time value of a particular economic stream, has not one but \( n \) distinct discount rates, \( r_i \), each of which is given relative weight \( \rho_i \) in the construction, by linear aggregation, of a composite discount function, i.e.,

\[
d(t) = \sum_{i=1}^{n} \rho_i \exp(-r_i t)
\]

It is easy to imagine examples of such a discount function such as when:
- a particular agent expresses a variety of different private discount rates over time, with \( d(t) \) being some sort of time average of these rates;
- an institution is indebted (equally) to a variety of different sources at several different interest rates; or
- a public good yields its value over time and each citizen values it according to her private discount rate.

When \( n \) is very large, as in the latter case, the discount function can be approximated by replacing the summation with integration as

\[
d(t) = \int_0^{\infty} R(r) \exp(-rt) \, dr
\]

where \( R(r) \) is a spectrum of discount rates and serves as a weighting function in the integral. This expression is, in fact, a generalization of the summation insofar as the discrete expression is obtainable from the continuous one through the use of the a weighting function of the form

\[
R(r) = \sum_{i=1}^{n} \rho_i \, \delta(r - r_i)
\]

where \( \delta(\cdot) \) is the Dirac delta function.

In order for a function to be a discount function it must possess the two properties described above. First, it must produce unit discount at time zero, i.e.
d(0) = 1 \Rightarrow \int_{0}^{\infty} R(r) \, dr = 1

There are, of course, an infinite number of functions which possess this property, including, for example, all probability density functions. Second, d(t) must be monotone non-increasing in t. It turns out that d(t) can be shown, by Bernstein’s theorem [Feller: 439], to have an even stronger property, that of complete monotonicity, whenever R(r) is greater than zero for all r, as it will always be if it is to have economic significance.

A particularly interesting class of functions satisfying both of the requirements on R(r) is

\[ R_m(r) = \frac{p^m r^{m-1}}{(m-1)!} \exp(-pr) \]

where m and p are parameters. When these functions are put into the discount function integral there results the hyperbolic form of order m with parameter p, that is

\[ d_m(t) = \int_{0}^{T} \frac{p^m r^{m-1}}{(m-1)!} \exp(-pr) \exp(-rt) \, dr = \frac{p^m}{(m-1)!} \int_{0}^{T} r^{m-1} \exp\left[-(p+t) r\right] \, dr = \frac{p^m}{(p+t)^m} \]

For a particular value of m we can be more concrete. With m = 1 we have the inverse exponential function,

\[ R_1(r) = p \exp(-pr) \]

appropriately weighted by the parameter p so that the area under the curve is unity; it is shown below for several values of the parameter:

---

5A function, f(x) which is completely monotone has the property that its derivatives alternate in sign, i.e. if f(x) ≥ 0 then f'(x) ≤ 0, f''(x) ≥ 0, and so on.

6Note that d(t) can be regarded as the Laplace transform of R(r) in the transform variable t. Although we shall not make direct use of this fact, suffice it to say that a large storehouse of theorems from Laplace transform analysis apply directly to d(t).
This distribution describes an economic agent having all discount rates, but considering larger ones to be always less important than smaller ones, with mean of $1/p$. Plugging this into the discount function

$$d_1(t) = \int_0^T p \exp(-pr) \exp(-rt) \, dr = p\int_0^T \exp[-(p+t) \, r] \, dr = \frac{p}{p+t}$$

we obtain $d_1(t)$. Thus, we find that the hyperbolic form can be put on something other than ad hoc grounds.

For the $m = 2$ case one obtains a single-peaked discount rate spectrum, i.e.

$$R_2(r) = p^2 r \exp(-pr)$$

which has the following graphical representation for several values of $p$:
Figure A.I.5

Such a distribution has a mean of $2/p$ and might represent an agent who keeps some medium term discount rate in primary view but also values both very small and very large rates to some lesser extent. When substituted into the discount function there results the hyperbolic form of order 2:

$$d_2(t) = \int_0^T p^2 r \exp(-pr) \exp(-rt) \, dr = p^2 \int_0^T r \exp(-(p+t)r) \, dr = \frac{p^2}{(p+t)^2}$$

This function has the general shape of the hyperbolic form, as described above. For $m > 2$, single peaked spectra always result, with mean $m/p$ and producing, as shown above, hyperbolic discount functions of order $m$.

However, the $m > 2$ hyperbolic functions have more in common with the exponential discount function than does the hyperbolic function of first order. This is so because the area under the second and higher order functions is finite, and thus value streams discounted with theses forms long into the future will obtain a stationary value, as is the case when $d_2(t)$ is used.
II. An Energy Investment Decision

Imagine two energy technologies, perfect substitutes, competing for market share and long run hegemony. One is an old, entrenched process, part of a technologically-mature industry--call it A and think of it as fossil fuel or perhaps nuclear power--while the other is a new technique, unestablished as yet and still in a state of rapid technological development, say B (read solar). In the long run forecasts predict that A will be displaced by B but it appears to be very difficult to predict when this will happen with any degree of confidence and/or precision.

Presently, B costs more per unit of electricity generated than does A, but this cost will surely decline as manufacturers' learn how to make it more efficiently and users learn how to integrate it into their processes better. Such learning is assumed to have already taken place for technology A. It is usual to model this incorporation and assimilation of knowledge into a particular production process through the use of the experience curve, commonly written as

\[ C(t) = A Y(t)^m \]

where \( C \) is the cost, \( Y \) is the cumulative production, \( t \) is time and \( A \) and \( m \) are parameters; usually \( 0.2 < m < 0.5 \). This relationship is shown graphically in the figure below for several values of \( m \) and with \( A \) set to unity:

![Cost as a function of total production, various m](image)

Figure A.II.1

Note that the cost never really plateaus, rather it continues decreasing at an ever-decreasing rate.

We say that technology A is 'entrenched' insofar as it is much further along its experience curve than is B. When we say that in the long
Run B will replace A we mean that \( m_A < m_B \), and there will thus come a
time when the cost of technology B will be less than the cost of A. There will
come a time, that is, if agents have enough foresight to realize B's long run
advantage and thus begin the process of moving away from A. Individual
agents will not, however, be able to alter demand significantly enough to
move either technology very far along its experience curve, and thus it is
only societally rational and not individually rational to adopt technology B.

In attempting to determine whether or not to adopt the more
expensive technology B in order to reduce long run costs, an individual
economic agent (firm or consumer) must discount the future cost stream.
The usual way to do this is through the exponential discount factor as
follows

\[
\int_0^T C(t) \exp(-rt) \, dt
\]

Call \( D_A(T) \) the discounted cost for technology A as a function of time
horizon, \( T \), similarly for \( D_B(T) \). Once this quantity is calculated for each
technology, that with the lowest discounted cost will, presumably, be
selected. We have executed this calculation for model technologies A and B,
with \( m_A = 0.4 \) and \( m_B = 0.5 \), assuming that each technology can satisfy a
demand (assumed fixed) of one unit per year, that 50 units of technology A
have been produced to date,\(^7\) and using cumulative production as a
surrogate for time. Whether \( D_B(T) \) is greater or lesser than \( D_A(T) \) is a
strong function of the discount rate, \( r \), employed and the starting position
assumed for technology B. The following figure shows \( D_B(T) - D_A(T) \) for \( r =
0.10 \) by letting the starting point of B vary.

\[\text{Discounted Cost Difference vs. B Starting Point}\]

\[\text{Figure A.II.2}\]

\(^7\)Thus technology A is already in the 'flat' portion of its experience curve.
Thus we find that for technology B sufficiently far down its experience curve— at least 20 or so units produced— \( DB(T) < DA(T) \), while if B is 'too new' the opposite occurs. Such a comparison rendered for any actual technologies will be intimately related to the parameters employed.

Now we undertake a calculation of \( DB(T) - DA(T) \) by allowing the total demand to vary. Since, by assumption, one unit is demanded per year, increasing the demand is equivalent to lengthening the time horizon. The results, for \( r = 0.05 \) and 0.10, are summarized in the figure below:

The lower discount rate results in relatively lower costs for technology B in the long run since the long time behavior is given relatively greater weight for small \( r \); thus the lower curve in the figure corresponds to \( r = 0.05 \). Overall, from the figure we find that \( DB(T) > DA(T) \) up through 10-12 years of time horizon, but the opposite condition obtains for all future time. Thus, for a sufficiently far-sighted society, technology B should be adopted at presently, i.e. at \( t = 0 \).

In the spirit of appendix I above, we might ask how hyperbolic discounting would change matters. In this case we would compute the discounted cost as

\[
\int_{0}^{T} C(t) \frac{p}{p + t} \, dt
\]

with \( p \) set to correspond to some discount rate and crossover time as read, for example, from the three dimensional plot of appendix I, above. The following figure gives a direct comparison of the exponentially discounted solution and the hyperbolic one, for two discount rates:
The hyperbolically discounted costs are smaller than the exponentially discounted ones. This is so because the hyperbolic discount function gives greater weight to future events, and it is in the future that technology B becomes relatively less expensive than technology A. Thus we find that, while we have designed the problem such that technology B is to be societally preferred to technology A, it may or may not be individually rational for particular economic agents to adopt technology B over A. In effect, technology A, insofar as it is used presently, is 'locked-in'.