FACOLTA' DI ECONOMIA
UNIVERSITA' DI BOLOGNA
SEDE DI FORLI

Percorso di Studi in Economia Sociale

Know Thyself: Self Awareness and Utility Misprediction in Discounting Models of Intertemporal Choice

Luca Savorelli

Working Paper n.57
Dicembre 2008

in collaborazione con

AICCON
CULTURA COOPERAZIONE | NONPROFIT

Luca Savorelli,
Università di Bologna
Università di Siena

First version: September 16, 2008
This version: December 19, 2008

Informazioni
Facoltà di Economia – Percorso di Studi in Economia Sociale
Tel. 0543 374673 - Fax 0543 374660 – e mail: segreteria.ecofo@unibo.it
Web: **www.ecofo.unibo.it**

KNOW THYSELF:
SELF AWARENESS AND UTILITY MISPREDICTION IN DISCOUNTING MODELS OF INTERTEMPORAL CHOICE

Luca Savorelli

First version: September 16, 2008
This version: December 19, 2008

Abstract

This review critically discusses the theoretical models of discounting through a selection of papers. It will focus on the comparison between the two major models, namely discounted utility and hyperbolic discounting. This paper differs from previous surveys in the attention given to the interpretation of models and suggestions for future research. First, it stresses how the economics literature is biased itself, since it considers only present-biased preferences on the basis of not well grounded evidence, while also future-biased preferences are worthwhile of attention to address relevant issues, like compulsive behaviors and work-life unbalance. Second, it makes qualifications on the meaning of self-awareness, distinguishing between cognitive boundaries and beliefs. Third, it suggests future direction of research to address this issue, fundamental for a sound policy analysis, by showing how the two main welfare criteria used in the literature to evaluate welfare of inconsistent individuals (long-run utility and Pareto criterion) have relevant shortcomings. Finally, it proposes a new perspective on misprediction of utility based on time perception.

JEL No. A12, D11, D90.
Keywords: future – biased preferences, intertemporal choice, partial naivete, self – control, sophistication, time inconsistency.

Luca Savorelli
Università di Siena
E-mail: savorelli@unisi.it

CONTENTS

Introduction

Who would prefer one hundred Euros to one hundred twenty, if they were delivered at the same moment? It is not a hazard to speculate that no one would. On the contrary, if one were asked to choose between an earning of one hundred Euros now and a higher sum next year, for instance one hundred twenty Euros, many people would be indifferent or would prefer the one hundred Euros offer. This fact suggests that individual choice is dependent on time peculiarities.

Economics has been explaining this phenomenon mainly by models of discounting. In these models the individual maximizes the intertemporal utility functional form:

$$U^t(c_t,...,c_T) = \sum_{i=0}^{T-t} D(i)u(c_{t+i}) \quad [1]$$

where $u(c_{t+i})$ represents the per period cardinal utility function of the individual and *D(i)* represents a discounting function, which associates to each time i a weight, attached to each per period utility function.

For example, in the so called discounted utility model (DU, Samuelson, 1934), the specification of *D*(*i*) is the following exponential function:

$$D(i) = \left(\frac{1}{1+\rho}\right)^i \quad . \quad [2]$$

Thus, assuming that $u(c_{t+i})$ is linear in the amounts considered, in the initial example people who are indifferent between 100 Euros now and 120 Euros next year exhibit a discount rate ρ =0.2.

Nevertheless, [2] is just one among the possible specifications of *D(i)*. Indeed, behind its formal elegance, the DU model bears many shortcomings, from the point of view of both normative and descriptive validity. In addition, psychological foundations and evidence do not seem to support heavily the DU model. Finally, there are several ways other than discounting in which time can influence individual behavior.

Excellent surveys about intertemporal choice and departures from the standard rational choice model already exist in the economics literature. In a very comprehensive review on intertemporal choice Frederick *et al.* (2002) examine the major characteristics of DU model, existing anomalies with respect to it and models addressing them, provide an historical framework, discuss both experimental and field evidence and techniques for measuring discounting. Tirole (2002) concisely reviews and discusses the major departures to rational choice models, giving particular attention to rational ignorance, memory, willpower and personal rules. O'Donoghue and Rabin (2002) review and extend the main results on quasi-hyperbolic discounting by means of simple models. From a different perspective, Frey and Stutzer (2006) consider the interrelation between utility misprediction (as individual errors) and happiness research to empirically evaluate people well-being, suggesting applications to smoking, obesity and watching TV . Finally, Della Vigna (2007) widely surveys anomalies in decision making focusing on field experiments.

The aim of this paper is to discuss models of discounting, with a particular focus on the comparison between the two major models: discounted utility and hyperbolic discounting. I am not surveying evidence on this models for two reasons: first, the reader can refer to excellent works as Frederick *et al.* (2003) and Della Vigna (2007); secondly, empirical results on intertemporal choice are not convergent in time and across studies (Frederick *et al.*, 2003), suggesting that a deductive approach to the problem based on everyday observation is still valid and useful.

This paper differs from previous surveys in the attention given to interpretation of models. First, it stresses how this literature is biased itself, since it considers only present-biased preferences on the basis of not well grounded evidence. On the contrary, I will argue that future-biased preferences are also worthwhile of attention and that research on this topic could address important topics, e.g. compulsive behaviors and work-life balance. Second, it makes qualifications on the meaning of self-awareness, distinguishing between cognitive boundaries and beliefs. Finally, it suggests future direction of research to address this issue, so much important for policy analysis, showing that the two main welfare criteria used in the literature to evaluate welfare of inconsistent individuals (long-run utility and the Pareto criterion) have relevant shortcomings.

I will proceed as follows. In section 1 I will show how time can affect individual choice and discuss whether discounting is an independent construct or whether it is confounded with other factors. In section 2 I will compare the DU, HD and quasi-HD models, introducing future-biased preferences. In section 3 I will discuss self-awareness, its role in misprediction of utility and criteria for welfare evaluation. In section 4 I will conclude with further suggestions for future research.

1 Factors influencing intertemporal choice

Time can affect individual choice in several ways. They can be sorted in two categories, depending whether they influence the expected *amount* of future utility or the *weight* of utility per se (Frederick *et al.* 2002). According to Frederick (1999), the expected amount of future utility can be influenced by

1) *Probability*: if the probability of the occurrence of an event which confers utility depends on time, then it is reasonable to evaluate differently the utility conferred to the same event at different in time points. Thus the maximization function can be written as
$$U^t = p(c_0)u(c_0) + ... + p(c_T)u(c_T) \quad [3]$$
The probability can be increasing or decreasing with time. A special role can be played by the probability of being alive (see section 1.3).
2) *Characteristic of the consumed good*: time can affect the characteristic of the good itself. A classic example is aging of wine or whisky. A correct formulation of the problem would take into account the characteristics of the good as follows:
$$U^t = u(c_0(t)) + ... + p(c_T)u(c_T(t)) \quad [4]$$
3) *Changes in the utility function*: the consumption of a good can confer different utility due to different taste changes.[1] For a classical discussion of the role of stable preferences see Stigler and Becker (1977). For instance, if I have just started learning Russian, the utility deriving from reading War and Peace it is going to be much lower with respect to the utility I would have after twenty years of study as a Russian – English translator:
$$U^t = u_0(c_0) + ... + u_T(c_T) \quad [5]$$
As Frederick (1999) notices, this holds also on smaller time scale, as the consumption of goods depends on the level of recent consumption (see section 1.2 for a discussion of connection between diminishing marginal utility and discounting).
4) *Anticipation*: the anticipation of the occurrence of an event that provides future utility can confer itself utility or disutility in the present. It is possible to enjoy the anticipation of a negative event (e.g. the last days of vacation before going back to school (Frederick, 1999)) or to suffer waiting for a positive events (e.g. waiting for a date). Sometimes anticipation can confer greater utility

[1] For a classical discussion of the role of stable preferences see Stigler and Becker (1977).

than the event itself.[2] It is important to distinguish between the utility conferred by anticipating the events and the change in utility derived from changing current consumption, because of the integration of expected future utility in consumption choice. Anticipation can also cause to delay or anticipate the consumption of the good.

$$U^t = u_0(c_0, u(c_1), ... u(c_T)) + ... + u_T(c_T) \quad [6]$$

5) *Memory*: The explanation advanced by Frederick (1999) to observation that people would like to enjoy a thing sooner, because they are going to enjoy its memory in the future for an higher number of years, seems quite weak. Memory is likely to change future consumption when the future utility deriving from memory is integrated in current choice. This is mainly done between consumption of goods that can keep track of memory, e.g. taking photos or writing a diary. For instance, every time we take a picture of a sunset we weight the pleasure of looking at the sunset without being bothered to pick up the camera with the utility we will experience in the future by looking at the photo and recalling that sunset. The intertemporal utility function taking into account the memory can be written as:

$$U^t = u_0(c_0) + ... + u_T(c_T, u(c_0), ... u(c_{T-1})) \quad [7]$$

6) *Opportunity costs* If instead of consuming goods we save money and invest it, a certain amount today can be exchanged for a higher amount in the future. For example, if you invest 100 Euros today at a market rate of 10%, next year you will have 110 Euros. Of course, this holds true only if the investment is effective, otherwise there is no reason for discounting its future value:

$$U^t = u_0(c_0) + ... + u_T(c_T, investment_0, ..., investment_{-1}) \quad [8]$$

The last way in which time can influence choice is *time discounting*.[3] Time discounting pertains to the weight given to the entire per period utility function. The formulation of a discounting model is based on none of the preceding reasons. What is postulated in any model of discounting, be it exponential or not exponential, is that the same good, coeteribus paribus (no change in utility function, not conferring utility from memory or anticipation neither having characteristics depending on time and being certain in consumption), confers less utility if consumed in the future rather than in the present.

[2] A more extreme, but valuable, perspective, is given, for example, by Giacomo Leopardi, one of the most important Italian poets, who argues in his "Village's Saturday" that enjoyment can be experienced only in the anticipation of events, not in the event itself. Interestingly, Piera degli Esposti, an Italian theatre actress, proposed a humoristic perspective on Leopardi's poem, based on backward induction logic: if enjoyment can be experienced only in the anticipation, then Saturday's enjoyment can be experienced only on Friday, and Friday only on Thursday, and so on.

[3] Of course all these factors can be combined. For example, time influences tasting wine by modifying the flavor of the good and providing utility from anticipation the sommelier waiting to taste an old wine. But also the decision of opening a certain aged bottle can be influenced by impatience.

There is no agreement on the nature of *time discounting*. Historically, the existence of discounting has been motivated by a series of psychological and non psychological reasons (for a summary see Frederick *et al*. 2003) which influence the intertemporal behavior: the uncertainty of life, the prospect of immediate consumption, the self-restraint and the bequests motive (Rae, 1934), the anticipation of future utility (Jevons, W. 1888), the role of abstinence (Senior, 1836), the ability of foresight and, last but not least, fashion (Fisher, 1930). The recent economics literature justifies discounting on the basis of an intertemporal multiple-selves perspective: the individual is made up of several selves, each following the other on the time line. Discounting is therefore justified on the basis of the psychological connectedness among these persons and is seen as a form of intra-personal altruism[4]. In this perspective, the weight attached by Self *i* to the utility of the other future selves is analogous to the weight she attaches to other people. The discounting function *D* is usually decreasing in *i*, but as we will see in section 2.1.3, there is no particular reason for not considering non-monotonic discounting.

While discounting models have been widely used since Fisher (1930) and Samuelson (1937), there is lack of strong evidence in supporting them. Frederick *et al.* (2002) advance doubts on the existence of the "discounting construct", and on the basis of Lowenstein (2001) propose to step back to the seminal 19th century literature and "unpack" time preference in three constituent motives: impulsivity, compulsivity and inhibition. Thus, justification for discounting has been found among psychological factors.

Nevertheless, some authors link discounting to decreasing marginal utility and probability. We will now briefly review and comment on these perspectives.

1.1 Discounting and decreasing marginal utility

Consider the following example. An individual's linear per period utility function has the form $u=x$, where x is the amount of money he can earn. Let's consider the functional form [1] and the discounting function [2]. If the individual is indifferent between 100 Euros and 200 Euros one year hence, this means that his discount rate is 100%. On the contrary, suppose that the utility function is concave, and not linear, for example it has the form $u = x^{1/2}$. In this case, if the individual is again indifferent between 100 and 200 Euros, then the discount rate is no more 100%. Since $u(100) = 10$ and $u(200) = 10\sqrt{2}$, $10 = \left(\frac{1}{1+\rho}\right) \cdot 10\sqrt{2}$ implies $\rho \cong 0.41$. Thus, even if from an observational perspective the "discount" rate in both cases is 100%, in the second case only 41% is to ascribe to *time*

[4] This view comes from the work of the philosopher Derek Parfit (Parfit, 1971).

discounting, while the remaining part depends on the concavity assumption. Moreover, diminishing marginal utility works also if the two choices are not sequential in time, differently from time discounting.

Opposite to this view, Rachlin (1992) explains diminishing marginal utility as a direct *consequence* of discounting. For example, if a person eats one apple every five minutes, the tenth apple would be eaten 45 minutes afterwards and its utility value would be discounted with respect to that of the first apple. Rachlin (1992) argues also that physiological satiation is not sufficient to explain decreasing marginal utility, since, in presence of satiation without discounting, everyone would space consumption so as to maximize his utility and diminishing marginal utility would not be observed. Firstly, this argument is weak, because usually is not possible to space out consumption at pleasure. Secondly, Rachlin (1992) recognizes that discounting could not account for all kinds of diminishing marginal utility (for example, quantity of sugar in a cup of tea). Thirdly, the argument is quite circular, since Rachlin assumes discounting as an empirical fact, while there is no agreement whether the construct of pure time discounting exists or whether it is a derivative phenomenon. Anyway, this discussion suggests that, when specifying examples like the one at the beginning of this paper there would be the need to specify preference orderings *without* discounting.

1.2 Discounting and life expectancy

Probability, in the particular form of life expectation, could be considered as the base for discounting (Stange and Summer, 1978). Apart from the correct Frederick (1999, fn 5) technical criticisms of their approach, a probability function of survival can be used as well in place of a discounting function. If we include probability considerations maximizing the expected value of the intertemporal utility function, it is reasonable to include the probability of being alive and of enjoying utility.

If the source of discounting would be probability, then discounting could not be anymore considered irrational anymore (see section 3.2.2).

2 Models of discounting: exponential and hyperbolic

Several economics models have been written to explain how time influences choice, mainly in a deductive way starting from introspection and everyday life. The majority of the economics literature uses of the DU model because of its appeal and tractability, but the discounting function in equation [2] can assume other different forms. The most widespread alternative to exponential discounting is hyperbolic discounting. In section 2.1 I will point out characteristics of the two models, and in section 2.2 I will show peculiarities of exponential discounting and in section 2.3 peculiarities of hyperbolic discounting.

2.1 Exponential and hyperbolic models, shared features

The class of models that assumes either exponential or hyperbolic discounting is characterized by different features. The first one is the underlying idea that the individual is made up of different selves. By metaphor, a person is analogue to a relay team, in which after his run each player passes the baton to the following player. This idea is often implicit in exponential models and explicit in hyperbolic discounting models. Then, in equation [2], *U* can be thought as the utility of person *t* (self *t*), which is influenced by the utility of other people (self *t+1*, self *t+2*…). Therefore, this model can be viewed as a particular model of altruism, in which the index of the per period utility function not only represents a different person, but also the time at which the person lives. Then, the discounting function values are the weights given to the other future people. As a consequence, the less the present person takes care of each of them, the more distant in the future the person "lives".

The second characteristic common to the two models is that, continuing the metaphor, all players are twins. In other terms, the per period utility function is stationary (and concave). Interestingly, even the discounting function is stationary (this is a necessary, but not sufficient condition for consistency, as we will see).

In addition, both models imply that the discount function is independent of what the person is going to discount: the discount function for apples is not different from the discount function for pears. Another feature shared by the two discount functions is that the discount rate ρ is positive in both models.

Lastly, in both models when the individual suddenly faces a new option of choice, he recalculates *all* his consumption profile taking into account of the new option, and chooses it only if the overall intertemporal utility is greater than integrating the option in his basket.

2.2 The DU model

The DU utility model has been described by equations [2] and [3]. Equation [2] can be expressed in continuous time as follows

$$U^t(\{c_\tau\}_{\tau\in[t,T]}) = \int_{\tau=t}^{T} e^{-\rho(\tau-1)}u(c_\tau) \qquad [9]$$

but for easiness I will refer to equations [2] and [3]. The most important feature of the DU model is that it implies time consistency. Function [3] is stationary and the discount rate ρ is constant, since the ratio *D(i)/D(i+1)* is constant, which implies $\rho_i = \rho, \forall i \in (t,...,T)$. Stationary discounting and constant discounting rate are necessary and sufficient conditions for time consistent preferences (Frederick *et al.* 2003). In the framework of a multiple selves approach, this means that later selves respect planned future choices of earlier selves. In other words, it is as if each runner's baton contained instructions for the other player on how to run his quarter and the other player followed strictly those instructions. Said it plain, the individual is going to do everything he earlier planned to do. This happens because with constant discounting a parallel shift in the dates of two events can not influence the preference accorded to them. An alternative way to look at consistency is given by this example. Suppose to ask me any day if this evening I prefer to listen to a live classical concert or to attend a basketball match. Suppose also that I am indifferent to the two options. Then you ask me my preference between a live classical concert one evening and a basketball match a week later and I answer that I prefer going to the music concert. In the DU model it is not important *when* this choice is made. You can ask my preference on December 10, 2008 or on March 27, 2020 and the answer is going to be the same. A way to express my preferences formally, calling X the option "to go to classical concert" and Y "to go to basketball match", is $X_i \succ Y_{i+d}, \forall i \in (t,...,T)$.

2.3 The hyperbolic discounting models

The HD discounting function can be written in this form:

$$D(t) = (1+\alpha t)^{-\frac{\rho}{\alpha}},\ \alpha > 0, \rho > 0 \qquad [10]$$

This function implies that the discount rate is declining over time. Thus here the discount rate is a *function* of time, the ρ in [10] not to be cofounded with that in [9]. Calibrating the parameter α of the model it is possible to shape qualitatively different discounting profiles. As α approaches 0 the function becomes exponential, $D(t) = e^{-\rho t}$. On the contrary, a high value of α causes the function to decline sharply in the near future, while afterwards the decline will be slower. For example (Bowles, 2004), you can be very impatient of having your next meal, but you can be less impatient at evaluating consequences due to global warming. The function is stationary, but the behavior of the individual is time inconsistent, since the discount rate is not constant. Going back to our metaphor, the instructions within the baton are not going to be followed by the succeeding runners. This means that a model of HD allows for preference reversal, too. For example, it can happen that at date *t* the individual prefers at t+10 outcome X over outcome Y, while approaching t+10 (for example, at t+8) the preference is reversed, becoming favorite Y over X.

Instead of using a hyperbolic discounting function, the literature on declining discounting often adopts a functional form known as "quasi-hyperbolic" function, introduced by Phelps and Pollak (1968) and widely used since Laibson (1994) among economists

$$D(i) = \begin{cases} 1 & if\ i = 0 \\ \beta\delta^{i} & if\ i > 0 \end{cases} \qquad [11]$$

The characteristic of this functional form is that the discount rate between period i and period $i+1$ is *different* from the discount rate between period $i+1$ and period $i+2$, but it is constant between any two future period since $i+2$. In the case $0 < \beta < 1$, the discount rate between i and $i+1$ is $\delta_i = \frac{(1-\beta\delta)}{\beta\delta}$, while between any two future period is $\delta_{i+k} = \frac{(1-\delta)}{\delta}$. Since $\delta_i > \delta_{i+k}$, that is the discount rate is *declining* with time, this kind of preferences are "present – biased" and the individual is said to be *impatient*.

Some qualification is needed. Although it is rarely explicitly mentioned, with some exception (Krusell, 2001), mathematically the functional form of "quasi-hyperbolic" discounting is not hyperbolic at all, as it can be seen simply looking at [11]. On the contrary, this functional form can be said more properly "quasi – geometric", because it differs from the geometric one just for the first period. The function has been denominated "quasi – hyperbolic" because, for $0 < \beta < 1$ it mimics qualitatively the behavior of decreasing discount rate in the hyperbolic case. The quasi- geometric function can be made "more hyperbolic" by adding one or more parameters in the functional form [12], for example (Diamond and Koszegi, 2003: Appendix D):

$$D(i)=\begin{cases}1 & if\ i=0\\ \beta\gamma\delta & if\ i=1\\ \beta\gamma^{2}\delta^{i} & if\ i>1\end{cases}$$

This makes discount rates different between the first and the second period, and between the second and the third period, constant afterwards. On the other side, calibrating the β parameter, analogously to α in [10], allows to shape qualitatively different discounting profiles. For small values of β, the decline in the discount rate is very sharp, while for $\beta=1$ the quasi-geometric discount function reduces to the exponential [3].

2.3.1 Future-biased preferences

The economics literature has considered only the case of present-biased preferences, with $0<\beta<1$. Probably, the reason is that everyday observation concerning intertemporal choices, like addictions, seems to exhibit an undervaluation of future consequences.

While, at first sight, it could be reasonable to postulate that a person weights his selves less, the farer in the future they are, this evidence in not well grounded. As Frederick *et al.* (2003) notice, the measure of discount rates across experiments (with real or hypothetical rewards) and field studies is highly variable. Moreover, variability and methodology in measurement seem not to have improved along years, since the range of results has not been shrinking over time. A great part of the studies they examine shows a discount rate less than one, which can be consistent also with values $\beta>1$.[5] Moreover, even though a different sample of studies show that the discounting rate on the average is declining with time, Read (2001) found no declining rates and two studies (Frederick, 1999 and Rubinstein, 2002) found *increasing* discount rates.

The following Frederick (1999) statement makes clear the idea that preferences are not necessarily present-biased:

> [The determination of] the degree of concern [a person] has for "her" future welfare might also depend how much she currently esteems future values and interests. She might, for example, ascribe more weight to "her" future welfare if she is told that she will become obsessed with chess. … [On the contrary] if we knew we were to be transformed into a frog, we might well be relatively unconcerned about further tortures applied to that frog.

[5] For example, if $\beta=1.1$ and $\delta=0.9$, for $i=1$ the overall value of the discounting function is 0.99.

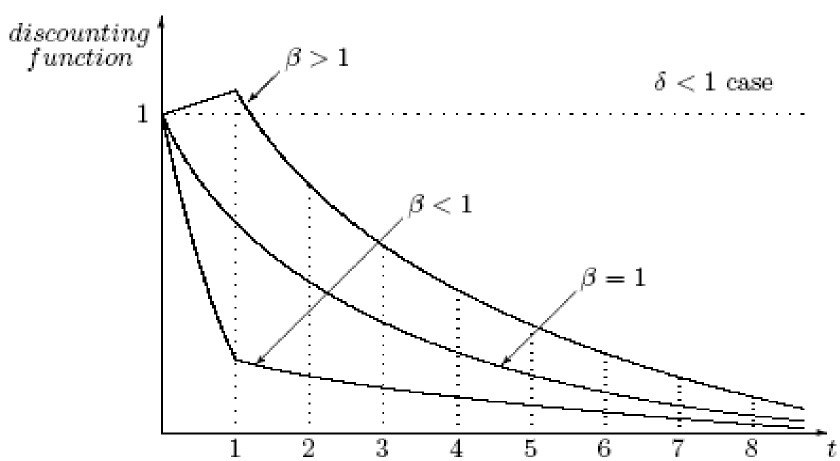
This is somehow different from changing tastes, since it is a matter of *weighting* future selves having in mind predictable changes in identity or in the budget constraint. There are many situations in which persons over-sacrifice their current pleasure in favor of future goals, which they weight too much. The feeling associate to this fact is regret. For example, many have experienced, after getting a good grade after an exam, the feeling that they could have studied less for the same result. As Kivetz and Keinan (2006) notice, the research on self-control

> … is premised on the notion that people are shortsighted (myopic) and easily tempted by hedonic "sins," such as overbuying (oniomania), splurging on tasty but unhealthy food, and indulging in luxuries … importantly suggests that people not only yield to temptation they had originally planned to resist but also subsequently reverse their preference and regret their myopic behavior. … While yielding to temptation can certainly be harmful, overcontrol and excessive farsightedness (hyperopia) can also have negative long-term consequences.

This overweighting of future consequences could involve central aspects of one's life, as career choices and life-work balance and could be at the basis of compulsive behaviors.

The economics literature has not been to attentive to regret and future-biased preferences. To our knowledge, no one has treated them, even though they are explicitly mentioned by Krusell (2001) presenting the quasi-geometric model, and by Salaniè and Treich (2006), showing that oversaving is possible also for hyperbolic discounters, independently of whether preferences are future or present-biased.

The following figure, taken from Krusell (2001), shows the behavior of the quasi-geometric function for $\beta > 1, \beta = 1, \beta < 1$.



As the figure shows, present-biased preferences and future-biased function are *not* symmetric.

2.4 Sub-additive discounting

Read (2001) explains this fact in terms of subadditive discounting. Discounting is subadditive when the value of the discounting function is decreasing in the length of the unit interval, i.e. the calculation of discount over an interval is greater if made in two steps calculating subintervals. On the contrary, hyperbolic discounting is not subadditive, but additive, i.e. the discounting over an interval is independent of its partition. Anyway, if decision are taken from the present subadditive discounting is equivalent to hyperbolic discounting (Frederick *et al.*, 2003), since from the present later dates implies larger intervals of time. Read (2001) punctually observes that subadditive discounting can be the source for hyperbolic discounting and that both hyperbolic and subadditive discounting can be at work. Noticeably, in his experiments he does not find declining rates and suggests to drop HD as a model of human behavior.

2.5 Summary

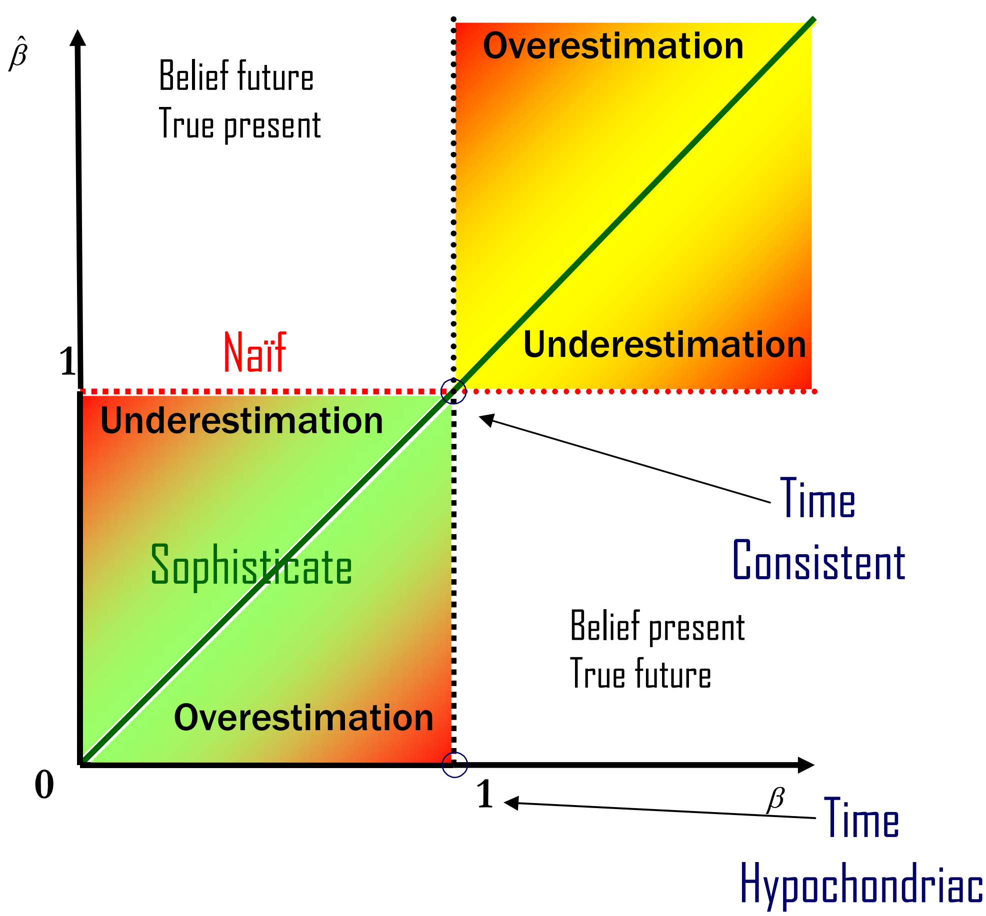
Up to now we have seen how the DU and the HD discounting models share several features but, holding the hypothesis of rationality (i.e. maximizing behavior) the HD discounting model introduces qualitative possibilities in economics models via non-constant discounting rate: inconsistency of behavior, misprediction of utility and identity. The following sections explore the role of self-knowledge and identity in consumer maximization problems.

3 Utility misprediction and self-knowledge: naivete, partial naivete and sophistication.

Economics models of HD assume that there is an objective "someone" and a "someone" believed to be. Then, the choices are undertaken by a person on the basis of the preferences he *believes* to have, but the utility he experiences depends on your objective utility function. This often results in misprediction of utility. In reality, HD could be just one among several sources of misprediction of utility (see Frey, 2006; Frederick *et al.* 2003; Read, 2001).

In the framework of quasi-HD models, people are perfectly aware of their per period utility function. This means, for example, that if a person's utility functions is homogeneous and accounts from memory (see [6]) , at time *t*, *relatively to all the other goods at t+1,* he correctly forecasts how much he is going to enjoy watching at his vacation pictures at time *t+1*. What he does not know exactly of yourself is the value of the parameters of his discounting function. Thus, the right interpretation of the quasi-hyperbolic model (but also of DU) is that people look at future utility in a diminished scale, but without modifying the relative gains in utility each good confers in the future.

Thus, the individual is associated to a couple of parameters $(\beta, \hat{\beta})$, where the first is the "true" value of the parameter, while the "hat" is the value the individual believes the parameter has. Thus, self-knowledge can be usefully represented in the $(\beta, \hat{\beta})$ space, as in the following figure.



Individuals can be grouped in four sets: *time consistent* with $\hat{\beta} = \beta = 1$ (point [1, 1]); *naïf*, with $\hat{\beta} = 1 \wedge \beta \neq 1$ (the line $\hat{\beta} = 1$); *sophisticated*, with $\hat{\beta} = \beta \wedge \beta \neq 1$ (the diagonal with the exception of point [1, 1]); *partial naïf* with $\hat{\beta} \neq \beta \wedge \hat{\beta} \neq 1$.[6] It is noteworthy that also individuals with $\beta = 1$ can have a believed value $\hat{\beta} \neq 1$, meaning that they are in a measure "time hypochondriac". The partial naïf can either undervalue or overvalue its true problem.

3.1 The maximization problem

The way in which the individual solve his maximization problem depends on whether $\hat{\beta} \neq 1$ or $\hat{\beta} = 1$.

3.1.1 Sophistication, partial naivete and partial sophistication

In the first case, $\hat{\beta} \neq 1$, the individual believes not to be time consistent, and that he can predict what his future selves are going to do. Then, he is a sophisticate and solves the problem by backward induction (or, equivalently, he finds a subgame-perfect equilibrium in a game among the selves). The major effect of sophistication is to make future selves choices consistent with earlier selves. As we will see in the next section about welfare evaluation, having a sophisticate behavior does not necessarily imply an improvement in individual's welfare. Moreover, the effects of sophistication depend on whether the individual is a partial naïf (O'Donoghue and Rabin, 2002) and on the characteristics of the economic environment.

As O'Donoghue and Rabin (2002) suggest, this approach is not very realistic, since also very aware people can have cognitive problems at solving all the round of backward induction and knowing how their preferences look like in the last period. Moreover, as for example Diamond and Koszegi (2003) notice studying the effect of quasi-hyperbolic discounting on retirement decision, when the number of periods is large, the individual often faces a very complicated intertemporal consumption schedule and it is not possible to say something general about the equilibrium path. To address this issue, O'Donoghue and Rabin (2002) suggest adopting a boundedly rational approach according to whom the individual performs just a small number of periods of backward induction. Thus, for example, the individual considers preferences in *t+2* and hence starts the backward induction. Therefore, this approach eliminates what they call "sophistication effects of higher order", meaning that the individual does not take into account the reaction of Self *t+2* to behavior of future selves. In addition, the

[6] In fact the naïf set is a subset of the partial naïf set.

individual needs to choose how often to compute again his consumption path. We could call this behavior *partial sophistication*.

The partial naïf solves the problem exactly in the same way as the sophisticate, but since he has a wrong belief about his value of β, he fails to predict the utility his future selves will experience, believing that they will behave like the sophisticate, but with a parameter $\hat{\beta}$. Of course, individuals can be partial naives and partial sophisticates at the same time.

3.1.2 Naivete and time consistence

In the second case, $\hat{\beta}=1$, each self *t* maximizes in a straight way the following intertemporal utility function (subject to his budget constraint):

$$U^{t}(u_{t},u_{t+1},...,u_{T})=u_{t}+\beta\sum_{\tau=t+1}^{T}\delta^{\tau-t}u_{t}\ ,\ \mathrm{t}=1,\dots,T \qquad [12]$$

believing that his future selves will be consistent with his choices. If $\beta=1$, they will and the behavior of the individual will be *time consistent*. If $\beta\neq 1$, they will not, and the individual will exhibit a naïf behavior, not taking into account any strategic consideration (because the naïf believes his future selves will be consistent with his choices).

A feature of this framework, usually not explicitly addressed, is that either future selves have no memory or there is no possibility of learning. Self *t* thinks that all his future selves will be consistent with him, although all his former selves were not consistent with earlier selves. This is in fact a commonly observed behavior, e.g. when people claim that tomorrow they will give up smoking, although they said it thousand of times in the past without giving it up.[7]

To conclude this section, a qualification on the interpretation of naivete and sophistication is needed. In fact, in this model people are naïf only with respect to the *value* of their β. They way in which they solve the maximization problem (backward induction or straight solution) is a consequence of having a certain belief $\hat{\beta}$, which in the case $\beta=1$ *implies* consistency of all future selves. Nevertheless, if the same naïf person were told to be in error, and that her value $\hat{\beta}\neq 1$, she would

[7] Another famous example is taken from "Zeno's Cosciousness" by Italo Svevo, where the protagonist writes many times on the calendar "L.S" beside date of the day after, meaning "last cigarette".

maximize in the same way as sophisticates. Implicitly, O'Donoghue and Rabin (2002) do not share this idea, since they suggest interpreting partial sophistication as an approximation of the naïf behavior. Future research should keep separate naivete arising from bounded rationality to naivete arising from misbelieves about one's preferences: self-control problems arising from cognitive problems at forecasting future behaviors could be addressed by very different policies from those aimed at solving self control problems arising from a scarce knowledge of preferences.

3.2 Criteria for individual welfare evaluation

The HD discounting framework is good at describing in a more realistic way people's behavior, while the analysis on the normative side is much more weaker. The two main criteria used in the economic literature are the long-run utility and the Pareto criterion, but both bear important shortcomings.

The long-run utility criterion adopts as optimal benchmark for welfare Self 0 utility valued at the solution of its maximization problem. The individual's long-run utility function is

$$U^0(u_1, u_2, ..., u_T) \equiv \beta \sum_{\tau=t+1}^{T} \delta^{\tau-t} u_t$$

The idea underlying choice of long-run utility is that discounting is rational (see a short discussion in section 3.2.2). The reasons to prefer this approach to welfare evaluation is preferred are several. Firstly, it is justified if decisions of the kind "I will quit smoking at time *i*" are repeated, because each earlier-than-*i* self has to agree on the statement. For example, all selves before *i* usually agree at quitting smoking, but Self *i* does not care and keep on smoking (Koszegi, 2005; O'Donoghue and Rabin, 2006).

Secondly, some authors consider departures from DU as mistakes (Bernheim and Rangel, 2004; Koszegi, 2005; Tversky and Kahneman, 1983, O'Donoghue and Rabin, 2003, 2006). On the contrary, Frederick *et al.* (2003) question the fact that DU "anomalies" are mistakes. This is to our knowledge an open field of research with respect to hyperbolic discounting. There is some experimental research on this topic which investigates departure DU other than HD (magnitude effect, Frederick and Read, 2002; preference for improving sequences, Loewenstein and Sicherman, 1991) and supports the idea that departures from DU are not anomalies and have normative value.

Thirdly, Gruber and Koszegi (2001) found that the closer the consumption level of future selves to that planned by earlier selves, the higher the discounted utility is for all selves. This idea connects the long-run utility to Pareto optimality, too.

The Pareto criterion applied to intertemporal choice (Goldman, 1980; Laibson 1997) states that the normative benchmark for welfare analysis is given by the consumption schedule preferred by all selves. The Pareto criterion is importantly weak, since in the case of multiplicity of equilibria there is the need to introduce another criterion to discriminate among them, and which does not allow ranking of Pareto optimal outcomes.

Nevertheless, the Pareto criterion overcomes a shortcoming of long-run utility criterion: since the same individual evaluates his well being differently at different points in time, in general there is no particular reason to consider Self 0 instead of Self *i* utility as the benchmark. In fact, for each self, his current choices are maximizing. Considering the same example in O'Donoghue and Rabin (2002), suppose that now the discounted utility of watching Sleepy Hollow tomorrow is equal to "2", lower than that of watching Ed Wood the day after, equal to "3". When tomorrow the individual is deciding whether to spend his money to watch Sleepy Hollow this evening or Ed Wood the following evening, it happens that, due to proximity, Sleepy Hollow confers him a utility equal to "4" and his preference reverses. Not only the choice is reversed, but also, in absolute value, he is going to enjoy Sleepy Hollow more than his yesterday self predicted he would have enjoyed Ed Wood. Thus, why should his idea of enjoyment today be used to value tomorrow choices?

To answer these important shortcomings of both Pareto and long-run criteria, probably useful suggestions could be found by considering criteria adopted in other field of research, e.g.. bioethics, investigating other domain of choice in which facts are apparently more clear cut, as it is for suicide or serious illnesses. The typical example is that of Ulysses-and-the-Sirens kind, namely at time t the individual writes down that at t+1 in case of serious disease that makes him loose the fundamental human capabilities he will go on living, while at t+1 he declares that he wants stop living. What Self should be taken into consideration at evaluating welfare? Do previous choices count more than present choices? And if so, why? For a bioethical approach to the problem, among different approaches, see Beauchamp and Childress (2008).

3.2.1 The role of self-awareness

O'Donoghue and Rabin (1999, 2000) show that, when considering simple choice environments with present-biased preferences, e.g. "when to do an activity", sophistication exacerbates misbehavior for pleasurable activities, while it mitigates misbehavior for onerous ones. The naïf behavior is symmetric, while the partial naïf behavior is characterized by a hedge value of $\hat{\beta}$, under which the partial naïf behaves like a sophisticate and over $\hat{\beta}$ like a naïf (O'Donoghue and Rabin, 2002).

On the contrary, in more complex environments, the effect of awareness is not so straightforward. In a simple three-period consumption-saving example, O'Donoghue and Rabin (2002) show that for three different CRRA utility functions the effects of sophistication (and partial naivete) are not straightforward. In particular, in one case it exacerbates over-consumption, in the other it mitigates it and in the third case it has no effect. Within the CRRA family, the entity of the effects of awareness is decreasing in the level of relative risk aversion. Risk aversion determines their direction, too: if it is less than one, awareness worsen over-consumption; if it is more than one, it mitigates over-consumption, while if the function is logarithmic, it has neutral effects (Pollack,1968).

From a different perspective, Tirole (2002) considers an agent who has to decide whether to perform a simple task and can acquire information at no cost. Also in this case information about one's characteristic does not necessarily have a positive value. Indeed, the quasi-HD model allows the agent to decide to be rationally ignorant.

3.2.2 Is discounting rational?

So far I have taken for granted that discounting, be it DU or HD, is rational. Frederick (1999) lists many authors (among them: Jevons (1888), Pigou (1920), Rawls (1971), Ramsey (1928)) contending the rationality of discounting. Stigler and Becker (1997) show that their discussion of stable preferences implies no discounting, and question the importance of time preference. Without entering the discussion, it must be noted that the rational basis for discounting depends on its source (psychological myopia or other). For example, if discounting is a construct derivative life expectancy (see section 1.2) it can be well considered rational.

In the case in which discounting would not be rational, the intertemporal utility function with no time preference should be used as an optimal benchmark. Nevertheless, so far the DU utility model has been taken as the normative benchmark for welfare analysis.

4 Concluding remarks and future research

The aim of this paper was to make some qualification on the existing literature on intertemporal choice, with a particular focus on discounting models. I chose to survey (hyperbolic) discounting models because, notwithstanding their shortcomings, they are parsimonious in not explaining phenomena with *ad hoc* extensions (as many models do introducing new arguments in the utility function), and in leaving room for normative analysis. Besides, they are not a mere translation of psychological literature in economics.

I discussed whether discounting can be seen as independent construct, or whether it can be derivative of other factors, stressing the role that life expectancy can play and the distinction of discounting from decreasing marginal utility. After comparing the major features shared by hyperbolic discounting and discounted utility models and their peculiarities, I discussed the introduction of future-biased preferences in order to extend the existing literature on self-control to "economics of regret". Then, I qualified the role played by self-awareness. I distinguished the misprediction of utility due to bounded rationality from that due to bad knowledge of preferences, suggesting that the nature of misprediction is an important factor at evaluating policies. I then concisely examined the two major criteria for welfare evaluation of inconsistent agents, namely the long-run utility and the Pareto criterion. Since these criteria bear important shortcomings, I suggested that they should be seriously challenged by future research if the literature on intertemporal choice is aimed at providing useful policy indications. In fact, evaluation of people welfare has been a difficult issue also in other domains, e.g. medicine now faces many bioethical issues. A possible direction of future research could look at how bioethics solved (or tried to solve) individual welfare analysis. Such a comparative analysis could suggest useful criteria for evaluating welfare also in the economic choice domain. Finally, I analyzed the role played by self-knowledge, showing how information on oneself does not always have a positive value. Future research should address *when* information has a positive or negative impact on people welfare: is the normative statement "know thyself" valid only when entering Delphi's temple?

To conclude this paper, I suggest a radical departure from models of discounting that anyway can be entailed in this framework. The literature on discounting is based on the premise on the fact that individuals are myopic or hyperopic at evaluating *future utility*. The fact that people can be myopic or hyperopic in *time perception* itself is never mentioned, but everyone has perceived at least once time as never passing (e.g. when listening to a boring classical music concert) or as passing too fast (e.g. on vacation). Along the same line it is common perception in life that years are passing faster when one gets older with respect to youth and childhood. This "subjective" time perception can be as well the source of misprediction of utility, if discounting is based on dishomogenous time intervals. For example, tomorrow Sleepy Hollow show can seem closer than one day if tomorrow is a vacation day rather than if tomorrow is a working day. Economics could embed its techniques of multiple selves analysis on the important work of the 21st century philosopher Henry Bergson, *Time and free will*. In his work Bergson (1889) explains human being as a two selves construct: the first Self, which regards time as homogenous, is a "shadow" Self of the real Self, which regards time as

> …a real duration whose heterogeneous moments permeate each other; below the numerical multiplicity of conscious states, a qualitative multiplicity: below Self with well-defined states, a Self in which succeeding each other means melting into one another and forming an organic whole…The Self thus refracted and thereby broken to pieces is much better adapted to the requirements of social life in general and language in particular: consciousness prefers it, and gradually loses sight of the fundamental Self.

References

Beauchamp, T. L., Childress, J. F. (2008), *Principles of Biomedical Ethics*, Oxford University Press, New York.

Bergson, H. (1889) *Time and Free Will: An Essay on the Immediate Data of Consciousness*, tr., F.L. Pogson (1910), Kessinger Publishing Company, Montana.

Bernheim, D., Rangel A. (2001), Addiction, Conditioning, and the Visceral Brain, Stanford U.

Bowles, S. (2004), *Microeconomics: Behavior, Institutions, and Evolution*, Russell Sage Foundation, Princeton.

Della Vigna, S. (2007), Psychology and economics: evidence for the field, Working Paper, University of California, Berkeley.

Diamond, P., Koszegi, B. (2003), Quasi-hyperbolic discounting and retirement, *Journal of Public Economics*, 87, 1839 - 1872

Fisher, I. (1930), *The Theory of Interest*, Macmillan, New York

Frederick, S., Loewenstein, G., O'Donoghue, T. (2002), Time Discounting and Time Preference: a Critical Review, *Journal of Economic Literature,* 15: 351 – 401.

Frederick S. (1999), *Discounting, time preference and identity*, Ph.D. dissertation.

Frey, B. S., Stutzer A., (2006), What Happiness Research Can Tell Us About Self – Control Problems and Utility Misprediction, IZA Discussion Paper No. 1952.

Goldman, S. (1979), Intertemporally Inconsistent Preferences and the Rate of Consumption, *Econometrica* 47:3, 621-626.

Gruber, J., Koszegi, B. (2001), Is addiction 'rational?' Theory and evidence, *Quarterly Journal of Economics,* 116:4, 1261-1305.

Jevons, William S. 1888. *The Theory of Political Economy*. London: Macmillan.

Kivetz, R., Keinan, A. (2006), Repenting Hyperopia: An Analysis of Self-Control Regrets, *Journal of Consumer Research,* 33, 272-286

Krusell P., Kuruscu, B., Smith A. A. (2001), Equilibrium Welfare and Government Policy with Quasi-Geometric Discounting, *Temi di discussione del Servizio Studi*, Banca d'Italia, n.413.

Koszegi B. (2005), On the Feasibility of Market Solutions to Self-Control Problems, *The Economic Council of Sweden's conference on "Regulating Unhealthy Consumption"*.

Laibson, D. (1994) *Essays in Hyperbolic Discounting*, Ph.D. Thesis, Massachusetts Institute of Technology.

Laibson, D. (1997), Golden Eggs and Hyperbolic Discounting, *Quarterly Journal of Economics,* 112, 443–77.

Loewenstein, G., Sicherman N. (1991), Do Workers Prefer Increasing Wage Profiles?, *Journal of Labor Economics,* 9:1, 67–84.

O'Donoghue, T., Rabin, M. (1999), Doing It Now or Later, *American Economic Review,* 89:1, 103-124.

O'Donoghue, T., Rabin, M. (2000), The Economics of Immediate Gratification, *Journal Behavioral Decision Making,* 13:2, 233-250.

O'Donoghue, T., Rabin, M. (2001), Choice and procrastination, *Quarterly Journal of Economics,* 116:1, 121 - 169.

O'Donoghue, T., Rabin, M. (2002), Self Awareness and Self Control, in *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. George Loewenstein, Daniel Read, and Roy Baumeister, eds., Russell Sage, New York.

O'Donoghue, T., Rabin, M. (2003), Studying optimal paternalism, illustrated by a model of sin taxes, *American Economic Review,* 93:2, 186-191.

O'Donoghue, T., Rabin, M. (2006), Optimal sin taxes, *Journal of Public Economics*, 90, 1825-1849.

Parfit, D. (1971), Personal Identity, *Philosophical Review*, 80:1, 3–27.

Pigou, A. C. (1920), *The Economics of Welfare,* Macmillan, London.

Pollak, R. A. (1968) ,Consistent Planning, *Review of Economic Studies,* 35, 201–208.

Rachlin, H. (1992), Diminishing marginal value as delay discounting, *Journal of the Experimental Analysis of Behavior*, 57:3, 407-415.

Rae, J. (1834), *The Sociological Theory of Capital* (*reprint 1834 ed.*), Macmillan, London.

Ramsey, F.P. (1928), A mathematical theory of saving. *The Economic Journal*, 38, 543-549

Rawls, J. (1971), *A Theory of Justice*, Harvard University Press.

Read, D. (2001), Is time discounting hyperbolic or subadditive? , *Journal of Risk Uncertainty,* 23, 5-32.

Rubinstein, A. (2000), Is It 'Economics and Psychology'? The Case of Hyperbolic Discounting, Tel Aviv U. and Princeton U.

Salanié, F., Treich, N. (2006), Over-savings and Hyperbolic Discounting, *European Economic Review,* 50, 1557-1570.

Samuelson, P. (1937), A Note on Measurement of Utility, *Review of Economic Studies*, 4, 155-161.

Senior, N. W. (1836), *An Outline of the Science of Political Economy*, Clowes & Sons, London.

Stigler, G. J., Becker, G. S. (1977), De gustibus non est disputandum, *The American Economic Review,* 67:2, 76-90.

Strotz, R.H. (1956) Myopia and inconsistency in dynamic utility maximization, *Review of Economic Studies.* 23:3, 165-180.

Tirole, J. (2002), Rational Irrationality: Some Economics of Self-Management, *European Economic Review,* 46, 633 -655.

Tversky, A., Kahneman D. (1983), Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological Review,* 90, 293–315.