Commitment and Welfare*

Frank N. Caliendo†
Utah State University
frank.caliendo@usu.edu

T. Scott Findley
Utah State University
tscott.findley@usu.edu

This Version: May 14, 2014. First Version: May 2, 2014.

Abstract

Welfare analysis under dynamically-inconsistent preferences is routinely criticized for being ad hoc or arbitrary because it is often based on the preferences of the time-zero self. We provide a counter argument. We show that the time-zero consumption-saving plan of a hyperbolic consumer, which we call the commitment allocation, is viewed by *all selves* as a better outcome than the consumption-saving path that is actually followed. Because we allow for the full range of assumptions about backward looking preferences, we resolve Bernheim and Rangel's (2009) concern about past studies that sought to invoke a similar Pareto argument. In the terminology of Bernheim and Rangel, we show that the commitment allocation *strictly and robustly multiself Pareto dominates* the actual allocation. Hence, welfare analysis that is based on the preferences of the time-zero self may not be all that arbitrary after all.

Key words: Behavioral Welfare Analysis, Commitment, Time-Zero Preferences, Dynamic Inconsistency, Hyperbolic Discounting.

JEL Classification: D91, C61.

*We thank Charlie Sprenger and participants at the Bay Area Behavioral and Experimental Economics Workshop at Stanford University.

†Corresponding author.

1 Introduction

The determination of individual welfare when preferences are dynamically inconsistent has been a key challenge for behavioral economics. The fundamental question is how to define a welfare improvement when all the intertemporal selves of a single individual seem to disagree on resource allocation. Brocas, Carrillo, and Dewatripont (2004) and others have argued that, while a Pareto criterion would be an ideal welfare measure, it is usually difficult or perhaps even impossible to satisfy given the disagreement among selves. Therefore, as a practical alternative, behavioral economists often assume that the goal of a policymaker is to maximize the utility of the time-zero self (e.g., O'Donoghue and Rabin (1999, 2000, 2001, 2003)).[1] The vast hyperbolic discounting literature is a prime example.

However, such behavioral welfare analysis is routinely characterized as ad hoc or arbitrary (Bernheim and Rangel (2009)). Although the typical approach seems intuitive—why not help individuals follow the plans that they themselves initially wanted to follow before they fell victim to self-control problems—critics argue that there is no obvious reason to favor the preferences of the time-zero self at the expense of all the other selves who, by their very actions, will choose different consumption-saving allocations if they are not already pre-committed to the original plan. For instance, Gul and Pesendorfer (2004, p.263) point out that welfare analysis based on the time-zero plan "has the planner forever guarding the perceived interests of the nonexistent former selves," which they later describe as "odd" (Gul and Pesendorfer (2008, p.38)). Likewise, Rubinstein (2006, p.248) says "One criticism made of behavioral economics is the arbitrariness of the welfare criterion...why should the utility of the first self be the basis for welfare considerations?"[2] And Brocas, Carrillo, and Dewatripont (2004, p.51) state that "there is no normative foundation" for equating welfare with time-zero preferences.

Before stating our thesis, we briefly fix terminology. The term "time-zero utility" is

[1] See Bryan, Karlan, and Nelson (2010) and Brocas, Carrillo, and Dewatripont (2004) for surveys.

[2] Of course, Rubinstein does not agree with this particular critique of behavioral economics.

sometimes used in different ways in the literature. For instance, sometimes it literally means the preferences of the time-zero self; other times it means the preferences of an exponential discounter (often called the "long-run self"); and, still other times it refers to a non-discounter. Although we did not distinguish between such concepts in the preceding introductory paragraphs—which are meant to broadly characterize the ongoing welfare debate—from this point forward in our analysis we use the term time-zero utility in the literal sense (preferences of the time-zero self). We feel that this usage of the term is most compatible with the concept of commitment, which implies sticking with one's initial plan rather than following a rule that was never the goal of any of the selves.[3]

The purpose of this paper is to argue that, even though the preferences of the many intertemporal selves are in conflict, behavioral welfare analysis that is based on the preferences of the time-zero self may not be all that arbitrary after all. We make this point with a continuous-time, life-cycle consumption-saving model with hyperbolic discounting. We show with numerical simulations that the time-zero consumption-saving plan (commitment allocation) is viewed by *all selves* as a better outcome than the consumption-saving path that is actually followed. By following the commitment allocation, the first-self attains the highest possible utility that he can attain, and all other selves are strictly better off as well. Thus, in the language of Bernheim and Rangel (2009), the commitment allocation *strictly multiself Pareto dominates* the actual allocation.[4]

The intuition for this powerful result is best understood by distinguishing between the concepts of *dynamic inconsistency* and *regret*. Of course, because of their dynamically-inconsistent preferences, older selves would naturally choose to jump off the original commitment path and instead commit themselves to less aggressive saving paths going forward.

[3]However, this is basically semantics anyway. For example, in the $\beta\delta$ model, the time-zero savings allocation and the savings allocation that comes from setting $\beta = 1$ (the long-run allocation) are so similar that we can refer to either as the "time-zero allocation." These two allocations only become significantly different when β is unrealistically low or when the time grid is very coarse (e.g., 3 periods). For settings with many periods and with typical values of β, the two savings allocations are virtually indistinguishable.

[4]Commitment can provide welfare gains to individuals in other behavioral models as well (temptation preferences, dual self models, etc.). See Bryan, Karlan, and Nelson (2010) for details. Commitment can also improve welfare in some environments with standard preferences (Brocas, Carrillo, and Dewatripont (2004)).

But the relevant question for welfare assessment is, what savings allocation would a given self like to commit all past, present, and future selves to following, assuming that he has full control rights over lifetime resources and can re-write the past. This savings allocation typically lies near or well above the original commitment savings allocation, from the perspective of any age vantage point. This in turn means that individuals tend to regret not having saved *more* than the commitment allocation, and moving from the actual savings allocation to the commitment allocation is a Pareto move. Hence, treating time-zero preferences as the welfare function is not as arbitrary as it may seem but instead has a logical foundation, at least in the context of the life-cycle setting that we consider.[5]

We quantify these Pareto gains in the following way. For each self over the life cycle, we compute the fraction of the *utility gap* between ideal and actual consumption allocations that is closed by following the commitment path. That is, we calculate the utility gain accruing to a given self that comes from commitment to the original (time-zero) plan, relative to the utility gain that would accrue to that same self if he could commit all selves (past and future) to the plan that he thinks is optimal. Of course, the welfare gains cap out at 100% for the time-zero self by definition, so the main quantitative questions relate to how the gains hold up as we consider older selves and how the gains vary across different parameterizations of the model.

We can summarize our main quantitative findings as follows:

1. The welfare gains from commitment to the time-zero allocation are always positive (and often significant) for all the different selves and for any assumption about the degree of backward (retrospective) discounting.

2. For our preferred parameterization in which the individual discounts all delays in the same manner, whether forward or backward, the welfare gains from commitment range

[5]Ever since Strotz (1956), economists have understood that consumers with dynamically inconsistent preferences will demand commitment devices (if they are aware of their self-control problem). We do not discuss how commitment devices could be designed and implemented. That is the subject of many other papers. See Bryan, Karlan, and Nelson (2010) for a survey. Instead, we focus on a more fundamental question of whether or not they are welfare improving in the first place.

between 23% to 100% across the many selves.

3. For the extreme parameterization in which the individual cares just as much about all past consumption realizations as he cares about present consumption, the welfare gains from commitment range between 30% to 100% across the many selves.

4. For the other extreme parameterization in which the individual does not care about the past at all (infinite backward discounting), the welfare gains from commitment range between 6% to 100% across the many selves.

5. For every vantage point (self), the welfare gains from commitment are strictly decreasing in the degree of backward discounting: commitment confers the largest gains under no backward discounting and commitment confers the smallest gains under infinite backward discounting.

These findings are a natural extension of the fundamental results proved by Caplin and Leahy (2004). They show that the consumption-saving path actually followed by dynamically-*consistent* individuals is the most impatient allocation of all the Pareto optima, where each member of the Pareto set is an allocation that is optimal from the perspective of a given vantage point in the life cycle. While Caplin and Leahy focus on dynamically-consistent behavior, the spirit of their result readily extends to our setting with dynamic inconsistency: while individuals will want to save less than their commitment assets if they can jump off the commitment path at a point in time, they will also wish they could constrain all selves to save *more* than the commitment allocation. This distinction forms the basis for the welfare gains that we document. Hence, all the selves agree that moving from the actual savings allocation to the commitment allocation is a move in the right direction.[6]

[6] It is important to emphasize that advocating for the time-zero allocation as a normative goal is not based on a paternalistic argument. Instead, it is choice-based in the sense that Bernheim and Rangel (2009) define choice-based welfare criteria. In fact, Bernheim and Rangel's choice-based logic leads us to choose the commitment allocation over the actual allocation because, even though the latter is actually chosen, the former would be chosen over the latter by all the selves if the former were always available (had younger selves not spent too much and saved too little).

Finally, our paper is a generalization of a point made by Laibson (1996). He argues that the saving rate that the time-zero individual would commit himself to follow in an infinite-horizon setting is Pareto superior to the equilibrium saving rate without commitment (also see Goldman (1979)). Similarly, Laibson, Repetto, and Tobacman (1998) consider the welfare gains of commitment to the plan that a dynamically-consistent (exponential) individual would follow, and they quantify these gains from the perspective of various selves in a finite-horizon life-cycle economy. Likewise, Laibson (1997) computes the welfare gains associated with "partial commitment"—the gains that result from a lack of access to instantaneous credit and a lack of ability to immediately liquidate asset holdings, and Cropper and Laibson (1999) compute the optimal capital subsidies that would perfectly replicate the consumption-saving allocation preferred by the time-zero self. In each of these studies, Laibson and his coauthors focus on the special case of infinite backward discounting (no weight is given to past consumption), whereas we generalize the multiself welfare calculations to allow for the full range of assumptions about backward looking preferences.[7]

By allowing for a wide variety of assumptions about backward discounting, we confront Bernheim and Rangel's (2009) concern that typical multiself Pareto analysis suffers from the "conceptual deficiency" that individuals are assumed to derive no utility at all from past consumption (infinite backward discounting). Bernheim and Rangel argue that there

[7] İmrohoroğlu, İmrohoroğlu, and Joines (2003) perform two related sets of welfare experiments (full commitment and partial commitment). Their first set is very similar to Laibson, Repetto, and Tobacman (1998) in that they also consider the welfare gains that accrue to each self from commitment to the plan that a dynamically-consistent (exponential) individual would follow in a finite-horizon life-cycle economy. However, unlike Laibson, Repetto, and Tobacman who consider the special case of infinite backward discounting, İmrohoroğlu, İmrohoroğlu, and Joines consider the special case of backward *inflation* at the exponential rate, which allows them to abstract from retrospective dynamic inconsistency and focus just on forward looking inconsistency. As mentioned above, our analysis generalizes these types of experiments by allowing for a full range of assumptions about backward preferences, and we focus on commitment to the time-zero allocation rather than to the allocation that comes from exponential discounting.

Their second set of experiments is similar to Laibson (1997) in that both studies consider the gains that accrue to each self from partial commitment, though the papers differ in that partial commitment in Laibson's paper relates to credit market frictions while social security is the partial commitment device in İmrohoroğlu, İmrohoroğlu, and Joines. Also, whereas Laibson considers the special case of infinite backward discounting, İmrohoroğlu, İmrohoroğlu, and Joines consider a range of assumptions about backward discounting. We differ from both studies in that we are focused exclusively on the value of committing to the time-zero consumption-saving allocation, which is distinctly different than partial commitment.

is no empirical basis for such an assumption, nor is there a solid basis for making any other *specific* assumption about backward discounting. Consequently, they advocate, in part, that we "adopt a notion of multiself Pareto efficiency that is robust with respect to a wider range of possibilities [about the nature of backward discounting]" (p.88). Because we show that the commitment allocation strictly multiself Pareto dominates the actual allocation for the full range of assumptions about the magnitude of backward discounting, we conclude that, in the terminology of Bernheim and Rangel, the commitment allocation *strictly and robustly multiself Pareto dominates* the actual allocation. Hence, we do not merely advocate time-zero preferences as the welfare criterion "because we lack critical information [about] backward looking preferences" (p.89), but rather we show that time-zero preferences emerge as a natural welfare criterion under the full range of assumptions about backward looking preferences.

2 Five Optimization Problems

We purposefully focus on a simple, finite-horizon setting with no uncertainty about income, returns, or longevity. In this setting regret is not simply due to bad realizations from some stochastic process but is instead a core feature of the preference structure of the individual. We likewise intentionally limit our attention to partial equilibrium analysis to capture the welfare effects that result from various reallocations of a fixed quantity of resources.

The forward-looking discount function for a delay of length τ is $F(\tau)$ and the backward-looking discount function is $B(\tau)$. Note that we use the term "delay" to mean the absolute value of the length of time between the current moment and some other moment, whether that other moment is in the future or in the past. The usual properties hold: $F(0) = B(0) = 1$, $F(\tau) > 0$, $B(\tau) > 0$, $F'(\tau) < 0$, $B'(\tau) < 0$.

Concerning notation: $c(t)$ is consumption, $k(t)$ is savings, r is the interest rate on savings, $y(t)$ is disposable income, and $\bar{T}$ is the lifespan. Period utility is isoelastic, $u(c) = c^{1-\sigma}/(1-$

σ).

2.1 P1. Time-Zero Self (Commitment Allocation)

The optimal consumption-saving allocation from the perspective of the time-zero self solves the following control problem:

$$\max : \int_0^{\bar{T}} F(t) \frac{c(t)^{1-\sigma}}{1-\sigma} dt,$$

subject to

$$\frac{dk(t)}{dt} = rk(t) + y(t) - c(t),$$

$$k(0) = k(\bar{T}) = 0.$$

Using the Maximum Principle it is straightforward to show that the solution to this problem for $t \in [0, \bar{T}]$ is

$$c_0^*(t) = \frac{\int_0^{\bar{T}} y(t) e^{-rt} dt}{\int_0^{\bar{T}} F(t)^{1/\sigma} e^{(1-\sigma)rt/\sigma} dt} F(t)^{1/\sigma} e^{rt/\sigma}, \quad k_0^*(t) = \int_0^t [y(s) - c_0^*(s)] e^{r(t-s)} ds.$$

2.2 P2. Dynamic Inconsistency (Strotz (1956))

Suppose the individual is now standing at vantage point v and has followed the initial plan from time zero until now. Taking his existing commitment assets as given, the consumption-saving allocation that he now perceives to be optimal solves the following problem:

$$\max : \int_v^{\bar{T}} F(t-v) \frac{c(t)^{1-\sigma}}{1-\sigma} dt,$$

subject to

$$\frac{dk(t)}{dt} = rk(t) + y(t) - c(t),$$

$$k(v) = k_0^*(v) = \int_0^v [y(t) - c_0^*(t)] e^{r(v-t)} dt, \; k(\bar{T}) = 0.$$

Again, with the Maximum Principle it is straightforward to show that for $t \in [v, \bar{T}]$, the solution is

$$c_v^*(t) = \frac{k_0^*(v) + \int_v^{\bar{T}} y(t) e^{-r(t-v)} dt}{\int_v^{\bar{T}} F(t-v)^{1/\sigma} e^{(1-\sigma)rt/\sigma + rv} dt} F(t-v)^{1/\sigma} e^{rt/\sigma},\ k_v^*(t) = k_0^*(v) e^{r(t-v)} + \int_v^t [y(s) - c_v^*(s)] e^{r(t-s)} ds.$$

Note that

$$c_v^*(v) e^{-rv/\sigma} = \frac{k_0^*(v) + \int_v^{\bar{T}} y(t) e^{-r(t-v)} dt}{\int_v^{\bar{T}} F(t-v)^{1/\sigma} e^{(1-\sigma)rt/\sigma + rv} dt},$$

so

$$c_v^*(t) = c_v^*(v) F(t-v)^{1/\sigma} e^{r(t-v)/\sigma}.$$

The Euler equation is

$$\frac{d}{dt} \ln c_v^*(t) = \left[r + \frac{F'(t-v)}{F(t-v)} \right] \frac{1}{\sigma},\ \text{for } t \in [v, \bar{T}].$$

Likewise, the Euler equation from P1 is

$$\frac{d}{dt} \ln c_0^*(t) = \left[r + \frac{F'(t)}{F(t)} \right] \frac{1}{\sigma},\ \text{for } t \in [v, \bar{T}].$$

If the discount function has the property

$$\frac{d}{d\tau} \left(-\frac{F'(\tau)}{F(\tau)} \right) < 0,$$

then

$$\frac{F'(t-v)}{F(t-v)} < \frac{F'(t)}{F(t)},\ \text{for all } t > v > 0,$$

and

$$\frac{d}{dt} \ln c_v^*(t) < \frac{d}{dt} \ln c_0^*(t),\ \text{for } t \in (v, \bar{T}] \Longrightarrow c_v^*(v) > c_0^*(v).$$

Now that the individual is standing at the new vantage point v, he prefers a new consumption profile with a slope that is less than the original plan, and hence he prefers to switch to a new plan that involves less saving for the future. The particular hyperbolic function that we use in our numerical examples later in the paper has this property.

If we were to stop the analysis here, we might conclude that committing people to the first plan amounts to asking future selves to save too much, because future selves would prefer a new saving plan that is strictly less aggressive than the initial plan. But such a conclusion would be a mistake because it fails to distinguish between the concepts of *dynamically inconsistent behavior* and *regret*. The former relates to the desire to switch plans, given their present state of affairs. But the latter is about wishing one could follow an entirely different plan from the start. In the next subsection we will show that even though the individual wants a less aggressive saving plan than the commitment plan going forward (conditional on existing assets), he would actually prefer a saving plan that is more aggressive than the commitment plan for the entire past, present, and future if he could re-write history.

2.3 P3. Regret (Caplin and Leahy (2004))

Suppose the individual is standing at v, but unlike the previous problem in which he takes existing assets as given, he now imagines the entire life-cycle consumption-saving program that he views as optimal. That is, he ignores the reality of how much (or little) he has already accumulated in his asset account by age v and instead imagines what might have been: he imagines the ideal plan for his past, present, and future consumption and savings. This program is the solution to a two-stage optimal control problem:

$$\max : \int_0^v B(v-t)\frac{c(t)^{1-\sigma}}{1-\sigma}dt + \int_v^{\bar{T}} F(t-v)\frac{c(t)^{1-\sigma}}{1-\sigma}dt,$$

subject to

$$\frac{dk(t)}{dt} = rk(t) + y(t) - c(t),$$

$$k(0) = k(\bar{T}) = 0.$$

Following the Two-Stage Maximum Principle (Tomiyama (1985)), we form a pair of Hamiltonians

$$\mathcal{H}_1 = B(v-t)\frac{c(t)^{1-\sigma}}{1-\sigma} + \lambda_1(t)[rk(t) + y(t) - c(t)], \text{ for } t \in [0, v],$$

$$\mathcal{H}_2 = F(t-v)\frac{c(t)^{1-\sigma}}{1-\sigma} + \lambda_2(t)[rk(t) + y(t) - c(t)], \text{ for } t \in [v, \bar{T}].$$

The first-order conditions include

$$\frac{\partial \mathcal{H}_1}{\partial c(t)} = B(v-t)c(t)^{-\sigma} - \lambda_1(t) = 0, \text{ for } t \in [0, v],$$

$$\frac{\partial \mathcal{H}_2}{\partial c(t)} = F(t-v)c(t)^{-\sigma} - \lambda_2(t) = 0, \text{ for } t \in [v, \bar{T}],$$

$$\frac{d\lambda_1(t)}{dt} = -\frac{\partial \mathcal{H}_1}{\partial k(t)} = -r\lambda_1(t), \text{ for } t \in [0, v],$$

$$\frac{d\lambda_2(t)}{dt} = -\frac{\partial \mathcal{H}_2}{\partial k(t)} = -r\lambda_2(t), \text{ for } t \in [v, \bar{T}],$$

$$\lambda_1(v) = \lambda_2(v).$$

Solving the costate equations gives

$$\lambda_1(t) = a_1 e^{-rt}, \text{ for } t \in [0, v],$$

$$\lambda_2(t) = a_2 e^{-rt}, \text{ for } t \in [v, \bar{T}],$$

for constants a_1 and a_2. The matching condition implies $a_1 = a_2$ and hence we can drop the subscripts

$$\lambda(t) = ae^{-rt}, \text{ for } t \in [0, \bar{T}].$$

Rewrite the Maximum Conditions

$$c(t) = [ae^{-rt}]^{-1/\sigma} B(v-t)^{1/\sigma}, \text{ for } t \in [0, v],$$

$$c(t) = [ae^{-rt}]^{-1/\sigma} F(t-v)^{1/\sigma}, \text{ for } t \in [v, \bar{T}].$$

From the state equation we have

$$k(t) = \begin{cases} \int_0^t \left\{ y(s) - [ae^{-rs}]^{-1/\sigma} B(v-s)^{1/\sigma} \right\} e^{r(t-s)} ds, \text{ for } t \in [0, v], \\ k(v) e^{r(t-v)} + \int_v^t \left\{ y(s) - [ae^{-rs}]^{-1/\sigma} F(s-v)^{1/\sigma} \right\} e^{r(t-s)} ds, \text{ for } t \in [v, \bar{T}]. \end{cases}$$

Evaluate $k(t)$ at $t = \bar{T}$ and use $k(\bar{T}) = 0$

$$\begin{aligned} 0 &= \int_0^v \left\{ y(s) - [ae^{-rs}]^{-1/\sigma} B(v-s)^{1/\sigma} \right\} e^{r(v-s)} ds \times e^{r(\bar{T}-v)} \\ &\quad + \int_v^{\bar{T}} \left\{ y(s) - [ae^{-rs}]^{-1/\sigma} F(s-v)^{1/\sigma} \right\} e^{r(\bar{T}-s)} ds \\ &= \int_0^v \left\{ y(s) - [ae^{-rs}]^{-1/\sigma} B(v-s)^{1/\sigma} \right\} e^{-rs} ds \\ &\quad + \int_v^{\bar{T}} \left\{ y(s) - [ae^{-rs}]^{-1/\sigma} F(s-v)^{1/\sigma} \right\} e^{-rs} ds \end{aligned}$$

and then solve for a

$$a^{-1/\sigma} = \frac{\int_0^{\bar{T}} y(s) e^{-rs} ds}{\int_0^v e^{rs/\sigma - rs} B(v-s)^{1/\sigma} ds + \int_v^{\bar{T}} e^{rs/\sigma - rs} F(s-v)^{1/\sigma} ds}.$$

Hence, looking backward over $t \in [0, v]$ the optimal consumption path is

$$c_v^{**}(t) = \frac{\int_0^{\bar{T}} y(s) e^{-rs} ds}{\int_0^v e^{rs/\sigma - rs} B(v-s)^{1/\sigma} ds + \int_v^{\bar{T}} e^{rs/\sigma - rs} F(s-v)^{1/\sigma} ds} e^{rt/\sigma} B(v-t)^{1/\sigma},$$

and looking ahead over $t \in [v, \bar{T}]$ the optimal consumption path is

$$c_v^{**}(t) = \frac{\int_0^{\bar{T}} y(s) e^{-rs} ds}{\int_0^v e^{rs/\sigma - rs} B(v-s)^{1/\sigma} ds + \int_v^{\bar{T}} e^{rs/\sigma - rs} F(s-v)^{1/\sigma} ds} e^{rt/\sigma} F(t-v)^{1/\sigma}.$$

Claim 1. *From any vantage point $v > 0$, the individual will regret not having saved* <u>*more*</u>

than his commitment assets, whether he does or does not discount the past.

Proof. We begin by assuming he discounts the past. Without loss of generality, we assume that he discounts the past in the same manner that he discounts the future to keep the math clean and simple ($B(\tau) = F(\tau)$), though any backward discount function will work. To make the point, it is enough to compare the consumption choice at $t = 0$,

$$c_0^*(0) = \frac{\int_0^{\bar{T}} y(t)e^{-rt}dt}{\int_0^{\bar{T}} F(t)^{1/\sigma}e^{(1-\sigma)rt/\sigma}dt} = \frac{\int_0^{\bar{T}} y(t)e^{-rt}dt}{\int_0^{v} F(t)^{1/\sigma}e^{(1-\sigma)rt/\sigma}dt + \int_v^{\bar{T}} F(t)^{1/\sigma}e^{(1-\sigma)rt/\sigma}dt},$$

to the consumption choice that the individual standing at age v wishes that he had taken at $t = 0$

$$\begin{aligned} c_v^{**}(0) &= \frac{\int_0^{\bar{T}} y(s)e^{-rs}ds}{\int_0^{v} e^{rs/\sigma-rs}B(v-s)^{1/\sigma}ds + \int_v^{\bar{T}} e^{rs/\sigma-rs}F(s-v)^{1/\sigma}ds}B(v)^{1/\sigma} \\ &= \frac{\int_0^{\bar{T}} y(t)e^{-rt}dt}{\int_0^{v} e^{rt/\sigma-rt}[B(v-t)/B(v)]^{1/\sigma}dt + \int_v^{\bar{T}} e^{rt/\sigma-rt}[F(t-v)/B(v)]^{1/\sigma}dt} \\ &= \frac{\int_0^{\bar{T}} y(t)e^{-rt}dt}{\int_0^{v} e^{rt/\sigma-rt}[F(v-t)/F(v)]^{1/\sigma}dt + \int_v^{\bar{T}} e^{rt/\sigma-rt}[F(t-v)/F(v)]^{1/\sigma}dt}. \end{aligned}$$

Note

$$F'(\tau) < 0 \Longrightarrow F(t-v) > F(t) \Longrightarrow \frac{F(t-v)}{F(v)} > F(t), \text{ for all } t > v > 0,$$

and hence

$$\int_v^{\bar{T}} e^{rt/\sigma-rt}\left[\frac{F(t-v)}{F(v)}\right]^{1/\sigma} dt > \int_v^{\bar{T}} F(t)^{1/\sigma}e^{(1-\sigma)rt/\sigma}dt.$$

Also note that

$$F'(\tau) < 0 \Longrightarrow \frac{F(v-t)}{F(v)} > 1 \Longrightarrow \frac{F(v-t)}{F(v)} > F(t), \text{ for any } v > t > 0,$$

and hence

$$\int_0^{v} e^{rt/\sigma-rt}\left[\frac{F(v-t)}{F(v)}\right]^{1/\sigma} dt > \int_0^{v} F(t)^{1/\sigma}e^{(1-\sigma)rt/\sigma}dt.$$

Thus,

$$c_v^{**}(0) < c_0^*(0).$$

Looking back, the individual wishes he had consumed less and saved more.

Next consider the case where he does not discount the past at all, $B(\tau) = 1 > F(\tau)$ for all $\tau > 0$. Simple examination of the equations above allows us to conclude that, again,

$$c_v^{**}(0) < c_0^*(0),$$

which implies that the individual wishes he had consumed less and saved more than the commitment plan. Notice that these results are invariant to the particular vantage point v.

■

2.4 P4. Actual Allocation under Naiveté

Suppose the individual is standing at vantage point v but, unlike Problem 2, he has *not* followed the initial plan from time zero until now. He has continually failed to follow past plans and has instead reoptimized in a naive way, always thinking future selves will follow his current plans. At age v, he makes a consumption-saving plan according to the following problem:

$$\max : \int_v^{\bar{T}} F(t-v) \frac{c(t)^{1-\sigma}}{1-\sigma} dt,$$

subject to

$$\frac{dk(t)}{dt} = rk(t) + y(t) - c(t),$$

$$k(v) = k_a(v) = \int_0^v [y(t) - c_a(t)] e^{r(v-t)} dt, \; k(\bar{T}) = 0,$$

where k_a and c_a stand for actual savings and consumption. Again, with the Maximum Principle it is straightforward to show that for $t \in [v, \bar{T}]$, the solution is

$$c(t) = \frac{k_a(v) + \int_v^{\bar{T}} y(t)e^{-r(t-v)}dt}{\int_v^{\bar{T}} F(t-v)^{1/\sigma} e^{(1-\sigma)rt/\sigma + rv}dt} F(t-v)^{1/\sigma} e^{rt/\sigma}.$$

However, the individual only follows this plan at the moment it is made, and not beyond. So actual consumption at vantage point v is found by setting $t = v$,

$$c_a(v) = \frac{k_a(v) + \int_v^{\bar{T}} y(t)e^{-r(t-v)}dt}{\int_v^{\bar{T}} F(t-v)^{1/\sigma} e^{(1-\sigma)rt/\sigma + rv}dt} e^{rv/\sigma}.$$

And because we are working in continuous time and the individual can reoptimize without constraints, all points in time are vantage points and we can replace v with t to get actual consumption at age t

$$c_a(t) = \frac{k_a(t) + \int_t^{\bar{T}} y(s)e^{-r(s-t)}ds}{\int_t^{\bar{T}} F(s-t)^{1/\sigma} e^{(1-\sigma)rs/\sigma + rt}ds} e^{rt/\sigma},$$

where $k_a(t)$ obeys

$$\frac{dk_a(t)}{dt} = rk_a(t) + y(t) - c_a(t),$$

$$k_a(0) = k_a(\bar{T}) = 0.$$

2.5 P5. Social Welfare (Caplin and Leahy (2004))

The solution consumption-saving allocation from P3 is optimal from the perspective of age v, but not from any other age.[8] Hence there is disagreement among the selves concerning how allocations should be valued. Each solution represents a Pareto optimum, or a cross-section of the Pareto surface.

Utility from the perspective of the individual standing at age v,

[8] This is true even if F and B are exponential. In fact, this is the point of Caplin and Leahy's paper.

$$U(v) \equiv \int_0^v B(v-t)u(c(t))dt + \int_v^{\bar{T}} F(t-v)u(c(t))dt.$$

"Social welfare" embodies equal treatment of the preferences of all the different selves (i.e., utilitarian),

$$\mathcal{SW} \equiv \int_0^{\bar{T}} U(v)dv.$$

If so, then the socially optimal consumption profile solves the following control problem:

$$\max : \int_0^{\bar{T}} \left(\int_0^v B(v-t)u(c(t))dt + \int_v^{\bar{T}} F(t-v)u(c(t))dt \right) dv,$$

subject to

$$\frac{dk(t)}{dt} = rk(t) + y(t) - c(t),$$

$$k(0) = k(\bar{T}) = 0.$$

This has the appearance of an intractable control problem, but we can make progress with a change in the order of integration of the iterated integrals in the objective functional,

$$\int_0^{\bar{T}} \left(\int_0^v B(v-t)u(c(t))dt \right) dv = \int_0^{\bar{T}} \left(\int_t^{\bar{T}} B(v-t)u(c(t))dv \right) dt,$$

$$\int_0^{\bar{T}} \left(\int_v^{\bar{T}} F(t-v)u(c(t))dt \right) dv = \int_0^{\bar{T}} \left(\int_0^t F(t-v)u(c(t))dv \right) dt,$$

so

$$
\begin{aligned}
& \int_0^{\bar{T}} \left(\int_0^v B(v-t)u(c(t))dt + \int_v^{\bar{T}} F(t-v)u(c(t))dt \right) dv \\
= & \int_0^{\bar{T}} \left(\int_t^{\bar{T}} B(v-t)u(c(t))dv \right) dt + \int_0^{\bar{T}} \left(\int_0^t F(t-v)u(c(t))dv \right) dt \\
= & \int_0^{\bar{T}} \left(\int_t^{\bar{T}} B(v-t)dv + \int_0^t F(t-v)dv \right) u(c(t))dt \\
= & \int_0^{\bar{T}} D(t)u(c(t))dt,
\end{aligned}
$$

where

$$D(t) \equiv \int_t^{\bar{T}} B(v-t)dv + \int_0^t F(t-v)dv.$$

Thus, the social optimization problem can be re-stated compactly

$$\max : \int_0^{\bar{T}} D(t)u(c(t))dt,$$

subject to

$$\frac{dk(t)}{dt} = rk(t) + y(t) - c(t),$$

$$k(0) = k(\bar{T}) = 0.$$

The solution to this optimal control problem for isoelastic utility is

$$c^*_{\mathcal{SW}}(t) = \frac{\int_0^{\bar{T}} y(t)e^{-rt}dt}{\int_0^{\bar{T}} e^{rt/\sigma}D(t)^{1/\sigma}e^{-rt}dt} e^{rt/\sigma}D(t)^{1/\sigma},$$

$$k^*_{\mathcal{SW}}(t) = \int_0^t [y(s) - c^*_{\mathcal{SW}}(s)]e^{r(t-s)}ds.$$

3 Numerical Examples

3.1 Parameter Values

Numerical examples allow us to characterize the quantitative magnitude of the welfare gains from commitment. We set $\bar{T} = 55$ to reflect an economic lifespan from ages 25 to 80. Assuming retirement occurs after 40 years of work, we set $y(t) = 1$ before 40 and $y(t) = 0.4$ afterwards to reflect social security benefits and other sources of income such as part-time work. We set $r = 1\%$ to align with typical risk-free returns on US treasuries and we parameterize the utility function to be logarithmic ($\sigma = 1$). Finally, we use standard hyperbolic discount functions $F(\tau) = (1 + \alpha_F \tau)^{-1}$ and $B(\tau) = (1 + \alpha_B \tau)^{-1}$. We assume a modest amount of forward discounting, $\alpha_F = 3.5\%$.

All that remains is to select the backward discount parameter α_B. We break into three parameterizations that vary in the value that we assign to α_B.

Parameterization 1 ($\alpha_B = \alpha_F$). Here, the present is salient and the individual discounts all delays, whether forward or backward, in the same manner. The memory of a great vacation last year provides the same utility as the anticipation of a great vacation next year. Because this seems like a logical place to start, we refer to this as our preferred parameterization. In this case $D(t)$ is strictly concave (quadratic) with a peak at $t = \bar{T}/2$:

$$D(t) = \frac{\ln[1 + \alpha(\bar{T} - t)]}{\alpha} + \frac{\ln[1 + \alpha t]}{\alpha}.$$

$$D'(t) = \frac{-1}{1 + \alpha(\bar{T} - t)} + \frac{1}{1 + \alpha t}.$$

$$D'(t) = 0 \Longleftrightarrow t = \frac{\bar{T}}{2}.$$

$$D''(t) = \frac{-\alpha}{[1 + \alpha(\bar{T} - t)]^2} - \frac{\alpha}{[1 + \alpha t]^2} < 0.$$

The quadratic shape of this function is because the midpoint is, on average, the closest to all the various vantage points and hence is discounted the least in an aggregate sense. On the other hand, the boundaries of the life cycle are, on average, the furthest from all the vantage points and so they get the least weight in the social optimization problem.

Parameterization 2 ($\alpha_B = 0$). The individual cares just as much about the past as he cares about the present. Here $D(t)$ is strictly concave and strictly decreasing with a peak at $t = 0$:

$$\begin{aligned}
\lim_{\alpha_B \longrightarrow 0} D(t) &= \lim_{\alpha_B \longrightarrow 0} \frac{\ln[1+\alpha_B(\bar{T}-t)]}{\alpha_B} + \lim_{\alpha_B \longrightarrow 0} \frac{\ln[1+\alpha_F t]}{\alpha_F} \\
&= \lim_{\alpha_B \longrightarrow 0} \frac{\frac{d}{d\alpha_B}\ln[1+\alpha_B(\bar{T}-t)]}{\frac{d}{d\alpha_B}\alpha_B} + \frac{\ln[1+\alpha_F t]}{\alpha_F} \\
&= \lim_{\alpha_B \longrightarrow 0} \frac{\bar{T}-t}{1+\alpha_B(\bar{T}-t)} + \frac{\ln[1+\alpha_F t]}{\alpha_F} \\
&= \bar{T} - t + \frac{\ln[1+\alpha_F t]}{\alpha_F}.
\end{aligned}$$

$$\frac{d}{dt}\left(\lim_{\alpha_B \longrightarrow 0} D(t)\right) = -1 + F(t) < 0.$$

$$\frac{d^2}{dt^2}\left(\lim_{\alpha_B \longrightarrow 0} D(t)\right) = F'(t) < 0.$$

Individuals have perfect memory recall in the sense that a fun vacation a year ago is just as valuable as a fun vacation today. The social optimum respects this, which gives the discount function its concave, decreasing shape.

Parameterization 3 ($\alpha_B = \infty$). The individual derives no utility at all from past consumption and only cares about the present and future. Here $D(t)$ is strictly concave and strictly increasing, with a peak at $t = \bar{T}$:

$$
\begin{aligned}
\lim_{\alpha_B \longrightarrow \infty} D(t) &= \lim_{\alpha_B \longrightarrow \infty} \frac{\ln[1+\alpha_B(\bar{T}-t)]}{\alpha_B} + \lim_{\alpha_B \longrightarrow \infty} \frac{\ln[1+\alpha_F t]}{\alpha_F} \\
&= \lim_{\alpha_B \longrightarrow \infty} \frac{\frac{d}{d\alpha_B} \ln[1+\alpha_B(\bar{T}-t)]}{\frac{d}{d\alpha_B}\alpha_B} + \frac{\ln[1+\alpha_F t]}{\alpha_F} \\
&= \lim_{\alpha_B \longrightarrow \infty} \frac{\bar{T}-t}{1+\alpha_B(\bar{T}-t)} + \frac{\ln[1+\alpha_F t]}{\alpha_F} \\
&= \frac{\ln[1+\alpha_F t]}{\alpha_F}.
\end{aligned}
$$

$$
\frac{d}{dt}\left(\lim_{\alpha_B \longrightarrow \infty} D(t)\right) = F(t) > 0.
$$

$$
\frac{d^2}{dt^2}\left(\lim_{\alpha_B \longrightarrow \infty} D(t)\right) = F'(t) < 0.
$$

Here individuals do not care at all about the past, but all of the selves care about utility when old (to varying degrees). The social optimum respects this "consensus" opinion about the importance of old-age consumption and places enormous weight on utility when old.

Figure 1 shows individual discount functions for the three parameterizations of α_B. We set the vantage point $v = 15$ for illustrative purposes, which corresponds to age 40 given that model time starts at age 25. Figure 2 reports the social discount function $D(t)$ for the same parameterizations of α_B.

Figure 3 reports savings allocations over the life cycle for P1-P5, for the case of $\alpha_B = \alpha_F$ (Parameterization 1) and $v = 15$ for illustrative purposes. The path the individual at time zero would like to follow is $k_0^*(t)$, but the individual's time-inconsistent preferences cause him to follow $k_a(t)$ instead. The path $k_v^*(t)$ is an example of how the individual at vantage point v would like to deviate from the commitment path, taking commitment assets at $t = v$ as given. The path $k_v^{**}(t)$ is the ideal path from the perspective of self v, given that he has full control rights over his entire lifetime resources and can therefore dictate past, present, and future decisions. Finally, $k_{\mathcal{SW}}^*(t)$ is the "socially" optimal (utilitarian) savings path that respects the preferences of all the selves.

Note the distinction between (i) what the age-v individual would like to commit all future selves to doing $k_v^*(t)$, given assets $k_v^*(v) = k_0^*(v)$, and (ii) what the age-v individual would like to commit all past, present, and future selves to saving $k_v^{**}(t)$, taking lifetime resources as given. The former is relevant for behavior; the latter is relevant for welfare. Notice that the commitment asset path $k_0^*(t)$ is a move in the right direction: it lies between the actual savings allocation $k_a(t)$ and the ideal allocation $k_v^{**}(t)$.

3.2 Individual Welfare Simulations

We report welfare changes at the individual level rather than the social level to make our point that the commitment allocation is a Pareto improvement over the actual allocation. For an individual standing at age v, we compute the fraction of the *utility gap* between ideal (P3) and actual (P4) consumption allocations that is closed by following the commitment path (P1):

$$\Delta(v) \equiv \frac{U(v)_{c(t)=c_0^*(t)} - U(v)_{c(t)=c_a(t)}}{U(v)_{c(t)=c_v^{**}(t)} - U(v)_{c(t)=c_a(t)}}.$$

Of course, the welfare gains are 100% for the time-zero self by definition. So the main quantitative questions relate to how the gains hold up as we consider older selves and how the gains vary across parameterizations of the backward discount parameter α_B.

Figure 4 plots $\Delta(v)$ for all v, and for each of the three numerical parameterizations mentioned above. From this figure we can summarize our main quantitative findings:

1. The welfare gains are always positive (and often significant) for all the different selves and for all three parameterizations of the backward discount parameter α_B.

2. For our preferred parameterization in which the individual discounts all delays, whether forward or backward, in the same manner ($\alpha_B = \alpha_F$), the welfare gains from commitment range between 23% to 100%.

3. For the extreme parameterization in which the individual cares just as much about all past consumption realizations as he care about present consumption ($\alpha_B = 0$), the welfare gains from commitment range between 30% to 100%.

4. For the other extreme parameterization in which the individual does not care about the past at all ($\alpha_B = \infty$), the welfare gains from commitment range between 6% to 100%.

5. For every vantage point v, the Pareto gains from commitment are strictly decreasing in the degree of backward discounting: commitment confers the largest gains under no backward discounting and commitment confers the smallest gains under infinite backward discounting.

4 Discussion

Welfare analysis in behavioral economics is controversial. Bryan, Karlan, and Nelson (2010, p.694) believe that "settling on a particular approach and providing empirical support or clear philosophical arguments for [that approach] are the hard questions....[that deserve] more thought and research." We have tried to make some progress toward a stronger philosophical argument for the standard approach of treating time-zero preferences as the welfare criterion. We have done this by extending the arguments of Caplin and Leahy (2004) and have paid particular attention to the consideration of a broad range of assumptions about backward looking preferences, thereby overcoming Bernheim and Rangel's (2009) concerns about the potential fragility of previous claims that commitment is Pareto efficient.

A deeper issue that is still very much unresolved relates to the difficulty in using choice data to distinguish between competing choice theories. A number of theorists have proposed a variety of alternatives to hyperbolic discounting, and these alternatives often make predictions about choices that are similar to hyperbolic discounting. Indeed, even standard dynamically-consistent preferences offer similar predictions along some dimensions. Hence, while we have shown that commitment to the time-zero plan represents a Pareto improvement over the actual (equilibrium) consumption-saving allocation, we still don't know that hyperbolic discounting is the correct preference structure in the first place.

References

Bernheim, Douglas B. and Antonio Rangel (2009). Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Analysis. *Quarterly Journal of Economics* 124(1), 51-104.

Brocas, Isabelle and Juan D. Carrillo (2004). Commitment Devices under Self-control Problems: An Overview. In *The Psychology of Economic Decisions: Volume II: Reasons and Choices*, edited by Isabelle Brocas and Juan D. Carrillo. Oxford: Oxford University Press, 49-65.

Bryan, Gharad, Dean Karlan, and Scott Nelson (2010). Commitment Devices. *Annual Review of Economics* 2, 671-698.

Caplin, Andrew and John Leahy (2004). The Social Discount Rate. *Journal of Political Economy* 112(6), 1257-1268.

Cropper, Maureen and David Laibson (1999). The Implications of Hyperbolic Discounting for Project Evaluation. In *Discounting and Intergenerational Equity,* edited by Paul R. Portney and John P. Weyant. Resources for the Future, 1999.

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

Gul, Faruk and Wolfgang Pesendorfer (2004). Self-Control, Revealed Preference, and Consumption Choice. *Review of Economic Dynamics* 7, 243-264.

Gul, Faruk and Wolfgang Pesendorfer (2008). The Case for Mindless Economics. In *Foundations of Positive and Normative Economics,* edited by Andrew Caplin and Andrew Schotter. Oxford University Press, 2008.

İmrohoroğlu, Ayşe, Selahattin İmrohoroğlu, and Douglas Joines (2003). Time-Inconsistent Preferences and Social Security. *Quarterly Journal of Economics* 118(2), 745–784.

Laibson, David I. (1996). Hyperbolic Discount Functions, Undersaving, and Savings Policy. NBER Working Paper 5635.

Laibson, David I. (1997). Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics* 112(2), 443-478.

Laibson, David I., Andrea Repetto, and Jeremy Tobacman (1998). Self-Control and Saving for Retirement. Brookings Papers on Economic Activity 1, 91-172.

O'Donoghue, Ted and Matthew Rabin (1999). Doing It Now or Later. *American Economic Review* 89(1), 103-124.

O'Donoghue, Ted and Matthew Rabin (2000). The Economics of Immediate Gratification. *Journal of Behavioral Decision Making* 13, 233-250.

O'Donoghue, Ted and Matthew Rabin (2001). Choice and Procrastination. *Quarterly Journal of Economics* 11(1), 121-160.

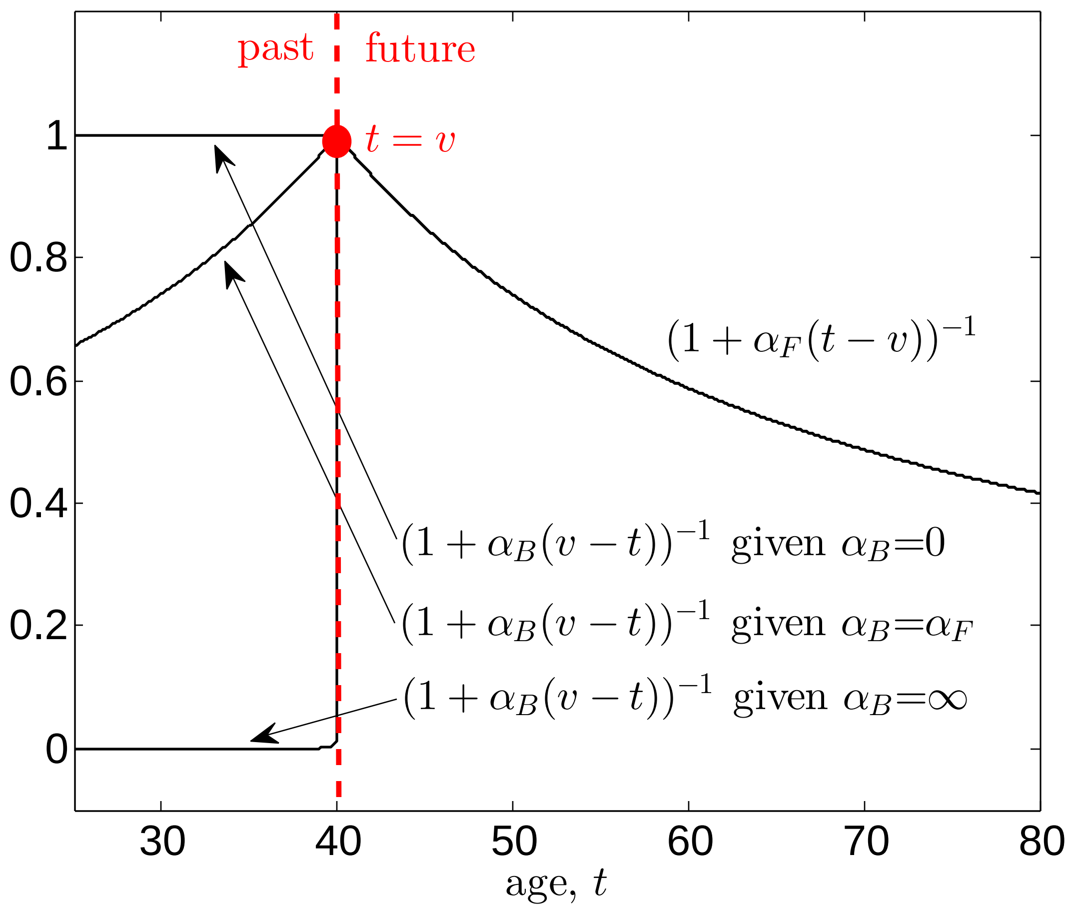
O'Donoghue, Ted and Matthew Rabin (2003). Self-Awareness and Self-Control. In *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, edited by George Loewenstein, Daniel Read, and Roy F. Baumeister. Washington DC: Russell Sage Foundation Publications, 217-243.

Rubinstein, Ariel (2006). Discussion of "Behavioral Economics". In *Advances in Economics and Econometrics: Theory and Application*, Volume 2, edited by Richard Blundell, Whitney K. Newey, and Torsten Persson. New York: Cambridge University Press, 246-254.

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

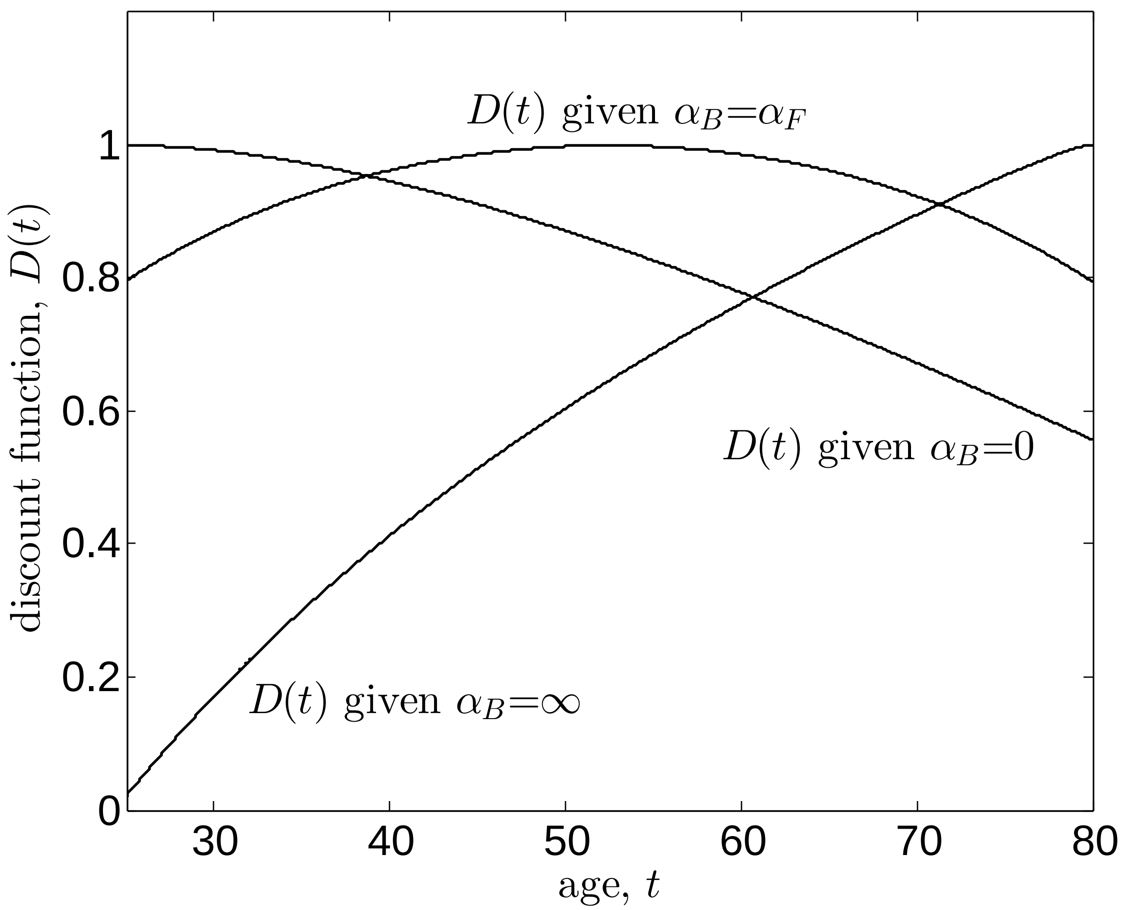
Tomiyama, Ken (1985). Two-Stage Optimal Control Problems and Optimality Conditions. *Journal of Economic Dynamics and Control* 9(3), 317-337.

Figure 1. Discount Functions from the Vantage Point of Age 40



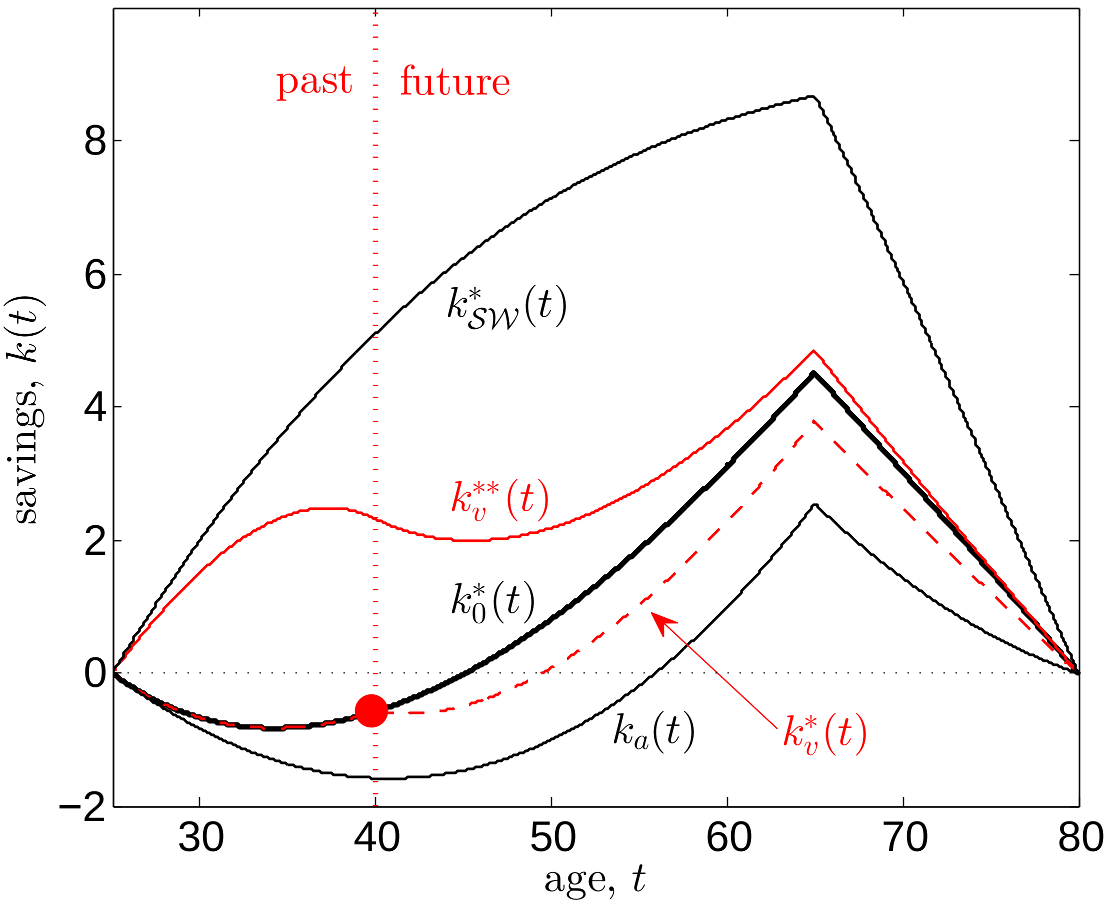
The forward discount parameter, α_F, is set to 0.035.

Figure 2. Social Discount Functions



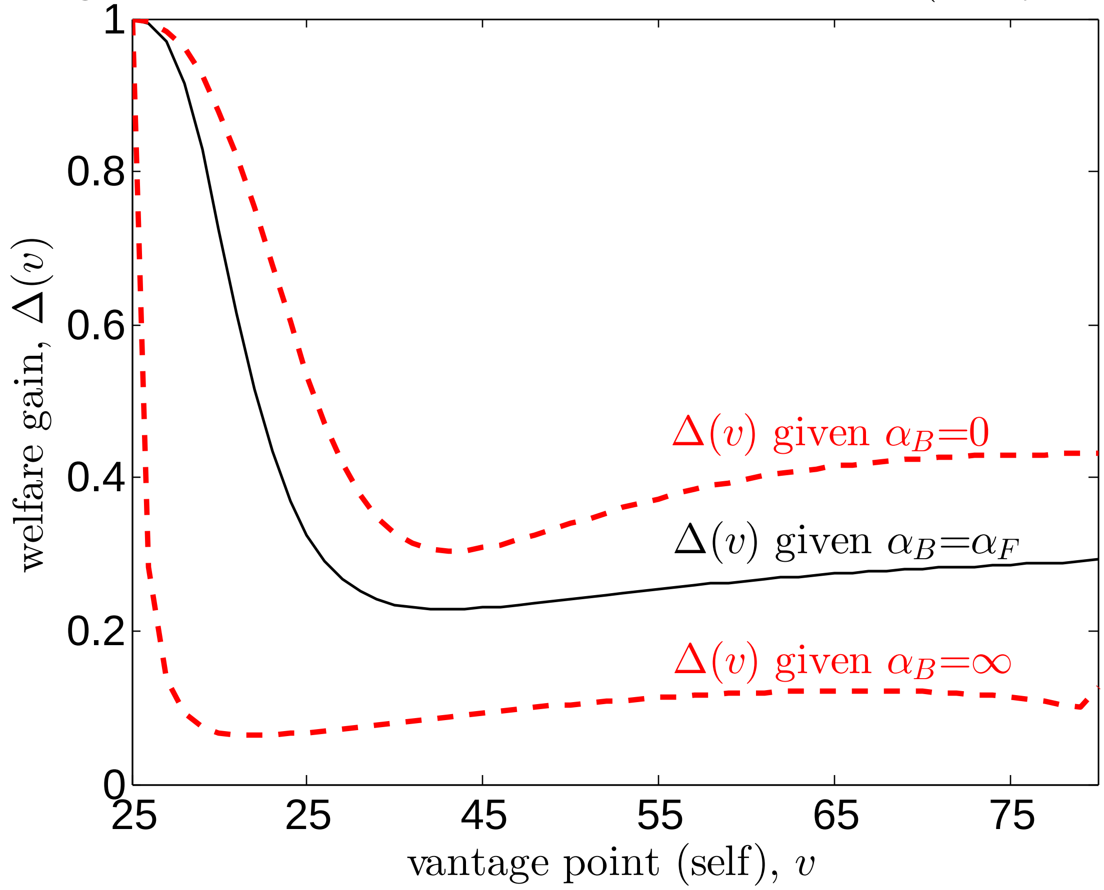
The forward discount parameter, α_F, is set to 0.035.
Each function is normalized to ensure a peak at unity.

Figure 3. Savings Allocations over the Life Cycle (5 Problems)



All allocations correspond to Example 1: $\alpha_B = \alpha_F$.

Figure 4. The Pareto Gains from Commitment (Many Selves)



The forward discount parameter, α_F, is set to 0.035.