



ELSEVIER

Journal of Public Economics 87 (2003) 1839–1872

JOURNAL OF
PUBLIC
ECONOMICS

www.elsevier.com/locate/econbase

Quasi-hyperbolic discounting and retirement

Peter Diamond^a, Botond Köszegi^{b,*}

^a*Department of Economics, MIT, Cambridge, MA 02139, USA*

^b*Department of Economics, University of California, Berkeley, CA 94720 USA*

Received 15 December 2000; received in revised form 29 January 2002; accepted 31 January 2002

Abstract

Some people have self-control problems regularly. This paper adds endogenous retirement to Laibson's quasi-hyperbolic discounting savings model [Quarterly Journal of Economics 112 (1997) 443–477]. Earlier selves think that the deciding self tends to retire too early and may save less to induce later retirement. Still earlier selves may think the pre-retirement self does this too much, saving more to induce early retirement. The consumption pattern may be different from that with exponential discounting. Other observational non-equivalence includes the impact of changing mandatory retirement rules or work incentives on savings and a possibly negative marginal propensity to consume out of increased future earnings. Naive agents are briefly considered.

© 2002 Elsevier B.V. All rights reserved.

Keywords: Laibson's model; Retirement; Quasi-hyperbolic discounting

1. Introduction

If you are one of the vast majority of people who think they are saving too little of their income.¹ The natural conclusion is that you have self-control problems. If, in addition, you argued to yourself that saving more today would only lead to spending more tomorrow, and thus there is no point in saving for retirement, at

*Corresponding author.

E-mail address: botond@econ.berkeley.edu (B. Köszegi).

¹Bernheim (1994) reports that people 'admits to' saving much less for retirement than they should. We don't know, though, how prevalent this is among academics.

least there is a small consolation: you are a *sophisticated* decision-maker with self-control problems. And self-control problems can extend beyond savings decisions. A thirty-something Italian, one of us met in Prague, had decided that it wasn't worth looking for a job anymore, because even if he got himself to do it and found one, he would quit shortly thereafter, anyway.

It is exactly these kinds of agents our paper is concerned with: people who have self-control problems but realize this and behave according to it. A very clean way to model such actors is through the introduction of quasi-hyperbolic discounting.² This form of discounting sets up a conflict between the preferences of different intertemporal selves. With assumptions of no commitment and that the agent takes into account her self-control problem, savings decisions can then be modeled as an equilibrium in a sequential game played by the different selves. This modeling paradigm avoids the common connection made between preference changes and cognitive failures,³ and is therefore closer to standard economic analysis. The agent in the model understands perfectly the consequences of her actions, and acts optimally within the constraints imposed by her discount function, which the psychological evidence seems to support at least some of the time,⁴ and the absence of easily available commitment.

Laibson (1997a) analyzed actors of the above kind in detail. His key result is that sophisticated actors with a quasi-hyperbolic discount structure undersave; that is, all intertemporal selves could be made better off if all of them saved a little bit more. Since each self consumes too much from earlier selves' point of view, each of them would agree to increase savings a little bit in exchange for later selves doing the same.

We adapt Laibson's basic setup for the analysis of the effect of endogenous retirement decisions on savings behavior. The addition is simply that in each of the models there is a single period (period 0) in which the agent can choose whether to work or retire. Working costs the agent some utility, but she is compensated for it with extra wealth. We assume that commitment is not possible: agents cannot precommit to a decision concerning retirement, nor to any consumption level. The paper characterizes the savings and retirement outcomes with these preferences as a function of lifetime income and of the additional earnings if retirement is delayed.

There are three types of individual outcomes. Saving and early retirement could be the same as in the situation where work in period 0 were not an option. Similarly, saving and delayed retirement could be the same as in the situation

²Quasi-hyperbolic instead of psychologically more accurate hyperbolic discounting is used only for computational tractability.

³For example, Mischel and Staub (1965) find that subjects fail to understand the contingencies involved in a decision about delay of gratification. See Ainslie and Haslam (1992) for further references.

⁴For example, Ainslie (1992).

where retirement in period 0 were not an option. Interestingly, the former early retirement outcome can be the equilibrium when the late retirement outcome would be the equilibrium if the prior self could commit the later self to work. We refer to the higher level of savings to accommodate such retirement as ‘resigned oversaving.’ A third and distinctive possibility is that savings could be just low enough to ‘force’ work, which we refer to as ‘strategic undersaving.’ This outcome does not match either of the outcomes where there is not a choice about retirement. It is not surprising that the removal of choice can change both the retirement age and savings, as is also true with time consistent preferences. What is different is that the savings to ‘force work’ can be lower than they would be if retirement were not an option, even though the retirement outcome is the same with and without a choice. With time-consistent preferences, removing an option that is not chosen cannot change behavior.

While the paper contrasts outcomes with and without a choice about work without explicitly modeling a change in the underlying economic environment, the results can be interpreted as relating to policy changes. The simplest interpretation is the introduction of mandatory retirement, thereby replacing a choice whether to retire or not by definite retirement. With time consistent preferences, a worker retiring before the new mandatory retirement age would not change behavior because of the introduction of mandatory retirement. The same is true with the quasi-hyperbolic discounting that we model.

The alternative of the disappearance of the opportunity to retire is more complicated to envision and more interesting. Consider a worker who could retire at the earliest age of eligibility for (illiquid) social security benefits, but chooses to work for one more period. Assume she is doing some saving in every period and so satisfies the first order conditions we analyze. If the earliest retirement age were increased, (with benefits unchanged at the later retirement age) a time consistent worker would not change behavior. However, a quasi-hyperbolic worker might respond by saving more while still retiring at the same age. In this case, greater savings did not happen when there was a choice because with greater savings, the later self would have chosen early retirement, while the earlier self preferred later retirement. Note that all the selves prefer the changed outcome when the early retirement option is removed. We defer a systematic analysis of the effects of social security to a later paper that recognizes liquidity issues.

The paper also considers a setting where the earlier self can commit the later self to a given retirement age, although no commitment is possible on future savings. In part, this is simply a way to pick out the interesting examples of removing options. In part it can be interpreted in terms of a choice between two different employers with different defined benefit plans. Consider a worker choosing between two firms. As a function of the length of career each firm offers a lifetime compensation level. A time-consistent worker would plan savings and retirement based on the maximal level of lifetime income for each retirement age (the outer envelope). However, a quasi-hyperbolic worker would also pay attention

to the incentives to work inherent in the lifetime earnings profile. Choosing between the firms might be equivalent to a commitment device on work if the firms differ in the payoff to the last period of work while not differing in lifetime compensation for the planned length of career. If one offers little, while the other offers such a large amount that work will be worthwhile at the optimal retirement age with optimal savings, a commitment to a firm is effectively a commitment on retirement age.

We start with the simplest model that is relevant in (quasi-)hyperbolic discounting: a three-period model in which the middle period is the retirement decision period, period 0. The crucial intuition is that part of the payoff from self 0's working accrues to self 1 through higher savings. But in a quasi-hyperbolic framework, self 0 cares less about self 1 relative to self 0 than self - 1 does. So there will be circumstances where self - 1 would want self 0 to work (for the benefit of self 1), but self 0 does not want to work. In order to avoid this outcome, self - 1 might save less (than she would if she could commit self 0 to work) to 'force' self 0 to work. On the other hand, if self - 1 would like self 0 to work, but it is too expensive to achieve that without commitment, she will save more (than if she could commit self 0 to work) to help finance self 0's unavoidable retirement. Note the qualitative distinction between a change in self - 1's saving (compared to a setting with commitment) to induce a retirement decision and to accommodate one. Here, we can get lower saving to block the 'threat' of retirement and higher saving to accommodate it.

Things get much more complicated when we allow for more periods before retirement. In a four-period model we show a possible conflict that a later self plans to retire too *late*, not too early, from earlier self's point of view. This is because with quasi-hyperbolic discounting successive selves agree in what the later selves should do, but they don't agree on how much it is worth to induce them to do it. And the earlier pre-retirement self will always prefer for the later pre-retirement self to save more than the later wants to save. Thus we can get higher saving to 'encourage' early retirement.

The paper also considers how a retirement decision affects the ability to observationally distinguish quasi-hyperbolic and exponential discounting.⁵ The most radical difference from the predictions of consistent preference models emerges when we consider the effect of an increase in wage level in the endogenous retirement period. In a situation of strategic undersaving, the need for lower savings to induce work is relaxed through higher earnings, so the agent will save more, giving a negative marginal propensity to consume out of changes in future earnings.

Also, we briefly discuss the potential outcomes under the assumption of naiveté, that each self falsely assumes that the others will comply with her plans. Since

⁵As Laibson (1997a) has noted, in the savings game the path of consumption can't be used to distinguish the two, only some comparative statics observations can.

there is no game in this case, the analysis is considerably simpler. One interesting implication of naivet e is the possibility that the selves before the deciding self plan to retire late, but the deciding self chooses to retire early, leading to an update in lifetime wealth and thus a drop in the consumption path at retirement.

2. The quasi-hyperbolic discounting setup

We adapt the structure recently used by Laibson for analyzing quasi-hyperbolic discounting issues. For a more detailed introduction, see Laibson (1997a), for example. The consumer’s instantaneous utility function is of the constant relative risk aversion (CRRA) class, that is:

$$u(c) = \frac{c^{1-\rho}}{1-\rho} \text{ if } \rho \neq 1 \text{ and } u(c) = \ln(c) \text{ if } \rho = 1 \tag{1}$$

ρ being the risk aversion parameter. A nice property of CRRA utility functions is the fact that for intertemporal maximizations of the form:

$$\begin{aligned} &\max_{c_1, c_2} u(c_1) + \kappa u(c_2) \\ &\text{s.t. } c_1 + \frac{1}{R} c_2 = W \end{aligned} \tag{2}$$

with κ a positive discount factor, the solution will always be $c_1 = \lambda(R, \kappa) W$ for some $0 < \lambda(R, \kappa) < 1$. Also, then, some easy manipulation shows that lifetime discounted utility can be written as $K(R, \kappa) u(W)$ (or $K(R, \kappa) + u(W)$ for $u(c) = \ln(c)$) for a positive function $K(R, \kappa)$. This allows us to collapse periods where we have already solved the problem and gotten linear answers into a single period, a shortcut extremely convenient for backward induction arguments. We will use this property a number of times in the paper.

In a T -horizon game, self t ’s discounted utility from present and future consumption is:

$$u(c_t) + \beta \sum_{i=1}^{T-t} \delta^i u(c_{t+i}) \tag{3}$$

with an expectation at front if there is uncertainty. β and δ (both between 0 and 1) are discount parameters meant to capture the essence of hyperbolic discounting, namely that the discount factor between adjacent periods close by is smaller than between similar periods further away. Indeed, the discount factor between periods t and $t + 1$ is $\beta\delta$, and between any two adjacent periods later it is δ .

Of course, the discount structure just described applies only to self t ; for example, self $t + 1$ ’s discount factor between $t + 1$ and $t + 2$ is $\beta\delta$. Therefore, there is a conflict between different selves about how much to consume (or whether to retire) in a given period, or, more formally, preferences are inter-

temporally inconsistent. We assume that commitment is not possible (so that each self controls her period's consumption, subject to a financial or wealth constraint,⁶ and possibly a decision concerning retirement), and model the behavioral decisions as a subgame-perfect equilibrium of the game played by the different selves.⁷ Finally, R is the constant and exogenous gross return on wealth.

3. Three-period model

We begin with a three-period model, the shortest possible that actually generates time inconsistency effects. The periods are labeled $-1, 0, 1$, and subscripts on c or W refer to the period in question. In the period -1 , the agent has to work; in period 0 , she can decide whether to work or retire; and in period 1 , she has to be retired.⁸ The agent incurs an additive constant utility cost of effort $e > 0$ if she works in period 0 , but she also gets an extra Δ amount of income if she does.

As usual when looking for subgame-perfect equilibria, we solve backwards. The decision is easy in period 1 : no work is done and all remaining wealth is consumed. Suppose, then, that the period 0 self inherits a wealth of W_0 . This will be her remaining wealth if she retires, and she will have $W_0 + \Delta$ if she works. As we have mentioned above, there is a $\lambda > 0$ such that self 0 will always consume a proportion λ of her wealth. Thus her discounted utility is:

$$u(\lambda W_0) + \beta \delta u(R(1 - \lambda)W_0) \quad (4)$$

if she doesn't work, and:

$$u(\lambda(W_0 + \Delta)) + \beta \delta u(R(1 - \lambda)(W_0 + \Delta)) - e \quad (5)$$

if she works. Therefore, she will work iff:⁹

$$u(\lambda(W_0 + \Delta)) - u(\lambda W_0) + \beta \delta u(R(1 - \lambda)(W_0 + \Delta)) - \beta \delta u(R(1 - \lambda)W_0) \geq e \quad (6)$$

⁶Depending on whether there are liquidity constraints.

⁷The game theory-based decision rule is basically equivalent to the assumption of sophistication on the part of the agent. An alternative assumption is naiveté, where each self naively assumes that others will follow her decisions. We will study naifs briefly in Section 7.

⁸We use this somewhat odd notation because we will add periods before retirement. To make it easier to compare results, in each of the models we assume that period 0 is the retirement decision period.

⁹We are assuming for now that the agent will work if she is indifferent. In the long-horizon models, we will more generally assume that an agent indifferent between two actions will choose the one the earlier selves would prefer. (With quasi-hyperbolic discounting, all earlier selves want the same thing.) It turns out that this gives the essentially unique subgame-perfect equilibrium—otherwise, the earlier self's maximization problem has no solution.

Since u is concave, there is a \bar{W}_0 such that self 0 will retire iff $W_0 > \bar{W}_0$.

Now let's look at this situation from the point of view of self -1. She will prefer self 0 to work if:

$$\beta\delta u(\lambda(W_0 + \Delta)) - \beta\delta u(\lambda W_0) + \beta\delta^2 u(R(1 - \lambda)(W_0 + \Delta)) - \beta\delta^2 u(R(1 - \lambda)W_0) \geq \beta\delta e$$

or

$$u(\lambda(W_0 + \Delta)) - u(\lambda W_0) + \delta u(R(1 - \lambda)(W_0 + \Delta)) - \delta u(R(1 - \lambda)W_0) \geq e \tag{7}$$

Notice that the left-hand side of 7 is greater than the left-hand side of 6; consequently, there is a range of wealth levels for which self 0 wouldn't work, but self -1 would like her to. In particular, for $W_0 = \bar{W}_0$ self 0 is indifferent between working and not working, but self -1 strictly prefers her to work. This effect arises simply because self -1 weighs the cost and the benefit of working in period 0 differently: for her, the cost is less salient. We assume that if self 0 is indifferent, self 0 chooses to work.

Fig. 1 displays the continuation utility for self -1 (her utility from periods 0 and 1) as a function of W_0 , the level of wealth self -1 leaves for self 0, for an example with logarithmic utility. The curve that starts off as a solid line and continues as a dotted one (U_w) is self -1's continuation utility assuming self 0 works, and the other curve (U_r) is her continuation utility assuming self 0 doesn't work. Only the solid part of each curve is available to self -1, as she has to take into account self 0's retirement decision, based on the relative sizes of W_0 and \bar{W}_0 . Nevertheless, the simplest way to understand self -1's maximization problem is through the continuation utilities U_w and U_r . Define s_i^* (i being r or w) to be the wealth received by self 0 in the solution to the maximization problems of self -1 that assume that retirement or work is exogenous:

$$\max_s u\left(W_{-1} - \frac{1}{R} s\right) + U_i(s) \tag{8}$$

Note that $s_w^* < s_r^*$ since work provides extra income in period 0, and some of that is consumed in period -1. If self -1 could commit self 0 to a decision on work (but not on consumption), she would choose one of these savings levels. We describe optimal savings levels using these constructs, and compare optimal savings with endogenous retirement, to those with exogenous retirement and to the savings that would result if self -1 could commit to self 0's retirement decision. We denote the optimal savings level with endogenous retirement by s^* .

Fig. 2 shows lifetime discounted utilities for self -1 as a function of W_0 assuming work and retirement in period 0 for the same example as in Fig. 1. Again, the solid part of each curve is available to self -1. s_w^* maximizes the work curve, s_r^* the retirement curve, and, as is clear from the figure, since U_i are concave, the best available point is one of the three points, s_w^* , s_r^* , and the level of savings that would just induce work, \bar{W}_0 . In this example, it seems to be \bar{W}_0 .

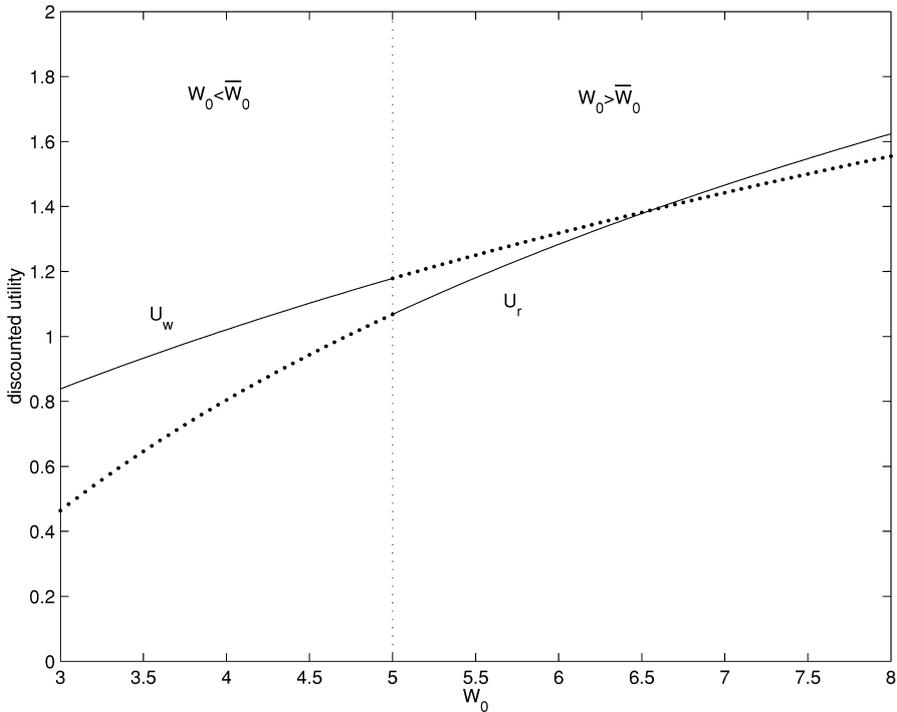


Fig. 1. Utility of self - 1 from periods 0 and 1 with and without work in period 0.

We turn now to some comparisons. Assume that there is mandatory retirement in period 0. This results in savings s_r^* . Assume that mandatory retirement is repealed and consider cohorts young enough to make their savings decisions in period -1. Some workers will continue to plan to retire in period 0 and will continue to save at the level s_r^* . Other workers will plan to work in period 0 and will change their savings level to one of the choices, s_w^* , or \bar{W}_0 .

There is an interesting contrast in the reverse comparison. Assume that work were mandated in period 0 (an odd assumption that will be justified in a moment), implying savings of s_w^* . Assume that the mandate is dropped. As a result, some workers might change their work plan and their savings plan—retiring in period 0 and saving s_r^* . What is different about the quasi-hyperbolic setup is that some workers will not change their work plan and will change their savings plan nevertheless—changing from s_w^* to \bar{W}_0 . Note that every self of a worker who changes savings without changing work would prefer the mandate to work. For a worker who continues to have late retirement, this scenario, of an end of a mandate to work, could follow from a decrease in the early entitlement age for social security (assuming a liquidity constraint blocked early retirement before

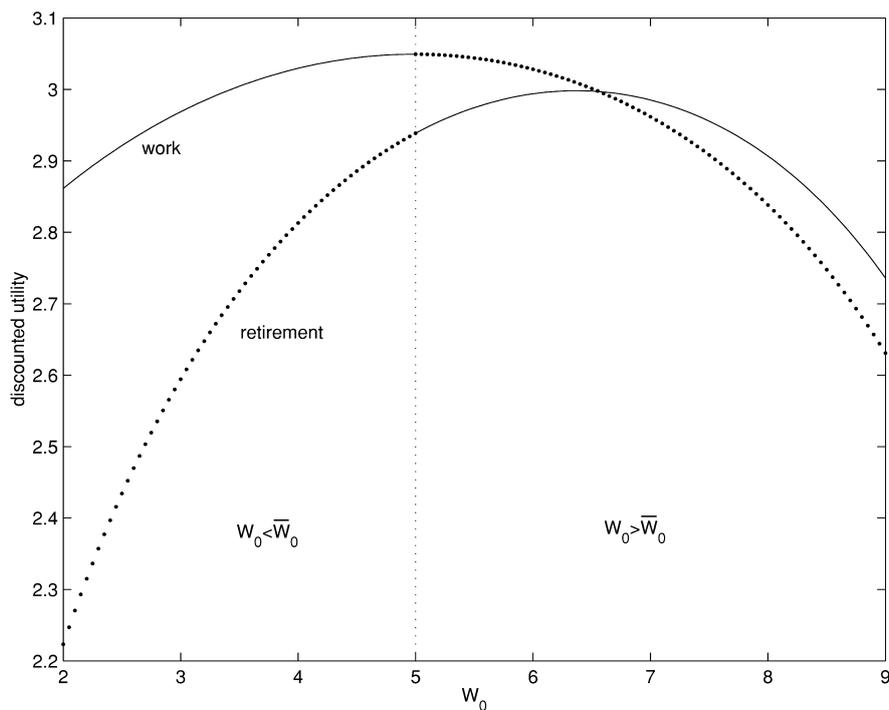


Fig. 2. Lifetime utility of self -1.

benefits). Alternatively, it might result from the repeal of a large implicit subsidy on work in period 0, as could occur with a change in a defined benefit pension plan that did not change the level of benefit at the previous equilibrium retirement age. Thus, with the end of a mandate to retire, savings only change if the work plan changes. In contrast, with the end of a mandate to work, savings can change even if the work plan does not change.

Another comparison of interest is between the outcome without commitment and the one where self -1 can make a commitment about work in period 0 (although not about savings in period 0). For example, self -1 might be choosing between two different firms with different defined benefit plans that have such powerful (and different) incentives that they are equivalent to choosing whether to work in period 0 or not. If the solution to the commitment problem is to have retirement in period 0, then that is also the solution to the problem with no-commitment. If the solution to the no-commitment problem is to have work in period 0 and if the level of savings is s_w^* , then that is also the solution to the problem with commitment. In contrast, if the solution to the no-commitment problem is to have work in period 0 and if the level of savings is \bar{W}_0 , then, with

the ability to commit, there would be continued work, but savings would rise to s_w^* .

Recapping, if, with a commitment mechanism, self -1 wants self 0 to retire, an inability to commit to retirement does not matter—removing the commitment device would change neither work nor savings. However if with a commitment mechanism, self -1 wants self 0 to work, there are three possibilities. Removal of the commitment device might have no effect on either savings or work, might change savings (from s_w^* to \bar{W}_0) while preserving work, or might change both savings (from s_w^* to s_r^*) and from work to retirement.

4. Comparative statics in the three-period model

4.1. Changing wealth

For low values of W_{-1} , $s_w^* < \bar{W}_0$, self 0 works in equilibrium and the inability of self -1 to commit self 0 to work has no effect. Then, there is a range of values for W_{-1} such that optimal savings equals \bar{W}_0 in order to just induce work. Over this range savings are less than they would be if self -1 could commit self 0 to work. We call this kind of equilibrium one of ‘strategic undersaving.’ In the next range of W_{-1} , self -1 accommodates self 0 ’s desire not to work, saving $s_r^* > \bar{W}_0$ even though self -1 would save less and commit self 0 to work if that were possible. This equilibrium type is called ‘resigned oversaving.’ For high enough values of W_{-1} , self -1 prefers that self 0 retire and there is, again, no effect from the inability to commit. This is shown in Fig. 3.

The marginal propensity to consume in period -1 out of a small increase in W_{-1} behaves differently in the different regions.¹⁰ In the lowest region, for a small increase in W_{-1} , the fraction of the increase consumed is λ_{-1} , just as in the case without a retirement decision. For a small increase in wealth in the strategic undersaving region (II), all of it is consumed so that self 0 continues to receive \bar{W}_0 . For small increases in wealth in the top two regions, again, the fraction λ_{-1} is consumed in period -1 .

When interpreting results in these short-horizon models, we have to be very careful not to confuse genuine quasi-hyperbolic discounting effects with effects that arise due to the fact that we have chosen a short horizon. In particular, you might notice that even if optimal savings satisfies $s^* = s_i^*$ for an i , we don’t have $c_{-1}/c_0 = c_0/c_1$ as we do for exponential discounting with CRRA utility functions. But this peculiarity occurs only because the marginal propensities to consume change from period to period, a property that disappears as the horizon after

¹⁰By a small increase we mean one that does not move self -1 into a different region.

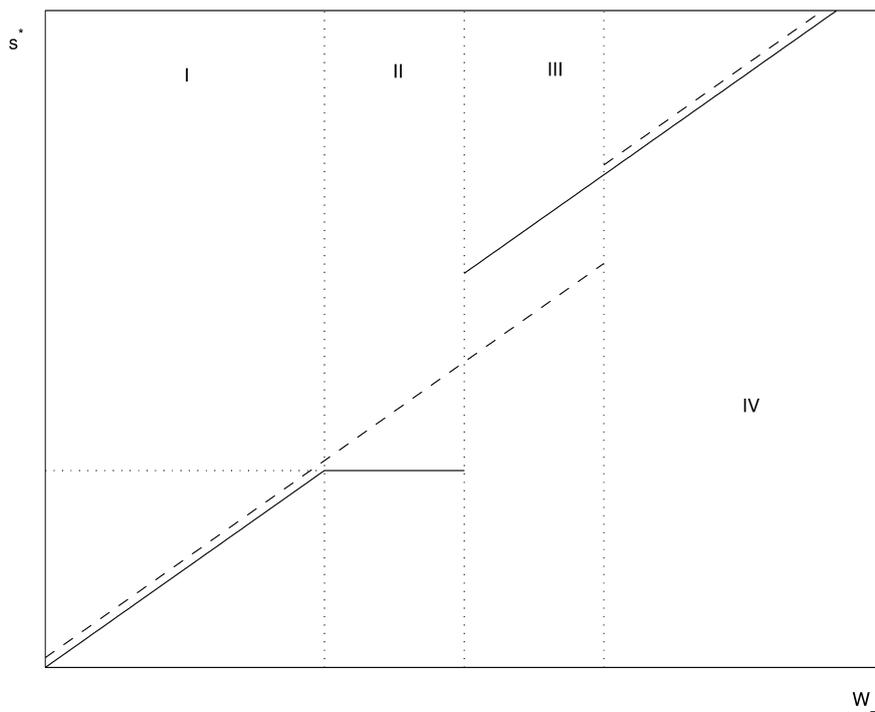


Fig. 3. Savings of self -1 with (dashed line) and without (continuous line) commitment. (Note: the figure is only qualitative; it is not meant to illustrate actual slopes or relative sizes for the regions.).

retirement is assumed to go to infinity.¹¹ In that case, only the equivalent of $s^* = \bar{W}_0$ will not satisfy the equivalent of $c_{-1}/c_0 = c_0/c_1$. Strategic undersaving is the only outcome observationally different from exponential discounting: it is the only case when self -1 uses non-optimal savings (in the sense of the consumption game) as a tool to change the retirement decision of self 0. And as Laibson has pointed out in the context without a retirement decision, long-horizon optimal savings with quasi-hyperbolic discounting is observationally equivalent to exponential discounting (Laibson, 1997a). Non-optimal savings, finally, is not

¹¹The marginal propensity to consume matters with quasi-hyperbolic discounting simply because the Euler equation contains it:

$$\frac{u'(c_t)}{u'(c_{t+1})} = R\delta \left(\beta \frac{\partial c_{t+1}}{\partial W_{t+1}} + 1 - \frac{\partial c_{t+1}}{\partial W_{t+1}} \right)$$

This is proved in (Laibson, 1997a) but also falls out as a special case of our analysis in Appendix D.

possible with exponential discounting, even in the presence of a retirement decision: in that case \bar{W}_0 is defined by the intersection of the curves U_w and U_r , so s_w^* and s_r^* both dominate it, and one dominance is strong.

Behaviorally, as opposed to just observationally, there is another, more subtle, difference between quasi-hyperbolic and exponential discounting: resigned oversaving. It is possible that self -1 would prefer to commit self 0 to work and give her s_w^* , but since that is not possible and she has to undersave too much to make self 0 work, she chooses s_r^* . This reason for choosing s_r^* , though unobservable, is unique to quasi-hyperbolic discounting: it arises from the conflict of self 0 's decisions and self -1 's wishes. However, the reason for higher saving in this case (clearly $s^* > s_w^*$) is very different from the reason for lower saving above:¹² it is not intended to change the retirement decision of self 0 . Quite the opposite: in recognition of the fact that it would be 'too expensive' to change self 0 's decision, self -1 will save more to offset the lower wealth level of self 0 due to the early retirement. In fact, self -1 can end up saving more than if the agent were time consistent, a result qualitatively different from the equilibrium with only a savings decision.

4.2. Changing earnings

The comparative statics for savings in period -1 with respect to Δ is illustrated in Fig. 4. Savings with and without commitment by self -1 are shown. For very low levels of Δ (region I) it is not worth working, so the agent just saves from her other wealth for retirement. These savings don't depend on Δ , as period 0 income is never realized. In region II, self -1 would prefer self 0 to work if she could commit her to do it, but, without it, it is better to retire early, resulting in a different savings level. The most interesting region is the next one, region III. Here, self -1 undersaves to make self 0 work, giving her exactly $s^* = W_0$. Since \bar{W}_0 increases with Δ , s^* is increasing; furthermore, this is the only region in which s^* is in general not a linear function of Δ . In contrast, with commitment to work, a higher wage leads to lower savings as consumption in all periods rises with lifetime earnings. And finally, for high levels of Δ , region IV, the equilibrium involves work, and it is once again equivalent to the commitment solution. Notice that these regions are in exactly the opposite order as in Fig. 3—higher levels of wealth and higher levels of earnings have opposite incentive effects for retirement. Indeed, it is the difference between marginal propensities to save out of earned (future) income and unearned income that most sharply distinguish the two types of behavior.

In a certainty setting, we have different cases with the separate possibilities of strategic undersaving and resigned oversaving. In Appendix A, we briefly examine

¹²The benchmark for all savings discussions at this point is still the savings level that would arise if self -1 could control self 0 's retirement decision.

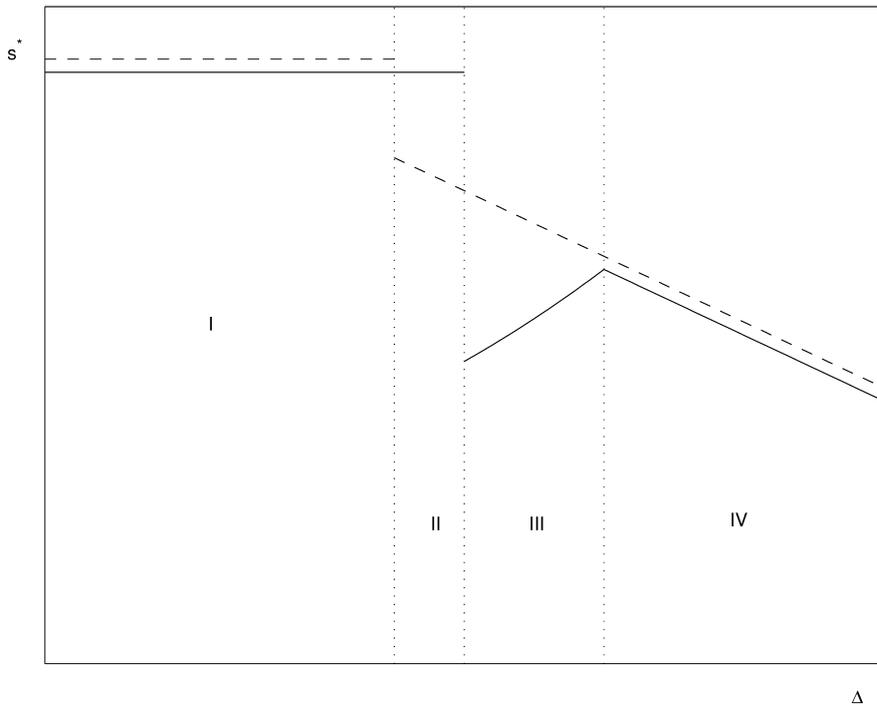


Fig. 4. Savings of self -1 with (dashed line) and without (continuous line) commitment as a function of self 0 's wage income. (Note: the figure is only qualitative; it is not meant to illustrate actual slopes or relative sizes for the regions.).

a setting with uncertainty. Adding uncertainty eliminates case analysis and transforms it into effects analysis, thereby also allowing a delineation of when higher or lower saving is likely to occur. With enough spread in the probabilities so that states arise where each of the effects might be present, the first order condition for savings includes both types of incentives.

In the savings game without a retirement decision, a long horizon tends to make marginal propensities to consume approximately equal across periods. This helps both in describing the quasi-hyperbolic equilibrium and in comparing it with the exponential discounting outcome. As we have mentioned, this is also the case in our model for consumption after retirement. But this would not affect any of our results, so we don't present it in this paper.

Though introducing a long horizon after retirement is of little consequence to the qualitative results of the impact of retirement choice on pre-retirement savings, a longer horizon before retirement does set up a novel distinction: how the effects play themselves out close to versus far from retirement. Unlike in the savings game, what happens at the end is no longer an empirically unattractive theoretical

nuisance; the behavior is not a response to nearby deterministic death, but to the approach of the end of working life—the central focus of this paper. Thus, we will ‘move backwards’ in the next section, and see what happens when the horizon before retirement is let to grow. Some distinctive new effects arise.

5. Four-period model

Unfortunately, there is very little we can say about the equilibrium in general with many periods of work. The bulk of the trouble stems from the fact that when later selves have decreasing marginal propensities to consume,¹³ the consumption schedules of a quasi-hyperbolic discounter become extremely complicated very quickly as we move to earlier periods. All we know is that the agent’s consumption schedule is piece-wise linear in wealth for each t , and, furthermore, the agent’s consumption path is as if she were going through a series of shorter Laibson problems.¹⁴ This is quite interesting in itself: the agent periodically acts as if she is liquidity constrained and/or impatient, even though she has perfect foresight and faces no constraints. But since we are unable to say much in general about the equilibrium, we will mostly restrict our attention to a model in which there are only two periods of exogenously mandated work before the period of endogenous decision.

For the three-period model, the inability of self -1 to commit self 0 to a retirement decision was reflected in savings. For earlier selves before retirement, the conflict is not only between the current self (say self t) and the self making the retirement decision—there are selves in-between with whom self t may also have a conflict. An important implication of this is that self t *might not want to commit to a retirement decision*. Commitment also allows other selves to behave differently, which self t might not like.¹⁵ Fortunately, this issue is not too critical if the earliest self considered is self -2 .

In the four-period model, the behavior of selves 1 , 0 , and -1 is the same as in Section 3. Let us now move back to self -2 and see what she thinks about the behavior of self -1 . (Notationally, we include the PDV of earnings in all periods

¹³Laibson (1997b) describes such an example in detail, though in the context of liquidity constraints. Here, since self 2 in Section 3 has a region where her marginal propensity to consume is 1 , in that region she behaves as if ‘liquidity constrained.’ This gives the jumps in consumption earlier on.

¹⁴See Appendix B for a formal statement and a proof. We will also take advantage of the fact that a Markov-perfect equilibrium in pure strategies exists for the game. Noticeably, that proof in the appendix uses similar methods to those below, but putting it there and just assuming existence for now makes the paper much easier to follow.

¹⁵We could say that we are comparing things to when self t is forced to make a commitment, but if that is against self t ’s will, the interpretation of the results is ambiguous.

except 0 in the wealth measure W , considering separately only the possible earnings Δ in period zero.) We view separately the three cases considered above, where self -1 chooses s_r^* , s_w^* , or \bar{W}_0 . The latter has the most interesting structure. Assume that self -1 is indifferent between having self 0 work or not. Then, self -2 strictly prefers that self -1 induce self 0 to retire. This result follows from the fact that self -2 views the cost of more saving by self -1 as less expensive than does self -1 , who finds the drop in consumption to finance early retirement more salient.

We begin with a lemma that extends the normality result to multiple periods—lesser wealth can not change the savings plan from one that induces work to one that induces retirement:

Lemma 1. *Suppose that $t \leq 0$ and $W_t > W'_t$. Then it is not possible that self t with wealth W_t behaves so that self 0 eventually works, and with wealth W'_t she behaves so that self 0 eventually retires.*

The formal proof is in Appendix C. It takes advantage of the concavity of consumption utility to show that savings is monotonically increasing in wealth for each self before zero. This implies that self 0's wealth is monotonically related to previous selves' wealth levels. And we know self 0 retires iff $W_0 > \bar{W}_0$ for a given \bar{W}_0 .

We turn now to the nature of the conflict between self -2 and self -1 . Self -2 will never use boundary (knife-edge) savings to get the working alternative, but it is possible she will use it to 'force' retirement. This was discussed above, and is exactly what the following lemma proves.

Lemma 2. *Let \bar{W}_t ($t < 0$) be the level of wealth at which self t is indifferent between behaviors that eventually lead to self 0 working or retiring. At this savings level, self $t - 1$ strictly prefers self t to choose to eventually make self 0 retire.*

Once again, the proof is in Appendix C, but its essence is simple: due to the different preferences, self t cares relatively more about consumption in period t than does self $t - 1$, so when self t is indifferent, self $t - 1$ wants her to go for the low-consumption (high-saving) option. And this is of course the early retirement option. Self -2 , then, might save more than with mandated early retirement to just induce self -1 to save so as to result in early retirement.

These lemmas can be used to illustrate self -2 's general qualitative savings behavior relative to wealth, which is done in Fig. 5. For very low levels of wealth, self -2 prefers late retirement, and this can be achieved with savings that satisfy

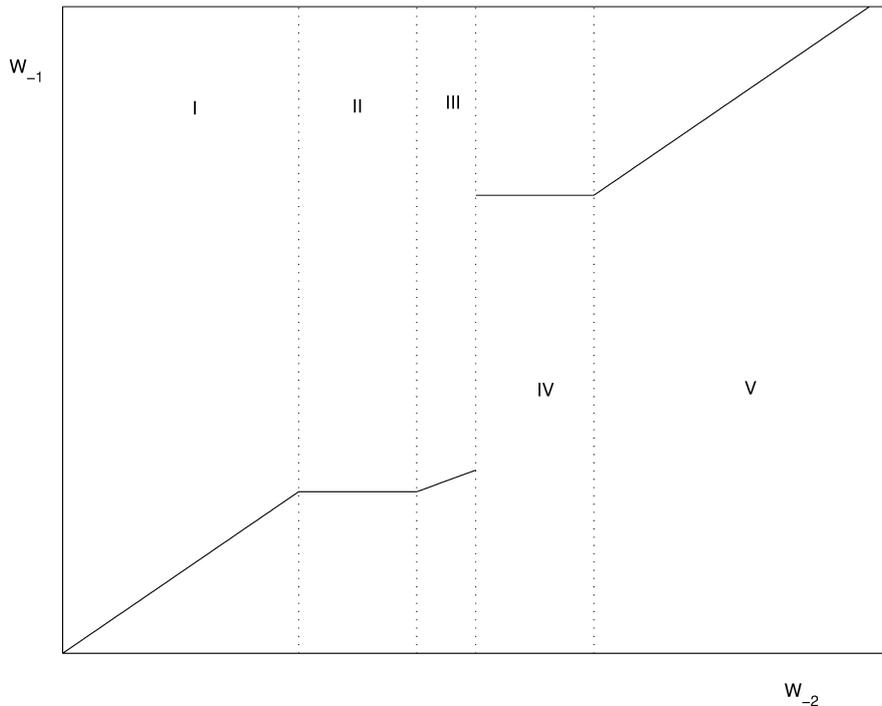


Fig. 5. Savings of self -2 (Note: the figure is only qualitative; it is not meant to illustrate actual slopes or relative sizes for the regions).

the Laibson consumption solution with late retirement and no retirement decision. That is, self -1 chooses the level of s_r^* appropriate for the level of wealth at the start of period -1 .

In the next two regions (II and III), selves -2 and -1 undersave to induce self 0 to work in the sense that self -1 chooses the level of savings to just induce work (\bar{W}_0). The two regions differ in how selves -1 and -2 contribute to this level of wealth at the start of period 0 . By Lemma 1, self -2 can split the undersaving with self -1 , while still inducing eventual late retirement. For relatively low wealth levels where there has to be undersaving done to induce self 0 to work (region II), all the undersaving will be done by self -2 . In this region, self -1 's marginal propensity to consume is 1 , so self -2 prefers to consume all extra wealth as long as $u'(c_{-2}) > \beta \delta u'(c_{-1})$, and her savings function is flat as a function of wealth. As self -2 gets richer, she will want to split the extra consumption with self -1 even though self -1 has a marginal propensity to

consume of 1. In this region (III), we have $u'(c_{-2}) = \beta \delta u'(c_{-1})$, and the savings function is positively sloped, although with a lower slope than in region I.¹⁶

Region IV is the content of Lemma 2—self -2 chooses early retirement, but in order for her to do that, she needs to save enough to ensure that self -1 ends up making self 0 retire. This is more savings than would be done if work were not an option. Again, in this region self -2 consumes all extra marginal wealth, until she is rich enough so that eventual early retirement results without oversaving. And in region V, self -2 consumes according to the Laibson solution, with early retirement and no choice.

This picture illustrates the result, that holds with more periods, that W_t with $t \leq -2$ such that self t wants self 0 to retire, self $t - 1$ wants her to retire as well. The converse of this is not true, that is, if self t chooses to save so that self 0 works, self $t - 1$ might not like that. Translating our intuition from the work equilibrium, we might be led to think that—due to the elimination of this conflict—if retirement were mandated, savings levels would be lower. Such a conclusion is true in the present setup, but not if we go back one more period. Imagine that with the mandate, W_{-1} is slightly above \bar{W}_{-1} , the cutoff wealth level for self -1 , and that self -2 is willing to bequeath higher savings to make self -1 choose early retirement. That is, even for some wealth levels below $1/(R1 - \lambda^*)\bar{W}_{-1}$, self -2 will choose to save \bar{W}_{-1} . Since self -2 overconsumes from the point of view of self -3 , self -3 might choose to lower her savings to self -2 once the mandate is removed. Then self -1 will end up with \bar{W}_{-1} —lower than with the mandate. The key intuition is that self -3 takes advantage of self -2 's efforts to control self -1 's decision for her own purposes.¹⁷

This highlights a key distinction between the early and late retirement outcomes.

¹⁶There are some things that can be said in greater generality. Assume that each self $t < 0$ already prefers early retirement for a low enough wealth level so that there are no jumps in self t 's consumption function on the late retirement section. (We expect the statements that follow to be true even without this assumption, but haven't been able to prove it.) Then it is easy to prove by backward induction and taking advantage of the above lemma that two things are possible. Either self t 's marginal propensity to consume is λ^* up to $1/((1 - \lambda^*)R) \bar{W}_0$, and (if she still prefers late retirement for higher wealth levels) then her marginal propensity to consume is 1 on some non-empty interval. Or self t 's marginal propensity to consume is λ^* up to some lower wealth level, above which she prefers early retirement. This immediately implies two things. First, if mandated work is acceptable to all selves (in the sense that they prefer late retirement at their mandate wealth level), then the outcome of a mandate is an equilibrium even with choice. Second, if this is not the case (the mandate is not acceptable to all selves), then savings for retirement in a work equilibrium without a mandate is lower. Also, if all selves prefer lower saving for at least some wealth levels, then small enough amounts of lower saving will all be done by the first self alive.

¹⁷The same counterexample works to show that the other statement from the late retirement case does not carry over, either: it is not true that if mandated retirement is acceptable to all selves, then the outcome of the mandate is an equilibrium.

When there is higher saving to be done to induce retirement, the earlier selves are by no means as eager to join in as when the task is lower saving. They are actually very happy to let later selves save more, as those selves consume too much from their point of view anyway. They will thus want to have them oversave a lot, often resulting in putting the self at her cutoff wealth level. (In more precise language: a self t who is leaving boundary saving but over her own cutoff wealth level usually has a marginal propensity to consume of 1, thus making the marginal rate of substitution for self $t - 1$ low.) As a consequence, small amounts of lower savings are ‘handled’ by the early selves, while higher saving is pushed on (in an exaggerated manner, in fact) to later ones.

Partly for this reason, it is important to focus on the lower saving outcome if we care about the well-being of the individual as a whole, that is, the set of her intertemporal incarnations. For such an analysis we can use similar tools as in welfare economics. If the outcome is ‘forced’ work, then mandating work in period 0 is a Pareto improvement. The Laibson consumption path with mandated work is already too high, and there is additional consumption done in the periods before retirement if there is not a mandate, making the equilibrium outcome without a mandate Pareto-inferior to a mandate: self -2 would benefit from a better consumption path, selves 0 and up from more savings, and self -1 (possibly) from both. In this strong sense, the equilibrium outcome is suboptimal, and can correctly be termed an undersaving outcome. Similarly unambiguous things cannot be said when the equilibrium has retirement in period 0. Higher saving by a self is in general good for both earlier and later selves but bad for that self. So, on the one hand, commitment might not be desirable, and on the other, its welfare implications are mixed.

In all these proofs we have *very strongly* used the particular structure of quasi-hyperbolic discounting.¹⁸ A troublesome occurrence of this was when we proved that in periods $t \leq -2$ lower saving is not possible in the boundary savings level sense (Lemma 2): the proof depended on the fact that selves t and $t - 1$ have two different weightings of the same utility tradeoff (c_t vs. K_t). Since Laibson introduced quasi-hyperbolic discounting as an approximation to hyperbolic discounting purely for analytical convenience, such results should be handled with some suspicion. In a true hyperbolic discount structure, from the point of view of self -2 , self -1 not only underweights effort in period 0 compared to consumption in period -1 , but she also overweights it compared to consumption after retirement. This results in self -2 choosing to undersave more often than in a quasi-hyperbolic model, where the second conflict is nonexistent. For a formal treatment, see Appendix D.

¹⁸Even the appendix’s proof of the existence of equilibria uses at a crucial point that with quasi-hyperbolic discounting all earlier selves would want a later self to do the same thing.

6. Notes on observational equivalence

One of the important caveats of quasi-hyperbolic discounting is that it is very hard to tell it apart from exponential discounting. Laibson (1997a) noted that an econometrician watching a quasi-hyperbolic discounter, but operating under the assumption of exponential discounting, will get a very good fit for her theory, as consumption paths of the two types of agents look the same. At the same time, she will radically misconstrue the agent's preferences, finding a one-period discount factor of 0.98 instead of 0.6 in a typical example. Only comparative statics involving the interest rate can be used to distinguish actors with self-control problems from the others.

Our models lend themselves to a number of convenient approaches to distinguishing exponential and quasi-hyperbolic discounters observationally. Both the consumption path and some comparative statics results can give a quasi-hyperbolic discounter away.

First, a consumption path that is smooth after retirement and not smooth leading up to it is a sign of quasi-hyperbolic discounting. This is of course due to the changing marginal propensities to consume in the periods preceding retirement. More interestingly, if equilibrium involves work in the period of decision, a lower average consumption rate after retirement than before is consistent with quasi-hyperbolic but not with exponential discounting.¹⁹

Interesting comparisons of a comparative statics nature also emerge. Consider a strategic undersaving equilibrium in the three-period model. If earnings in period 0 (Δ) increase, the period -1 self will save *more*: the extra earnings gives self 0 more incentive to work, lowering the amount of undersaving needed to induce work. Thus, self -1 's marginal propensity to consume out of changes in future earnings is negative. This could never happen with an exponential discounter.

Similarly, if the option to retire in period 0 is eliminated in some way, and self -1 would have undersaved before, she will save more. This is again impossible with exponential discounting: there the elimination of a non-chosen alternative doesn't change the optimum. Also, agents who work in period 0 in equilibrium but don't undersave, will not change their behavior. To check this effect it is, however, necessary to identify those who would have worked had the option been available.

Finally, notice that in the long-horizon equilibrium there is a range of wealth levels where richer people save disproportionately more of their wealth for retirement: as one switches from lower savings and work to retirement, the total

¹⁹It is tempting at first to try to use this as an explanation for the drop in consumption at retirement. There a number of problems, though: first, the drop in consumption occurs at period $t=0$ the latest, that is, *before* retirement. Also, the drop is much too general of a finding for this theory: it happens to almost all groups of people, irrespective of wealth or when they retire (Bernheim et al., 2001).

wealth that self 0 gets switches from \bar{W}_0 to something that is greater than $\bar{W}_0 + \lambda^* \Delta$, a change that is not warranted by the difference in lifetime wealth. Thus, controlling for income, on average the richer people (who retire early) have higher savings rates. While this can be explained by exponential discounting with individual heterogeneity in time preference, it could be explored with an independent measure of time preference.

7. Partially naive agents

As economists, we often assume too much rational capability on the part of humans. Our assumption of full sophistication of the agents is not immune from this criticism. Thus, the literature also considers the opposite extreme assumption of naiveté. (For a contrast of sophistication and naiveté in the context of quasi-hyperbolic discounting, see O'Donoghue and Rabin (1999)). An agent is called fully naive if each of her intertemporal selves assumes that future selves will make the same consumption and retirement decisions as she would. There is no game in this case, and the 'plans' (current decisions and expectations about future decisions) are simply updated each period.

First, let us assume that the agent is naive only about the retirement decision, not consumption. That is, she still plays a Laibson game with respect to consumption, but each self $t < 0$ assumes that self 0 (and others) will make the same retirement decision as she would. This assumption is mostly for analytical convenience, so that the discussion fits more naturally into what we have been doing. But it might also be interesting empirically, because retirement is (mostly) a one-time decision, so people should have less of a chance of learning about their intertemporal conflicts in this area than regarding consumption.²⁰ The assumption implies that self $-n < -1$ expects to work in period 0 iff:

$$V\left(W + \frac{1}{R^n} \Delta\right) - \beta \delta^n e \geq V(W) \quad (9)$$

where $V(W)$ is lifetime utility from consumption when starting with wealth W .

We have the following result:

Theorem 1. *Suppose the horizon after retirement is long. If self $-n < -1$ plans to work in period 0, so does self $-n + 1$.*

This theorem is the consequence of two considerations, one specific to quasi-hyperbolic discounting and one not. First, the Euler equation for consumption

²⁰Note that with this assumption consumption decisions in each period are the same as in the commitment case.

implies that the marginal utility of wealth today is less than $R\delta$ times the marginal utility tomorrow, so an extra amount of income that is R times as much in the future as today should be worth more than $1/\delta$ times in the future as today. And since in the future the cost will be perceived to be $1/\delta$ times as much, the future self is more likely to want to work. Second, the future self is additionally motivated to want to work as the current self consumes part of the planned income in period 0. The latter argument does not rely on quasi-hyperbolic discounting, while the former one does.²¹ The proof of Theorem 1 takes advantage of both in a tricky way. It could be simplified, but the given form allows for two generalizations.

It is easy to show that if n is sufficiently large, then self 0 will actually work. Also, though the problem seems different on the surface, the theorem is exactly the same if the agent is also naive about consumption decisions.

The converse of Theorem 1 is not true—if self $-n < -1$ wants to retire early, self $-n + 1$ might change her mind.

It is, however, true that if self -1 wants to retire in period 0, self 0 will actually do so. To see this, note that if self 0 were to work, that would be better for self -1 as well, and with optimal consumption it would be better still. Again, the converse of this is not true: it could happen that self -1 plans to work in period 0, but self 0 decides to retire. In this case, lifetime wealth is updated downwards (self -1 believes that period 0 earnings are a part of wealth), so there is a downward jump in the consumption path. In contrast to the sophisticated case, this occurs *exactly* at retirement, as actually observed empirically (Bernheim et al., 2001).

8. Conclusion

This paper makes an addition to the classic quasi-hyperbolic discounting savings model. Its technical contributions are minor—most of the analysis is possible with little more than the tools developed by David Laibson. However, the interaction of two decisions, with the one (savings) available as a tool to influence the other (retirement), changes the classic model in a few interesting ways.

One is the possibility of additional undersaving with the eventual consequence of making the self with a choice poor enough so that she will want to work. This strategic undersaving occurs *in addition* to the undersaving that characterizes the equilibrium without a retirement decision. It therefore aggravates an already inefficient outcome, and is likely to be bad for all selves.

The other, and perhaps more novel, effect is the possibility of higher saving than

²¹Without the assumption that the horizon after retirement is long, the statement of the theorem may not be true. As the agent approaches the last period, there is less reason to work, since there are fewer periods in which to consume. This effect acts opposite to those discussed above. However, it is unlikely to be important in reality.

when commitment is possible. Higher saving can occur for two reasons: either because it is too costly in terms of discounted utility to make the deciding self work, and thus one would rather finance her retirement, or because self $t \leq -1$ is too eager to work long and it is worth saving more to make her choose early retirement. Unlike undersaving, it is not in general bad for the individual—it can mitigate the overconsumption equilibrium of the classic model. In fact, higher saving seems never to be Pareto-worsening: the later selves, at least, should be happy about getting more savings.

We also noted some effects of mandates that are not present with exponential discounting. It might be possible to find ‘natural experiments’ changing work and retirement opportunities.

The theoretical model would benefit from two major extensions. One is the introduction of more periods when the agent can choose whether to work. We have solved a model of this sort without savings: in each period, the agent can decide whether or not to retire (the retirement decision being final.) and consumption just equals income or benefits. To make it an interesting problem, one has to assume, for example, a benefit profile that increases with the age of retirement. In equilibrium, the agent retires too early: the retirement date is Pareto-dominated by a later retirement date. No such results emerge in our models with savings, but they might if there are more periods of retirement decisions.²²

Another useful extension would be the investigation of liquidity constraints in this context. They are clearly important in practice, and they change the nature of equilibria with quasi-hyperbolic discounting considerably. They would play an important role in the analysis of social security since the payment of benefits as an annuity can have independent effects from the mandate to save.

A perplexing aspect of quasi-hyperbolic discounting models is a question that is very hard to answer: why don't people take advantage of annuity-type commitment devices to overcome their undersaving problem? These financial tools are readily available but rarely used. Some modestly satisfactory reasons can be brought up. First, if there is a bequest motive, then, just like in many exponential discounting models, annuities look less attractive than without a bequest motive. Second, the annuities market is quite complicated, and there are good reasons for boundedly rational people not to enter markets they know little about. The latter seems to indicate that as people learn about annuities they may come into broader use. Even if that happens, the commitment is unlikely to be full, leaving at least some room for quasi-hyperbolic discounting effects. In the absence of annuities, there is of course a wide-spread institutional structure that serves as a commitment device for agents happy or unhappy about it: social security. We plan to study the

²²The Pareto-improving retirement date is at least two periods later than the equilibrium date t : otherwise self t wouldn't want to retire. Then it is not a major surprise that our models don't generate too early retirement.

implications of the joint mandates of savings and receipt of social security benefits as a real annuity in a later paper.

Acknowledgements

We are grateful to the National Science Foundation for research support (under grant SBR-9618698).

Appendix A. Uncertainty

We could introduce uncertainty in period 0 labor income (Δ), and in period 0 cost of effort (e). The two give similar results, and the latter is somewhat nicer to present, so we present only that one. Assume therefore that self -1 doesn't know e , but knows its continuous density function f (and the cumulative distribution function F). This standard assumption is made plausible by the possibility that the agent does not know how healthy or how thrilled she will be to work in the future. We assume that the support of f is wide enough to encompass all of the regions above.

We start again from self 0's problem, who has inherited a wealth W_0 . Define $\bar{e}(W_0)$ as the level of effort cost at which self 0 is indifferent to work:

$$u(\lambda(W_0 + \Delta)) + \beta\delta u(R(1 - \lambda)(W_0 + \Delta)) - \bar{e}(W_0) = u(\lambda W_0) + \beta\delta u(R(1 - \lambda)W_0) \tag{A.1}$$

Self 0 will work if $e \leq \bar{e}(W_0)$. Therefore self 0 will work with probability $F(\bar{e}(W_0))$ and retire with probability $1 - F(\bar{e}(W_0))$. For simplicity, let K be the constant such that $\beta\delta u(\lambda W) + \beta\delta^2 u(R(1 - \lambda)W) = Ku(W)$. As we have mentioned, such a constant always exists for CRRA utility functions.²³ Now the maximand for self -1 is:

$$u\left(W_{-1} - \frac{1}{R}W_0\right) + K[F(\bar{e}(W_0))u(W_0 + \Delta) + (1 - F(\bar{e}(W_0)))u(W_0)] - \beta\delta \int_0^{\bar{e}(W_0)} ef(e) de \tag{A.2}$$

The first-order condition is:

²³When the utility function is logarithmic, the correct expression is $\beta\delta u(\lambda W) + \beta\delta^2 u(R(1 - \lambda)W) = K + u(W)$. The analysis is the same in this case.

$$\begin{aligned} \frac{1}{R} u' \left(W_{-1} - \frac{1}{R} W_0 \right) &= K[F(\bar{e}(W_0)) u'(W_0 + \Delta) + (1 - F(\bar{e}(W_0))) u'(W_0) \\ &+ f(\bar{e}(W_0)) \bar{e}'(W_0) u(W_0 + \Delta) - f(\bar{e}(W_0)) \bar{e}'(W_0) u(W_0)] - \beta \delta \bar{e}(W_0) f(\bar{e}(W_0)) \bar{e}'(W_0) \end{aligned}$$

which is equivalent to:

$$\begin{aligned} \frac{1}{R} u' \left(W_{-1} - \frac{1}{R} W_0 \right) &= K[F(\bar{e}(W_0)) u'(W_0 + \Delta) + (1 - F(\bar{e}(W_0))) u'(W_0)] \\ &+ f(\bar{e}(W_0)) \bar{e}'(W_0) [Ku(W_0 + \Delta) - \beta \delta \bar{e}(W_0) - Ku(W_0)] \end{aligned} \tag{A.3}$$

A similar first-order condition would arise if self -1 could commit self 0 to a state-contingent retirement decision,²⁴ except that $\bar{e}(W_0)$ should be replaced by $\tilde{e}(W_0)$, where $\tilde{e}(W_0)$ is defined by

$$\begin{aligned} \beta \delta u(\lambda(W_0 + \Delta)) + \beta \delta^2 u(R(1 - \lambda)(W_0 + \Delta)) - \beta \delta \tilde{e}(W_0) &= \beta \delta u(\lambda W_0) \\ + \beta \delta^2 u(R(1 - \lambda) W_0) \end{aligned} \tag{A.4}$$

(This just defines the cutoff cost level under which self -1 would want self 0 to work.) Then, by definition, $Ku(W_0 + \Delta) - \beta \delta \tilde{e}(W_0) - Ku(W_0) = 0$, so the first-order condition is:

$$\frac{1}{R} u' \left(W_{-1} - \frac{1}{R} W_0 \right) = K[F(\tilde{e}(W_0)) u'(W_0 + \Delta) + (1 - F(\tilde{e}(W_0))) u'(W_0)] \tag{A.5}$$

Neither of these two first-order conditions is well-behaved, and we have not found simple conditions on f that would make them well-behaved. If f and f' are ‘small enough’ (though it is hard to give meaning to this phrase), the problem is well-behaved.²⁵ For example, a uniform distribution with a large enough support will do. This is certainly a sufficient condition, albeit not necessary.

Having said that, we assume that unique solutions to the FOCs exist, in which

²⁴A commitment device conditional on the realized e is not very realistic, but as a comparison it is useful for highlighting the tradeoffs self -1 faces. If self -1 could only commit to a specific decision (one not conditional on e), she would never commit to retirement, and to work only if that is not too costly on the high- e end.

²⁵What we would like is for the right-hand sides of Eqs. (A.3) and (A.5) to be decreasing in W_0 . Then we would have unique solutions to the first-order conditions, which would be global maxima. Notice that the derivative of the right-hand side of (A.3) is of the form:

$$\begin{aligned} K[F(\bar{e}(W_0)) u''(W_0 + \Delta) + (1 - F(\bar{e}(W_0))) u''(W_0)] + f(\bar{e}(W_0)) [Z] \\ + f'(\bar{e}(W_0)) \bar{e}''(W_0) [Ku(W_0 + \Delta) - \beta \delta \bar{e}(W_0) - Ku(W_0)] \end{aligned} \tag{A.7}$$

where the expression Z multiplied by $f(\bar{e}(W_0))$ is complex and not worth writing down for our purposes. The derivative of the right-hand side of Eq. (A.5) is very similar, the difference being that $\bar{e}(W_0)$ is replaced by $\tilde{e}(W_0)$ and there is no term multiplied by f' (The term multiplied by $f(\bar{e}(W_0))$ is also simpler).

case they define the maximum. We are interested in the difference of the right-hand-sides of the first-order conditions (A.3) and (A.5):²⁶

$$\begin{aligned}
 & \overbrace{K[F(\tilde{e}(W_0)) - F(\bar{e}(W_0))][u'(W_0) - u'(W_0 + \Delta)]}^{\text{higher saving}} \\
 & + f(\bar{e}(W_0)) \bar{e}'(W_0) \underbrace{[Ku(W_0 + \Delta) - \beta\delta\bar{e}(W_0) - Ku(W_0)]}_{\text{lower saving}}
 \end{aligned} \tag{A.6}$$

Notice that since $\tilde{e}(W_0) > \bar{e}(W_0)$ for any W_0 , the overbraced product is positive, so it indeed encourages higher saving. On the other hand, we know that for $\bar{e}(W_0)$, self 0 is indifferent between working and not working, and also that in that case self -1 would prefer her to work. Thus the underbraced term is positive. But $\bar{e}'(W_0)$ is negative, so the given effect in fact tends to lower savings.

The intuition behind these two effects is straight-forward enough. First, since there is a chance that self 0 will retire when self -1 prefers that she work, she'll need more money than if she worked. Thus, self -1 saves more. Second, since saving less induces work in some additional states, self -1 has an incentive to save less.

It should be clear that these are just translations of the cases analyzed in the certainty model into the uncertainty setting. This setup, in addition, also allows for convenient analysis of when higher or lower saving is likely to occur. For example, if $f(\bar{e}(W_0^c))$ (where W_0^c is optimal savings with commitment) is high compared to $F(\tilde{e}(W_0^c)) - F(\bar{e}(W_0^c))$, we will get lower saving. That is, if self -1 feels that she can exert a lot of influence on self 0's decision through savings, she will save less. On the other hand, if $f(\bar{e}(W_0^c))$ is close to zero, while $F(\tilde{e}(W_0^c)) - F(\bar{e}(W_0^c))$ is fairly large, there will be higher saving. In simpler terms, if self -1 can't exert much influence on self 0, she will just accept that self 0 might retire too early, and give the now poorer self more savings.²⁷

Appendix B. Existence and uniqueness of equilibria

In this section, we outline a proof of the existence of equilibrium for the long-horizon game. It just requires pulling together much of what we have already shown.

²⁶If the difference is positive at the optimal savings with commitment, then the optimal savings without commitment is higher. This is trivial if the problem is well-behaved in the above sense. But the assumption that the first-order condition has a unique solution, together with the observation that for low W_0 the right-hand side of Eq. (A.3) is greater than the left-hand-side, and vice versa if W_0 is close to RW_{-1} , is also sufficient. Similarly, the opposite is the case if the difference is negative.

²⁷Notice that making the size assumptions on f and f' does not make the comparison of the two effects an irrelevant exercise. Though f and f' are small (compared to 1), there is no restriction on their relative size, so $f(\bar{e}(W_0^c))$ and $F(\tilde{e}(W_0^c)) - F(\bar{e}(W_0^c))$ might compare in any number of ways.

For the game after retirement, the existence and uniqueness of the subgame-perfect equilibrium has been established by David Laibson. For earlier periods, we prove the following general theorem:

Lemma 3. *A Markov-perfect subgame-perfect equilibrium exists with the following properties. For $t \leq 0$, the domain $(0, \infty)$ of the consumption rule $c_t(W_t)$ can be divided into finitely many disjoint intervals such that in the interior of each interval,*

1. *the eventual period 0 work/retirement decision is the same,*
2. *the equilibrium consumption schedules $c_s(W_s)$ for $s > t$ are all differentiable in W_s , and*
3. *self t has a constant marginal propensity to consume;*
4. *further, at an interval endpoint a , self t is indifferent between following the limit of the two neighboring intervals' consumption rules, and utility is continuous in wealth at a .*

Proof. Starting from $t = 0$, use the following backward induction type of construction for finding the equilibrium: given the next self's strategy, maximize utility for self t . If for some wealth self t is indifferent between a number of consumption levels, assign to her the strategy that the earlier self would prefer.

Of course, we have to prove that this construction works and yields an equilibrium with the above properties. We do this by backward induction.

The case is clear for $t = 0$. Now suppose the statement is true for $t = m + 1$. We will prove it for $t = m$.

Suppose W_m is given. For self $m + 1$, let the intervals in question be divided the by points $0 < a_1 < \dots < a_M$. For any $\epsilon > 0$, self m 's maximization problem has a solution if her savings level is restricted to lie in the interval $[a_i + \epsilon, a_{i+1} - \epsilon]$. Since there are only finitely many intervals, a maximum on the union of these intervals and the points $\{a_i\}$ also exists. It is easy to see that as ϵ approaches zero, eventually the maximum doesn't change. For otherwise there would be a point a_i such that as W_{m+1} approaches a_i from one of the sides, self m 's utility is greater than at savings level a_i , which contradicts that when indifferent, self $m + 1$ chooses the consumption level self m prefers.²⁸

This shows that for each wealth level W_m , self m 's problem has a solution. Now

²⁸More precisely, there is a sequence $W_{m+1,n}$ approaching a_i from one side such that discounted utility for self m is increasing on that sequence, and the limit of the discounted utilities is more than discounted utility at a_i . But if at wealth level a_i self $m + 1$ consumes $\lim c_{m+1}(W_{m+1,n})$, by point 4 the discounted utility of self m should be the limit of the discounted utilities when leaving savings $W_{m+1,n}$. But this is impossible by construction as we have assumed that when indifferent, self $m + 1$ does what self m prefers.

define $0 = b_0 < b_1 < \dots < b_N < \infty$ such that for each $i = 0, \dots, N - 1$, if $W_m \in (b_{2i}, b_{2i+1})$, then $W_{m+1} \in (a_i, a_{i+1})$, and if $W_m \in (b_{2i+1}, b_{2i+2})$, then $W_{m+1} = a_{i+1}$.²⁹

By definition, point 1 is satisfied for each (b_j, b_{j+1}) . It is also clear that for any (b_{2i+1}, b_{2i+2}) , points 2 and 3 are satisfied as well. Therefore let us concentrate on the case $W_m \in (b_{2i}, b_{2i+1})$. Since all future consumption schedules are differentiable at $W_{m+1}(W_m)$, the discounted utility of self m as a function of c_m is differentiable at $c_m(W_m)$. Now $c_m(W_m)$ maximizes this utility, so the derivative at that point is zero. Taking the derivative for selves m and $m + 1$, as in Laibson (1997a), and substituting leads to the Euler equation:

$$\frac{u'(c_m)}{u'(c_{m+1})} = R\delta(\beta\lambda_{m+1} + 1 - \lambda_{m+1}) \tag{B.1}$$

where λ_{m+1} is self $m + 1$'s marginal propensity to consume. Then self m 's marginal propensity to consume λ_m on (b_{2i+1}, b_{2i+2}) is constant and is given by the equation:

$$\frac{\lambda_m}{1 - \lambda_m} = \frac{R\lambda_{m+1}}{[R\delta(\beta\lambda_{m+1} + 1 - \lambda_{m+1})]^{1/\rho}} \tag{B.2}$$

(This is just Laibson's recursion for the λ s).

Also, clearly, utility is continuous in wealth at each interval endpoint, otherwise the agent would 'jump' to the other interval at a different place. Finally, we need to show the agent is indifferent between the limits of the two neighboring consumption rules. Suppose by contradiction that, say, consuming $\lim_{W_t \searrow a} c_t(W_t)$, doesn't yield the limit of the utilities. This could only be because one of the future selves jumped at an interval endpoint. Then self m 's utility actually increased, because when indifferent future selves do what self m wants them to (with quasi-hyperbolic discounting, all previous selves want the same thing). But in this case near a self m 's choice of consumption wasn't optimal, a contradiction. \square

Since sequential equilibria in finite extensive-form games with perfect information are generically unique (see for example Myerson (1991)), the above equilibrium is essentially unique.

Appendix C. Proofs of some claims

To prove Lemma 1, we need the following preliminary result.

Lemma 4. For $t \leq -1$, savings is monotonically increasing in wealth.

²⁹Of course, some of the intervals (b_j, b_{j+1}) may be empty.

Proof. Suppose by contradiction that $W_t > W'_t$ but that the corresponding savings levels satisfy $W_{t+1} < W'_{t+1}$. Let the consumption levels be c_t and c'_t and denote the continuation utilities from leaving wealth levels W_{t+1} and W'_{t+1} by K and K' , respectively. Furthermore, define $c_t = W_t - (1/R)W'_{t+1}$, $c'''_t = W'_t - (1/R)W_{t+1}$. Then:

$$u(c''_t) + K' \leq u(c_t) + K \tag{C.1}$$

$$u(c'''_t) + K \leq u(c'_t) + K' \tag{C.2}$$

We can add these and eliminate K and K' to get:

$$u(c''_t) + u(c'''_t) \leq u(c_t) + u(c'_t) \tag{C.3}$$

But notice that $c_t > c''_t$, $c'''_t > c'_t$ and $c_t + c'_t = c''_t + c'''_t$. Since u is concave, the inequality (C.3) is impossible. This completes the proof. \square

Lemma 1. *Suppose that $t \leq 0$ and $W_t > W'_t$. Then it is not possible that self t with wealth W_t behaves so that self 0 eventually works, and with wealth W'_t she behaves so that self 0 eventually retires.*

Proof. We prove by backward induction. The statement is clearly true for $t = 0$.

Suppose the statement is true for $t = m + 1$. We will prove by contradiction that it is true for $t = m$. Suppose it isn't. Then there are wealth levels W_m and W'_m such that $W_m > W'_m$ and with wealth W_m self 0 eventually works, and with wealth W'_m self 0 eventually retires. Since our statement is true for $t = m + 1$, we then need to have $W_{m+1} < W'_{m+1}$. But this is impossible by Lemma 4. \square

Lemma 2. *Let \bar{W}_t ($t < 0$) be the level of wealth at which self t is indifferent between behaviors that eventually lead to self 0 working or retiring.³⁰ At this savings level, self $t - 1$ strictly prefers self t to choose to eventually make self 0 retire.*

Proof. We again prove by backward induction, although, as the reader will see, the need for that is little more than technical. Let c_t^w, K_t^w and c_t^r, K_t^r be the consumption levels and continuation utilities for self t with wealth level \bar{W}_t in the working and retirement cases, respectively.

³⁰Though this fact is not necessary here, it should be said that \bar{W}_t exists and is unique. That it exists can be seen from the consideration that both the set of savings levels where eventual early retirement is (weakly) preferred and where eventual late retirement is preferred are closed. This can be proven easily using backward induction. That it is unique follows from a variant of Lemma 1 (the proof of which didn't use strict preferences) along with backward induction.

Suppose first that $t = -1$. We have $c_{-1}^w > c_{-1}^r$, since otherwise self -1 would have to leave \bar{W}_0 for self 0 , which would not make him indifferent between working and retiring. Also

$$u(c_{-1}^r) + \beta\delta K_{-1}^r = u(c_{-1}^w) + \beta\delta K_{-1}^w \tag{C.4}$$

To see what self -2 would want, we have to compare $\beta\delta u(c_{-1}^r) + \beta\delta^2 K_{-1}^r$ and $\beta\delta u(c_{-1}^w) + \beta\delta^2 K_{-1}^w$. This is easy:

$$\begin{aligned} &\beta\delta u(c_{-1}^r) + \beta\delta^2 K_{-1}^r - \beta\delta u(c_{-1}^w) - \beta\delta^2 K_{-1}^w \\ &= \beta\delta(u(c_{-1}^r) - u(c_{-1}^w)) + \beta\delta^2(K_{-1}^r - K_{-1}^w) \\ &= \beta\delta(u(c_{-1}^r) - u(c_{-1}^w)) + \delta(u(c_{-1}^w) - u(c_{-1}^r)) = (\delta - \beta\delta)(u(c_{-1}^w) - u(c_{-1}^r)) > 0 \end{aligned}$$

If the statement is true for $t = m + 1$, then since self $m + 1$ is not indifferent between self $m + 2$ working and retiring at \bar{W}_{m+2} , we have $c_{m+1}^w > c_{m+1}^r$. Then the same proof as above works. \square

Theorem 1. *Suppose the horizon after retirement is long. If self $-n < -1$ plans to work in period 0 , so does self $-n + 1$.*

Proof. With a long horizon, the marginal propensity to consume out of wealth is a constant over time, λ^* . For notational simplicity, redefine W to include the present value of earnings in period 0 . If self $-n$ plans to work, next period’s wealth is $(1 - \lambda^*)RW$. Now:

$$\begin{aligned} &\left[V((1 - \lambda^*)RW) - V\left((1 - \lambda^*)RW - \frac{1}{R^{n-1}}\Delta\right) \right] \delta \\ &\gg \left[V((1 - \lambda^*)RW) - V\left((1 - \lambda^*)RW - \frac{1}{R^{n-1}}\Delta\right) \right] \delta(\lambda^*\beta + 1 - \lambda^*) \end{aligned} \tag{C.5}$$

since $\beta < 1$. Using that $R\delta(\lambda^*\beta + 1 - \lambda^*) = 1/((1 - \lambda^*)/R)^{-\rho}$ ³¹ this equals:

$$\begin{aligned} &\frac{1}{R}\left(\frac{1}{1 - \lambda^*}\frac{1}{R}\right)^{-\rho} \left[V((1 - \lambda^*)RW) - V\left((1 - \lambda^*)RW - \frac{1}{R^{n-1}}\Delta\right) \right] \\ &= \frac{1}{R}\left(\frac{1}{1 - \lambda^*}\frac{1}{R}\right)^{-\rho} \int_0^{\frac{1}{R^{n-1}}\Delta} V'((1 - \lambda^*)RW - x) dx \end{aligned} \tag{C.6}$$

Since V is concave, the above is greater than:

³¹This is an easy consequence of the sophisticates’ first-order condition for consumption levels in adjacent periods (Laibson, 1997a).

$$\frac{1}{R} \left(\frac{1}{1 - \lambda^*} \frac{1}{R} \right)^{-\rho} \int_0^{\frac{1}{R^{n-1}} \Delta} V'((1 - \lambda^*) RW - (1 - \lambda^*)x) dx \tag{C.7}$$

which, since $V(W) = Vu(W)$ for each W , equals:

$$\begin{aligned} \frac{1}{R} \int_0^{\frac{1}{R^{n-1}} \Delta} V' \left(W - \frac{1}{R} x \right) dx &= \int_0^{\frac{1}{R^n} \Delta} V'(W - x) dx \\ &= V(W) - V \left(W - \frac{1}{R^n} \Delta \right) \end{aligned} \tag{C.8}$$

through a change in variables. Since self $- n$ plans to work, this is greater than or equal to $\beta \delta^n e$. But then $V((1 - \lambda^*) RW) - V((1 - \lambda^*) RW - \frac{1}{R^{n-1}} \Delta) \geq \beta \delta^{n-1} e$, implying the claim. \square

The proof of Theorem 1 really only used that:

$$\frac{((1 - \lambda^*) R)^\rho}{R \delta} < 1 \tag{C.9}$$

where λ^* is each self’s marginal propensity to consume. Even for agents naive about consumption decisions, marginal propensity to consume is equal across periods with a value of:

$$\lambda^* = \frac{1 - (\delta R^{1-\rho})^{1/\rho}}{1 - (1 - \beta^{1/\rho})(\delta R^{1-\rho})^{1/\rho}} \tag{C.10}$$

Assuming $\delta R^{1-\rho} < 1$, which is necessary for the naive maximization problem to have a solution, it is easily established that the above satisfies inequality C.9. The proofs of the other claims in the text carry over quite effortlessly as well.

Appendix D. A more hyperbolic discount structure

The only change we make is to introduce an additional discount parameter $\gamma < 1$ into Laibson’s model, which is effective for two periods. Thus, self t ’s discounted utility from consumption is:

$$u(c_t) + \beta \gamma \delta u(c_{t+1}) + \beta \gamma^2 \delta^2 \sum_{i=0}^{\infty} \delta^i u(c_{t+2+i}) \tag{D.1}$$

Of course, we have to start from ground zero and solve the savings equilibrium before we can get into questions concerning retirement. The analysis is similar to Laibson (1997a), and we will only go through an accelerated version of it.

Backwards induction along with a repeated use of property 2 of CRRA utility functions proves that in each period, consumption is a linear (and thus differentiable) function of wealth. Then in equilibrium self t will choose c_t to satisfy:

$$u'(c_t) = \beta\gamma\delta R \frac{\partial c_{t+1}}{\partial W_{t+1}} u'(c_{t+1}) + \beta\gamma^2\delta^2 R^2 \sum_{i=0}^{\infty} R^i \delta^i \frac{\partial c_{t+2+i}}{\partial W_{t+2+i}} \prod_{j=0}^i \left(1 - \frac{\partial c_{t+1+j}}{\partial W_{t+1+j}}\right) u'(c_{t+2+i}) \quad (D.2)$$

The similar equation for period $t + 1$ is:

$$u'(c_{t+1}) = \beta\gamma\delta R \frac{\partial c_{t+2}}{\partial W_{t+2}} u'(c_{t+2}) + \beta\gamma^2\delta^2 R^2 \sum_{i=0}^{\infty} R^i \delta^i \frac{\partial c_{t+3+i}}{\partial W_{t+3+i}} \prod_{j=0}^i \left(1 - \frac{\partial c_{t+2+j}}{\partial W_{t+2+j}}\right) u'(c_{t+3+i}) \quad (D.3)$$

Combining the two we get:

$$u'(c_t) = \beta\gamma\delta R \frac{\partial c_{t+1}}{\partial W_{t+1}} u'(c_{t+1}) + \beta\gamma^2\delta^2 R^2 \frac{\partial c_{t+2}}{\partial W_{t+2}} \left(1 - \frac{\partial c_{t+1}}{\partial W_{t+1}}\right) u'(c_{t+2}) + \delta R \left(1 - \frac{\partial c_{t+1}}{\partial W_{t+1}}\right) \left(u'(c_{t+1}) - \beta\gamma\delta R \frac{\partial c_{t+2}}{\partial W_{t+2}} u'(c_{t+2})\right)$$

Putting this into a more convenient form leads to the following lemma.

Lemma 5. *The Euler equation for the choice of consumption at time t is:*

$$u'(c_t) = \delta R \left(\beta\gamma \frac{\partial c_{t+1}}{\partial W_{t+1}} + \left(1 - \frac{\partial c_{t+1}}{\partial W_{t+1}}\right) \right) u'(c_{t+1}) - \beta\gamma\delta^2 R^2 \frac{\partial c_{t+2}}{\partial W_{t+2}} \left(1 - \frac{\partial c_{t+1}}{\partial W_{t+1}}\right) u'(c_{t+2})(1 - \gamma). \quad (D.4)$$

It is easily seen that for $\gamma = 1$ this reduces to Laibson’s Euler equation.

Using this Euler equation, we can show that in a game with horizon T , the consumption rule is $c_t = \lambda_{T-t} W_t$, where the λ ’s are determined by the recursion:

$$\left(\frac{\lambda_{n+2}}{1 - \lambda_{n+2}}\right)^{-\rho} = \delta R^{1-\rho} (\beta\gamma\lambda_{n+1} + (1 - \lambda_{n+1})) \lambda_{n+1}^{-\rho} - \beta\gamma\delta^2 R^{2(1-\rho)} (1 - \gamma) \lambda_n^{1-\rho} (1 - \lambda_{n+1})^{1-\rho} \quad (D.5)$$

with initial value $\lambda_0 = 1$. Though we haven’t shown that this converges, it seems to do so: in computer simulations it converged for all values of the parameters that

we have tried.³² The existence of a constant marginal propensity to consume far from the end is not technically necessary for what we are going to do, but it is nice to work off a benchmark that has smooth consumption. We will therefore assume that for our parameter values the long-horizon case has a constant marginal propensity to consume of λ^* .³³

As before, we introduce variously discounted value functions. We will need three this time:

$$\begin{aligned}
 V(W) &= u(\lambda^*W) + \beta\gamma\delta u(\lambda^*(1 - \lambda^*)RW) + \beta\gamma^2\delta^2 \sum_{i=2}^{\infty} \delta^{i-2} u(\lambda^*R^i(1 - \lambda^*)^i W) \\
 Z(W) &= u(\lambda^*W) + \gamma\delta \sum_{i=1}^{\infty} \delta^{i-1} u(\lambda^*R^i(1 - \lambda^*)^i W) \\
 D(W) &= \sum_{i=0}^{\infty} \delta^i u(\lambda^*R^i(1 - \lambda^*)^i W) \tag{D.6}
 \end{aligned}$$

It is easy to see that one period before retirement we get the same undersaving possibility as with quasi-hyperbolic discounting. In that case \bar{W}_{-1} is defined by:

$$u\left(\bar{W}_{-1} - \frac{1}{R} \bar{W}_0\right) + \beta\gamma\delta Z(\bar{W}_0 + \Delta) - \beta\gamma\delta e = V(\bar{W}_{-1}) \tag{D.7}$$

To see what self -2 wants self -1 to do at this wealth level we want to look at the difference:

$$\beta\gamma\delta u\left(\bar{W}_{-1} - \frac{1}{R} \bar{W}_0\right) + \beta\gamma^2\delta^2 D(\bar{W}_0 + \Delta) - \beta\gamma^2\delta^2 e - \beta\gamma\delta Z(\bar{W}_{-1}) \tag{D.8}$$

Using that $V(W) = \beta\gamma Z(W) + (1 - \beta\gamma)u(\lambda^*W) + \beta\gamma\delta(1 - \beta\gamma)u(\lambda^*(1 - \lambda^*)RW)$ and $Z(W) = \gamma D(W) + (1 - \gamma)u(\lambda^*W)$, along with Eq. (D.7), the above becomes:

$$\begin{aligned}
 &\beta\gamma\delta u(c_{-1}^w) + \beta\gamma^2\delta^2 D(\bar{W}_0 + \Delta) - \beta\gamma^2\delta^2 e - \beta\gamma^2\delta^2 D(\bar{W}_0 + \Delta) \\
 &- \beta\gamma\delta^2(1 - \gamma) u(c_0^w) + \beta\gamma\delta^2 e + \delta(1 - \beta\gamma) u(c_{-1}^r) + \beta\gamma\delta^2(1 - \gamma) u(c_0^r) \tag{D.9}
 \end{aligned}$$

where the subscripts on c denote the period in question and the superscripts stand for whether retirement or work is chosen. Dividing by δ and regrouping we get:

$$\begin{aligned}
 &-(1 - \gamma)[(u(c_{-1}^w) + \beta\gamma\delta u(c_0^w) - \beta\gamma\delta e) - (u(c_{-1}^r) + \beta\gamma\delta u(c_0^r))] \\
 &- \gamma(1 - \beta)(u(c_{-1}^w) - u(c_{-1}^r)) \tag{D.10}
 \end{aligned}$$

Using that self -1 is indifferent between working and retiring:

³²It must be said, though, that we haven't tried very many values.

³³In this case, we also get the familiar undersaving outcome.

$$(1 - \gamma)[\beta\gamma^2\delta^2D((1 - \lambda^*)R(\bar{W}_0 + \Delta)) - \beta\gamma^2\delta^2D((1 - \lambda^*)^2R^2\bar{W}_{-1})] - \gamma(1 - \beta)(u(c_{-1}^w) - u(c_{-1}^r)) \tag{D.11}$$

Dividing by γ , we finally get that the difference (D.8) has the same sign as:

$$\beta\gamma\delta^2(1 - \gamma)\overbrace{[D((1 - \lambda^*)R(\bar{W}_0 + \Delta)) - D((1 - \lambda^*)^2R^2\bar{W}_{-1})]}^{\text{I}} - (1 - \beta)\overbrace{(u(c_{-1}^w) - u(c_{-1}^r))}^{\text{II}} \tag{D.12}$$

The second term in this sum is always negative (II is positive), while the other one can be either positive or negative, though it seems it is more often positive (for that we only need $\bar{W}_0 + \Delta > R(1 - \lambda^*)\bar{W}_{-1}$). For $\gamma = 1$, the first term drops out, so the expression is negative, which means that self -2 would want self -1 to retire at this wealth level. This is just what we had before. On the other hand, with $\gamma \neq 1$ and no degeneracy, the first term can be positive, so we do not necessarily get a negative sum. In particular, if the first term is positive and $\beta = 1$, we can only get lower saving (that is, self -2 wants self -1 to work at \bar{W}_{-1}).

In general, both for γ and $1 - \gamma$ close to 0 (both relative to $1 - \beta$), we will get higher saving. This will be clear intuitively as soon as we understand that Eq. (D.12) contrasts two conflicts between selves -1 and -2. First, from the point of view of self -2, self -1 discounts too much between periods -1 and 0, as we had before (term II). But also, self -1 discounts too much between periods 0 and 1, that is, she doesn't appreciate the extra consumption from working as much as she should (term I). For β close to 1, the first effect is negligible. For γ close to 0 or 1, the second one is: close to 1 because then the conflict is small, and close to 0 because then the effect is 'too far in the future' (it is very discounted).

This is only a simple extension of the quasi-hyperbolic setup, but it still indicates that lower saving is more likely with hyperbolic discounting. It also captures what appears to be the two most important conflicts between selves -1 and -2 regarding retirement: that from the perspective of self -2, self -1 overweights consumption in period -1 but underweights consumption after period 0 relative to effort in period 0. Their conflicts about consumption in periods after period 0 are likely to be unimportant. Of course, for earlier selves, this discount structure might not be sufficient: it would be interesting to see better approximations. It won't be easy: genuine hyperbolic discount functions generate equilibria that are extremely hard to analyze.

References

Ainslie, G., 1992. *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*. Cambridge University Press, Cambridge, UK.

- Ainslie, G., Haslam, N., 1992. Hyperbolic discounting. In: Loewenstein, G., Elster, J. (Eds.), *Choice Over Time*. Russell Sage Foundation, New York, p. 59.
- Bernheim, D.B., 1994. Do households appreciate their financial vulnerabilities? An analysis of actions, perceptions, and public policy. Mimeo. Princeton University.
- Bernheim, D.B., Skinner, J., Weinberg, S., 2001. What accounts for the variation in retirement wealth among US. households? *American Economic Review* 91 (4), 832–857.
- Laibson, D., 1997a. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* 112, 443–477.
- Laibson, D., 1997b. Hyperbolic discounting and time preference heterogeneity. Mimeo. Harvard University.
- Mischel, W., Staub, E., 1965. Effects of expectancy on working and waiting for longer rewards. *Journal of Personality and Social Psychology* 1, 625–633.
- Myerson, R.B., 1991. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, MA.
- O'Donoghue, T., Rabin, M., 1999. Doing it now or later. *American Economic Review* 89 (1), 103–124.