Commitment vs. Flexibility*

Manuel Amador†
Iván Werning‡
George-Marios Angeletos§

November 17, 2003

Abstract

This paper studies the optimal trade-off between commitment and flexibility in an intertemporal consumption/savings choice model. Individuals expect to receive relevant information regarding their own situation and tastes - generating a value for flexibility - but also expect to suffer from temptations - generating a value for commitment. The model combines the representations of preferences for flexibility introduced by Kreps (1979) with its recent antithesis for commitment proposed by Gul and Pesendorfer (2002), which nests the hyperbolic discounting model. We set up and solve a mechanism design problem that optimizes over the set of consumption/saving options available to the individual each period. We characterize the conditions under which the solution takes a simple threshold form where minimum savings policies are optimal. Our analysis is also relevant for other issues such as situations with externalities or the problem faced by a "paternalistic" planner, which may be important for thinking about some regulations such as forced minimum schooling laws.

Introduction

If people suffer from temptation and self-control problems, what should be done to help them? Most analysis lead to a simple and extreme conclusion: it is optimal to take over the individual's choices completely. For example, in models with hyper-

*We'd like to thank comments and suggestions from Daron Acemoglu, Andy Atkeson, Peter Diamond and especially Pablo Werning.

†Stanford GSB
‡MIT, NBER and UTDT
§MIT and NBER

bolic discounting preferences it is desirable to impose a particular savings plan on individuals.

Indeed, one commonly articulated justification for government involvement in retirement income in modern economies is the belief that an important fraction of the population would save "inadequately" if left to their own devices (e.g. Diamond, 1977). From the workers perspective most pension systems, pay-as-you-go and capitalized systems alike, effectively impose a minimum saving requirement. One purpose of this paper is to see if such minimum saving policies are optimal in a model where agents suffer from the temptation to "over-consume".

In a series of recent papers Gul and Pesendorfer (2001, 2002a,b) have given preferences that value commitment an axiomatic foundation and derived a useful representation theorem. In their representation the individual suffers from temptations and may exert costly self-control. This formalizes the notion that commitment is useful as a way to avoid temptations that either adversely affect choices or require exerting costly self-control. On the opposite side of the spectrum, Kreps (1979) provided an axiomatic foundation for preferences for flexibility. His representation theorem shows that they can be represented by including taste shocks into an expected utility framework.

Our model combines Kreps' with Gul and Pesendorfer's representations. Our main application modifies the intertemporal taste-shock preference specification introduced by Atkeson and Lucas (1995) to incorporate temptation. In their model the individual has preferences over random consumption streams. Each period an i.i.d. taste shock is realized that affects the individual's desire for current consumption. Importantly, the taste shock at time-t is assumed to be private information. We modify these preferences by assuming that agents suffer from the temptation for higher present consumption. This feature generates a desire for commitment.

The informational asymmetry introduces a trade-off between commitment and flexibility. Commitment is valued because it reduces temptation while flexibility is valued because it allows the use of the valuable private information. We solve for the optimal incentive compatible allocation that trades-off commitment and flexibility. One can interpret our solution as describing the optimal commitment device.

In addition to Gul and Pesendorfer's framework, models with time-inconsistent preferences, as in Strotz (1956), also generate a value for committment. In particular,

the hyperbolic discounting model has proven useful for studying the effects of a temptation to 'over-consume' as well as the desirability of committment devices (Phelps and Pollack, 1968, and Laibson, 1994). As Krusell, Kuruscu and Smith (2002) have pointed out, however, the temptation framework provided by Gul and Pesendorfer effectively generalizes the hyperbolic discounting model: it results in the limiting case when the agent cannot exert any self-control, giving in fully to his temptations. For expositional purposes we first treat the hyperbolic discounting case in detail and then show that the results extend to Gul and Pesendorfer's framework.

We begin by considering a simple hyperbolic discounting case with two possible taste shocks. By solving this case, we illustrate how the optimal allocation depends critically on the strength of the temptation for current consumption relative to the dispersion of the taste shocks. For the resulting second-best problem there are two important cases to consider.

For low levels of temptation, relative to the dispersion of the taste shocks, it is optimal to separate the high and low taste shock agents. If the temptation is not too low, then in order to separate them the principal must offer consumption bundles that yield somewhat to the agent's temptation for higher current consumption. Thus, both bundles provide more present consumption than their counterparts in the first best allocation. When temptation is strong enough, separating the agents becomes too onerous. The principal then finds it optimal to bunch both agents: she offers a single consumption bundle equal to her optimal uncontingent allocation. This solution resolves the average over-consumption issue at the expense of foregoing flexibility.

In this way, the optimal amount of flexibility depends negatively on the strength of the temptation relative to the dispersion of the taste shocks. These results with two shocks are simple and intuitive. Unfortunately, with more than two shocks, these results are not easily generalized. We show that with three shocks there are robust examples where 'money burning' is optimal: it is optimal to have one of the agents consuming in the interior of his budget set. Moreover, bunching can occur between any pair of agents. The examples present a wealth of possibilities with no obvious discernible pattern.

Fortunately, strong results are obtained in the case with a continuum of taste shocks. Our main result is a condition on the distribution of taste shocks that is necessary and sufficient for the optimal mechanism to be a simple threshold rule: a

minimum savings level is imposed, with full flexibility allowed above this minimum. The optimal minimum savings level depends positively on the strength of temptation. Thus, the main insight from the two type case carries over here: flexibility falls with the strength of temptation and this is accomplished by increased bunching.

We extend the model to include heterogeneity in temptation of current consumption. This is important because it is reasonable to assume that agents suffer from temptation at varying degrees. Indeed, perhaps some agents do not suffer from temptation at all. Allowing for heterogeneity in temptation would imply that those individuals that we observe saving less are more likely to be the ones suffering from higher temptation. However, we show that the main result regarding the optimality of a minimum saving policy is robust to the introduction of this heterogeneity.

The rest of the paper is organized as follows. In the remainder of the introduction we briefly discuss the related literature. Section 1 lays out the basic intertemporal model using the hyperbolic discounting model. Section 2 analyzes this model with two and three taste shocks while Section 3 works with a continuum of shocks. Section 4 extends the analysis to arbitrary finite time horizons and Section 5 extends the results to the case where agents are heterogenous with respect to their temptation. Section 6 contains the more general case with temptation and self-control proposed by Gul and Pesendorfer (2001,2002a,b). Section 7 studies the case where agents discount exponentially at a different rate than a 'social planner' and preferences are logarithmic. Section 8 diverges to discuss some alternative interpretations and applications of our main results regarding the optimal trade-off between committment and flexibility. The final Section concludes. An appendix collects some proofs.

Related Literature

At least since Ramsey's (1928) moral appeal economists have long been interested in the implications of, and justifications for, socially discounting the future at lower rates than individuals. Recently, Caplin and Leahy (2001) discuss a motivation for a welfare criterion that discounts the future at a lower rate than individuals. Phelan (2002) provides another motivation and studies implications for long-run inequality of opportunity of a zero social discount rate. In both these papers the social planner and agents discount the future exponentially.

Some papers on social security policies have attempted to take into account the possible "undersaving" by individuals. Diamond (1977) discussed the case where agents may undersave due to mistakes. Feldstein (1985) models OLG agents that discount the future at a higher rate than the social planner and studies the optimal pay-as-you-go system. Laibson (1998) discusses public policies that avoid undersaving in hyperbolic discounting models. Imrohoroglu, Imrohoroglu and Joines (2000) use a model with hyperbolic discounting preferences to perform a quantitative exercise on the welfare effects of pay-as-you-go social security systems. Diamond and Koszegi (2002) use a model with hyperbolic discounting agents to study the policy effects of endogenous retirement choices. O'Donahue and Rabin (2003) advocate studying paternalism normatively by modelling the errors or biases agents may have and applying standard public finance analysis.

Finally, several papers discuss trade-offs similar to those emphasized here in various contexts not related to the intertemporal consumption/saving problem that is our focus. Since Weitzman's (1974) provocative paper there has been great interest in the efficiency of the price system compared to a command economy, see Holmstrom (1984) and the references therein. In a recent paper, Athey, Atkeson and Kehoe (2003) study a problem of optimal monetary policy that also features a trade-off between time-consistency and discretion. Sheshinski (2002) models heterogenous agents that make choices over a discrete set of alternatives but are subject to random errors and shows that in such a setting reducing the set of alternatives may be optimal. Laibson (1994, Chapter 3) considers a moral-hazard model with a hyperbolic-discounting agent and shows that the planner may reward the agent for high output by tilting consumption towards the present.

1 The Basic Model

For reasons of exposition we first study a consumer whose preferences are time-inconsistent. Following Strotz (1956), Phelps and Pollack (1968), Laibson (1994) and many others we model the agent in each period as different *selves* and solve for subgame perfect equilibria of the game played between *selves*. In section 6 we show that all our results go through when we use the more general framework provided by Gul and Pesendorfer (2001,2002a,b) which, in addition, does not require an

intrapersonal game interpretation.

Consider first the case with two periods of consumption, $t = 1, 2$, and an initial period $t = 0$ from which we evaluate expected utility. Section 4 extends the analysis to arbitrary finite horizons. Each period agents receive an i.i.d. taste shock θ, normalized so that $E\theta = 1$ which affects the marginal utility of current consumption: higher θ make current consumption more valuable. The taste shock is observed privately by the agent at time t.[1] We think of the taste shock as a catch-all for the significant variation one actually observes in consumption and saving data after conditioning on the available observable variables. We denote first and second period consumption by c and k, respectively.

The utility for *self-1* from periods $t = 1, 2$ with taste shock θ is

$$\theta U(c) + \beta W(k).$$

where $U(\cdot)$ and $W(\cdot)$ are increasing, concave and continuously differentiable[2] and $\beta \leq 1$. The notation allows $W(\cdot) \neq U(\cdot)$, this generality facilitates the extension to N periods in section 4.

The utility for *self-0* from periods $t = 1, 2$ is

$$\theta U(c) + W(k).$$

Agents have quasi-geometric discounting: *self-t* discounts the entire future at rate $\beta \leq 1$ and in this respect, there is *disagreement* among the different *t-selves* and $1-\beta$ is a measure of this disagreement or bias. On the other hand, there is *agreement* regarding taste shocks: everyone values the effect of θ in the same way. Below we often associate the value of β to the strength of a 'temptation' for current consumption; thus, we say that temptation is stronger if β is lower.

An alternative interpretation to 'hyperbolic' discounting is available if we consider only periods 1 and 2. One can simply work with the assumption that the correct welfare criterion does not discount future utility at the same rate as agents do, although both do so exponentially. Although this alternative interpretation is available for

[1] With exponential CARA utility income shocks are equivalent to taste shocks.

[2] Note that a taste shock for period $t = 2$ is not included in this expression. However, its absence is only apparent since k cannot depend on θ_2 and $E\theta_2 = 1$.

two-periods we will see that in general it does not permit a straightforward extension of the analysis to more periods. In section 7 we discuss a case in which it does generalize.

We investigate the optimal allocation from the point of view of *self-0* subject to the constraint that θ is private information of *self-1*. The essential tension is between tailoring consumption to the taste shock and the *self-1*'s constant higher desire for current consumption. This generates a trade-off between commitment and flexibility from the point of view of *self-0*.

To solve the allocation preferred by *self-0* with total income y we now set up the optimal direct truth telling mechanism given y.

Two Periods

$$v_2(y) \equiv \max_{c(\theta),k(\theta)} \int \left[\theta U(c(\theta)) + W(k(\theta))\right] dF(\theta)$$

$$\theta U(c(\theta)) + \beta W(k(\theta)) \geq \theta U(c(\theta')) + \beta W(k(\theta')) \text{ for all } \theta, \theta' \in \Theta \qquad (1)$$

$$c(\theta) + k(\theta) \leq y \text{ for all } \theta \in \Theta$$

where $F(\theta)$ is the distribution of the taste shocks with support Θ.

This problem maximizes, given total resources y, the expected utility from the point of view of *self-0* (henceforth: the principal) subject to the constraint that θ is private information of *self-1* (henceforth: the agent). The incentive compatibility constraint (1) ensures that it is in agent-θ's self interest to report truthfully, thus obtaining the allocation that is intended for him. In the budget constraints the interest rate is normalized to zero for simplicity.

The problem above imposes a budget constraint for each $\theta \in \Theta$, so that insurance across θ-agent's is ruled out. The planner cannot transfer resources across different agent's types. This choice was motivated by several considerations.

First, it may be possible to argue that the case without insurance is of direct relevance in many situations. This could be the case if pooling risk is simply not possible or if insurance contracts are not available because of other considerations outside the scope of our model.

Second, the cardinality of the taste shocks plays a more important role in an analysis with insurance. The taste shock θ definitely affects ordinal preferences between

current and future consumption, c and k. However, we would like to avoid taking a strong stand on whether or not agents with high taste for current consumption also have a higher marginal utility from total resources as the expression $\theta u + w$ implictly assumes. Focusing on the case without insurance avoids making our analysis depend strongly on such cardinality assumptions.

Third, without temptation ($\beta = 1$) incentive constrained insurance problems such as Mirrlees (1971) or Atkeson and Lucas (1995) are non-trival and the resulting optimal allocations are not easily characterized. This would make a comparison with the solutions with temptation ($\beta < 1$) more difficult. In contrast, without insurance the optimal allocation without temptation ($\beta = 1$) is straightforward – every agent chooses their tangency point on the budget set – allowing a clearer disentangling of the effects of introducing temptation.

Finally, we hope that studying the case without insurance may yield insights into the case with insurance which we are currently pursuing.

Once the problem above is solved the optimal allocation for *self-0* solves a standard problem:

$$\max_{c_0} \left\{ \theta_0 U\left(c_0\right) + \beta v_2\left(y_0 - c_0\right) \right\}$$

where y_0, c_0 and θ_0 represents the initial $t = 0$, income, consumption and taste shock, respectively. In what follows we ignore the initial consumption problem and focus on non-trivial periods.

2 Two Types

In this section we study the optimal commitment with only two taste shocks, $\theta_h > \theta_l$, occuring with probabilities p and $1 - p$, respectively.

Without temptation, $\beta = 1$, there is no disagreement between the planner and the agent and we can implement the ex-ante *first-best allocation* defined by the solution to $\theta U'\left(c_{fb}\left(\theta\right)\right) / W'\left(k_{fb}\left(\theta\right)\right) = 1$ and $c_{fb}\left(\theta\right) + k_{fb}\left(\theta\right) = y$. For low enough levels of temptation, so that β is close enough to 1, the first-best allocation is still incentive compatible. Intuitively, if the disagreement in preferences is small relative to the dispersion of taste shocks then, at the first best, the low shock agent would not envy the high shock agent's allocation.

Proposition 1 *There exists a $\beta^* < 1$ such that for $\beta \in [\beta^*, 1]$ the first-best allocation is implementable.*

Proof. At $\beta = 1$ the incentive constraints are slack at the ex-ante first-best allocation. Define $\beta^* < 1$ to be the value of β for which the incentive constraint of θ_l holds with equality at the first best allocation. The result follows. ■

This result relies on the discrete difference in taste shocks and no longer holds when we study a continuum of shocks in Section 3.

For higher levels of temptation, *i.e.* lower β, the first best allocation is not incentive compatible. If offered, agent-θ_l would take the bundle meant for agent-θ_h to obtain a higher level of current consumption. The next proposition characterizes optimal allocations in such cases.

Proposition 2 *The optimum can always be attained with the budget constraint holding with equality: $c^*(\theta) + k^*(\theta) = y$ for $\theta = \theta_h, \theta_l$. We have that $\theta_l/\theta_h < \beta^*$ and:*
(a) if $\beta > \theta_l/\theta_h$ separation is optimal, i.e. $c^*(\theta_h) > c^*(\theta_l)$ and $k(\theta_h) < k(\theta_l)$
(b) if $\beta < \theta_l/\theta_h$ bunching is optimal, i.e. $c(\theta_l) = c(\theta_h)$ and $k(\theta_l) = k(\theta_h)$
(c) if $\beta = \theta_l/\theta_h$ separating and bunching are optimal

Proof. First, $\beta^* > \underline{\beta}$ follows since

$$\beta^* \equiv \theta_l \frac{U(c^*(\theta_h)) - U(c^*(\theta_l))}{W(y - c^*(\theta_l)) - W(y - c^*(\theta_h))}$$

$$> \theta_l \frac{U'(c(\theta_h))(c(\theta_h) - c(\theta_l))}{W'(y - c(\theta_h))(c(\theta_h) - c(\theta_l))} = \theta_l \frac{U'(c(\theta_h))}{W'(y - c(\theta_h))} = \frac{\theta_l}{\theta_h} \equiv \underline{\beta}$$

where c^* is the first best allocation.

Now, consider the case where $\beta > \underline{\beta}$ and suppose that $c(\theta_h) + k(\theta_h) < y$. Then an increase in $c(\theta_h)$ and a decrease in $k(\theta_h)$ that holds $(\theta_l/\beta) U(c(\theta_h)) + U(k(\theta_h))$ unchanged increases $c(\theta_h) + k(\theta_h)$ and the objective function. Such a change is incentive compatible because it strictly relaxes the incentive compatibility constraint of the high type pretending to be a low type and leaves the other incentive compatibility constraint unchanged. It follows that we must have $c(\theta_h) + k(\theta_h) = y$ at an optimum.

This also shows that separating is optimal in this case, proving part (a). Analogous arguments establish parts (b) and (c). ■

Proposition 2 shows that for $\beta < \beta^*$ the resulting non-trivial second-best problem can be separated into essentially two cases. For intermediate levels of temptation, i.e. $\theta_l/\theta_h < \beta$, it is optimal to separate the agents. In order to separate them the principal must offer consumption bundles that yield somewhat to the agent's ex-post desire for higher consumption giving them higher consumption in the first period than the first best.

For higher levels of temptation, i.e. $\beta < \theta_l/\theta_h$, separating the agents is too onerous. bunching them is then optimal at the best uncontingent allocation – with $U(\cdot) = W(\cdot)$ this implies $c(\cdot) = k(\cdot) = y/2$. bunching resolves the disagreement problem at the expense of flexibility. In this way, the optimal amount of flexibility depends negatively on the size of the disagreement relative to the dispersion of the taste shocks as measured by θ_l/θ_h.

Proposition 2 also shows that it is always optimal to consume all the resources $c(\theta) + k(\theta) = y$. In this sense, 'money burning', i.e. setting $c(\theta_h) + k(\theta_h) < y$, is not required for optimality. As discuss below, with more than two types this is not a foregone conclusion.

Figure 1 below shows a typical case that illustrate these results. We set $U(c) = c^{1-\sigma}/(1-\sigma)$, $U(\cdot) = W(\cdot)$, and $\sigma = 2$, $\theta_h = 1.2$, $\theta_l = .8$, $p = 1/2$ and $y = 1$. The figure shows consumption in the first period, $c(\theta)$, as a function of β. For comparison we also plots the optimal ex-post consumption for both types (i.e. the full flexibility outcome). Note that these are always higher than the optimal allocation: the principal does manage to lower consumption in the first period.

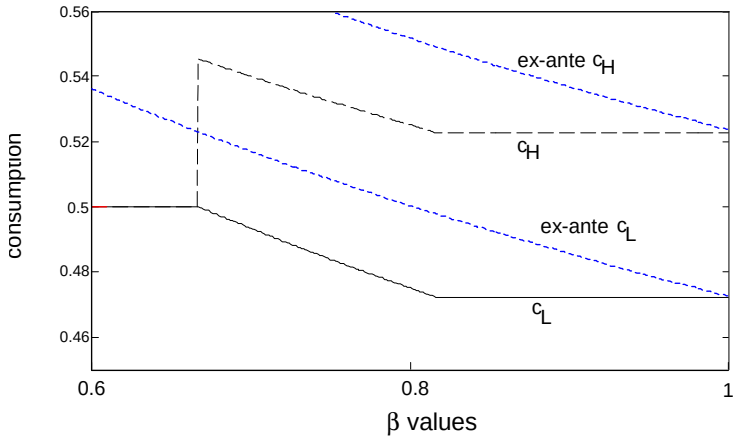

Figure 1: Optimal first period consumption (c) with two shocks as a function of β.

The figure illustrates Proposition 1 and 2 in the following way. For high β the first best allocation is attainable so the optimal allocation does not vary with β in this range. For intermediate β consumption in the first period rises as β falls. In this way the principal yields to the agent's desire for higher consumption. For low enough β bunching becomes optimal and $c(\theta) = y/2$.

To summarize, with two types we are able to characterize the optimal allocation which enjoys nice properties. In particular, the budget constraint holds with equality and we found simple necessary and sufficient conditions for a bunching or separating outcome to be optimal.

Unfortunately, with more than two types extending these conclusions is not straightforward. For example, with three taste shocks, $\theta_h > \theta_m > \theta_l$, it is simple to construct robust examples where the optimal solution has the following properties: (i) the budget constraint for agent θ_m is satisfied with strict inequality – i.e. 'money burning' is optimal; (ii) although $\beta < \theta_m/\theta_h$ remains a sufficient condition for bunching m and h, it is no longer necessary: there are cases with $\beta > \theta_m/\theta_h$ where bunching θ_m and θ_h is optimal; (iii) bunching can occur between θ_l and θ_m, with θ_h is separated. The examples seem to show a variety of possibilities that illustrate the difficulties in characterizing the optimum with more than two types.

Fortunately, with a continuum of types more progress can be made. In the next section we find conditions on the distribution of θ which allows us to characterize the optimal allocation fully.

3 Continuous Distribution of Types

Assume that the distribution of types is represented by a density $f(\theta)$ over the interval $\Theta \equiv [\underline{\theta}, \overline{\theta}]$. We find it convenient to change variables from (c, k) to (u, w) where $u = U(c)$ and $w = W(k)$ and we term either pair an 'allocation'. Let $C(u)$ and $K(w)$ be the inverse functions of $U(c)$ and $W(k)$, respectively, so that $C(\cdot)$ and $K(\cdot)$ are increasing and convex.

To characterize the incentive compatibility constraint (1) in this case consider the problem faced by agent-θ when confronted with a direct mechanism $(u(\theta), w(\theta))$:

$$V(\theta) \equiv \max_{\theta' \in \Theta} \left\{ \frac{\theta}{\beta} u(\theta') + w(\theta') \right\}.$$

If the mechanism is truth telling then $V(\theta) = \frac{\theta}{\beta} u(\theta) + w(\theta)$ and integrating the envelope condition we obtain,

$$\frac{\theta}{\beta} u(\theta) + w(\theta) = \int_{\underline{\theta}}^{\theta} \frac{1}{\beta} u(\tilde{\theta}) d\tilde{\theta} + \frac{\underline{\theta}}{\beta} u(\underline{\theta}) + w(\underline{\theta}) \tag{2}$$

(see Milgrom and Segal, 2002). Incentive compatibility of (u, w) also requires u to be a non-decreasing function of θ. Thus, condition (2) and the monotonicity of u are necessary for incentive compatibility. It is well know that these two conditions are also sufficient (e.g. Fudenberg and Tirole, 1989).

The planner's problem is thus,

$$v_2(y) \equiv \max_{u,w} \int_{\underline{\theta}}^{\overline{\theta}} [\theta u(\theta) + w(\theta)] f(\theta) d\theta,$$

subject to (2), $C(u(\theta)) + K(w(\theta)) \leq y$ and $u(\theta') \geq u(\theta)$ for $\theta' \geq \theta$. This problem is convex since the objective function is linear and the constraint set is convex. In particular, it follows that $v_2(y)$ is concave in y.

We now substitute the incentive compatibility constraint (2) into the objective function and the resource constraint, and integrate the objective function by parts. This allows us to simplify the problem by dropping the function $w(\theta)$, except for its value at $\underline{\theta}$. Consequently, the maximization below requires finding a function $u : \Theta \to \mathbb{R}$ and a scalar $\underline{w}$ representing $w(\underline{\theta})$.

Continuous Distribution of Types

$$v_2(y) \equiv \max_{u(\cdot),\underline{w}} \left\{ \frac{\underline{\theta}}{\beta} u(\underline{\theta}) + \underline{w} + \frac{1}{\beta} \int_{\underline{\theta}}^{\bar{\theta}} \left[(1 - F(\theta)) - \theta (1-\beta) f(\theta) \right] u(\theta) \, d\theta \right\}$$

$$K^{-1}(y - C(u(\theta))) + \frac{\theta}{\beta} u(\theta) - \frac{\underline{\theta}}{\beta} u(\underline{\theta}) - \underline{w} - \int_{\underline{\theta}}^{\theta} \frac{1}{\beta} u(\tilde{\theta}) d\tilde{\theta} \geq 0$$

$$u(\theta') \geq u(\theta) \text{ for } \theta' \geq \theta$$

3.1 Bunching

For any feasible allocation u it is always feasible to modify the allocation so as to bunch some upper tail of agents. That is, the allocation $\hat{u}$ given by $\hat{u}(\theta) = u(\theta)$ for $\theta < \hat{\theta}$ and $\hat{u}(\theta) = u(\hat{\theta})$ is feasible for any $\hat{\theta}$. Thus bunching the upper tail is always feasible, we now show that it is always optimal.

To gain some intuition, note that agents with $\theta \leq \beta\bar{\theta}$ share the ordinal preferences of the planner with a higher taste shock equal to θ/β. That is, the indifference curves $\theta u + \beta w$ and $\theta/\beta u + w$ are equivalent. Informally, these agents can make a case for their preferences. In contrast, agents with $\theta > \beta\bar{\theta}$ display a blatant over-desire for current consumption from the principal's point of view, in the sense that there is no taste shock that would justify these preferences to the planner. Thus, it is intuitive that these agents are bunched since separating them is tantamount to increasing some of these agents consumption, yet they are already obviously "over-consuming".

The next result shows that bunching goes even further than $\beta\bar{\theta}$.

Proposition 3 *Define θ_p as the lowest value in Θ such that for $\hat{\theta} \geq \theta_p$:*

$$\frac{E\left[\theta | \theta \geq \hat{\theta}\right]}{\hat{\theta}} \leq \frac{1}{\beta}$$

Note that $\theta_p \leq \beta\bar{\theta}$ and $\theta_p < \beta\bar{\theta}$ as long as $f > 0$. An optimal allocation u^* has $u^*(\theta) = u^*(\theta_p)$ for $\theta \geq \theta_p$ (i.e. it bunchs all agents above θ_p)

Proof. The contribution to the objective function from $\theta \geq \theta_p$ is

$$\frac{1}{\beta}\int_{\theta_p}^{\bar{\theta}} \left(\left(1-F\left(\theta\right)\right)-\theta\left(1-\beta\right)f\left(\theta\right)\right)u\left(\theta\right)d\theta.$$

Substituting $u = \int_{\theta_p}^{\theta} du + u\left(\theta_p\right)$ and integrating by parts we obtain,

$$u\left(\theta_p\right)\frac{1}{\beta}\int_{\theta_p}^{\bar{\theta}} \left(\left(1-F\left(\theta\right)\right)-\theta\left(1-\beta\right)f\left(\theta\right)\right)d\theta$$
$$+\frac{1}{\beta}\int_{\theta_p}^{\bar{\theta}}\left(\int_{\theta}^{\bar{\theta}}\left(\left(1-F\left(\tilde{\theta}\right)\right)-\tilde{\theta}\left(1-\beta\right)f\left(\tilde{\theta}\right)\right)d\tilde{\theta}\right)du.$$

Note that,

$$\int_{\theta}^{\bar{\theta}}\left(\left(1-F\left(\tilde{\theta}\right)\right)-\tilde{\theta}\left(1-\beta\right)f\left(\tilde{\theta}\right)\right)d\tilde{\theta} = \left(1-F\left(\theta\right)\right)\theta\left(\frac{1}{\beta}-\frac{E\left[\tilde{\theta}|\tilde{\theta}\geq\theta\right]}{\theta}\right)\leq 0,$$

for all $\theta \geq \theta_p$ so it is optimal to set $du = 0$, or equivalently $u\left(\theta\right) = u\left(\theta_p\right)$, for $\theta \geq \theta_p$. ■

With two types Proposition 2 showed that bunching is strictly optimal whenever $\theta_h/\theta_l < 1/\beta$. Proposition 3 generalizes this result since with two types when $\theta_h/\theta_l < 1/\beta$ then according to the definition essentially $\theta_p = \theta_l$.

If the support Θ is unbounded then θ_p may not exist. This occurs, for example, with the Pareto distribution. One can show that in this case it is either optimal to allow full flexibility or bunch all agents depending on the Pareto parameter.

3.2 Assumption A

To solve for the optimal allocation for $\theta \leq \theta_p$ we require the following condition on the density f and β.

A. *The density $f\left(\theta\right)$ is differentiable and satisfies*

$$\theta\frac{f'\left(\theta\right)}{f\left(\theta\right)} \geq -\frac{2-\beta}{1-\beta}$$

for all $\theta \leq \theta_p$.

Assumption A places a negative lower bound on the elasticity of f that is continuous and decreasing in β. The highest lower bound of -2 is attained for $\beta = 0$ and as $\beta \to 1$ the lower bound goes off to $-\infty$. Note that A does not impose the bound on the whole support Θ, only for $\theta \leq \theta_p$.

For any density f such that $\theta f'/f$ is bounded from below assumption A is satisfied for β close enough to 1. Moreover, many densities satisfy assumption A for all β. For example, it is trivially satisfied for all density functions that are non-decreasing and also holds for the exponential distribution, the log-normal, Pareto and Gamma distributions for a large subset of their parameters.

3.3 Minimum Saving Policies

Define $u^*(\theta), w^*(\theta)$ to be the unconstrained optimum for agent-θ:

$$(u^*(\theta), w^*(\theta)) \equiv \arg\max_{\hat{u},\hat{w}} \left\{ \frac{\theta}{\beta}\hat{u} + \hat{w} \right\}$$
$$\text{s.t. } C(\hat{u}) + K(\hat{w}) \leq y$$

Our next result shows that under assumption A agents with $\theta \leq \theta_p$ are offered their unconstrained optimum and agents with $\theta \geq \theta_p$ are bunched at the unconstrained optimum for θ_p. That is, the optimal mechanism offers the whole budget line to the left of some point (c^*, k^*), given by the ex-post unconstrained optimum of the θ_p-agent.

Define the Lagrangian function as:

$$L(\underline{w}, u|\Lambda) \equiv \frac{\underline{\theta}}{\beta}u(\underline{\theta}) + \underline{w} + \frac{1}{\beta}\int_{\underline{\theta}}^{\bar{\theta}} \left[(\beta - 1)\theta f(\theta) + (1 - F(\theta))\right] u(\theta)\, d\theta$$
$$+ \int_{\underline{\theta}}^{\bar{\theta}} \left(K^{-1}(y - C(u(\theta))) + \frac{\theta}{\beta}u(\theta) - \left(\frac{\underline{\theta}}{\beta}u(\underline{\theta}) + \underline{w}\right) - \int_{\underline{\theta}}^{\theta} \frac{1}{\beta}u(\tilde{\theta})d\tilde{\theta} \right) d\Lambda(\theta)$$

where the function Λ is the Lagrange multiplier associated with the incentive compatibility constraint. We require the Lagrange multiplier Λ to be non-decreasing (see Luenberger, 1969, Chapter 8).

Intuitively, the Lagrange multiplier Λ can be thought of as a cumulative distrib-

ution function[3]. If Λ happens to be differentiable with density λ then the continuum of constraints can be incorporated into the Lagrangian as the familiar integral of the product of the left hand side of each constraint and the density function $\lambda(\theta)$. Although this is a common approach in many applications, in general, Λ may have points of discontinuity and these mass points are associated with individual constraints that are particularly important. In such cases, working with a density λ would not be valid. As we shall see, in our case the multiplier Λ is indeed discontinuous at two points: $\underline{\theta}$ and θ_p.

Consider the allocation (u, w) given by $(u(\theta), w(\theta)) = (u^*(\theta), w^*(\theta))$ for $\theta < \theta_p$ and $(u^*(\theta), w^*(\theta)) = (u^*(\theta_p), w^*(\theta_p))$ for $\theta \geq \theta_p$.

Proposition 4 *The allocation (u, w) is optimal if and only if assumption A holds.*

Proof. Without loss of generality set $\Lambda\left(\bar{\theta}\right) = 1$. We will construct Λ to be left continuous. Integrating the Lagrangian by parts:

$$L(\underline{w}, u|\Lambda) \equiv \left(\frac{\underline{\theta}}{\beta}u(\underline{\theta}) + \underline{w}\right)\Lambda(\underline{\theta}) + \frac{1}{\beta}\int_{\underline{\theta}}^{\bar{\theta}}\left((\beta-1)\theta f(\theta) - F(\theta) + \Lambda(\theta)\right)u(\theta)\,d\theta + \int_{\underline{\theta}}^{\bar{\theta}}\left(K^{-1}(y - C(u(\theta))) + \frac{\theta}{\beta}u(\theta)\right)d\Lambda(\theta)$$

Note that the Lagrangian is a sum of integrals of concave functions of $\bar{w}$ and $u(\theta)$. This implies that the Gateaux differential exists and is easily computed (see the Lemma on Gateaux differentiability in the Appendix). In particular, at the proposed

[3]Except for the integrability condition. Also, for notational purposes, we make Λ a left-continuous function, instead of the usual right-continuous convention for distribution functions.

allocation for $\bar{w}, u$ the Gateaux differential is given by:

$$\begin{aligned}\partial L\left(\underline{w}, u; h_{\underline{w}}, h_u|\Lambda\right) = & \left(\frac{\underline{\theta}}{\beta} h_u\left(\underline{\theta}\right) + h_{\underline{w}}\right)\Lambda\left(\underline{\theta}\right) \\ & + \frac{1}{\beta}\int_{\underline{\theta}}^{\bar{\theta}}\left(\left(\beta-1\right)\theta f\left(\theta\right) - F\left(\theta\right) + \Lambda\left(\theta\right) - 1\right) h_u\left(\theta\right) d\theta \\ & + \frac{\theta_p}{\beta}\int_{\underline{\theta}}^{\bar{\theta}}\left(\frac{\theta}{\theta_p} - 1\right)\chi_{\left[\theta_p,\bar{\theta}\right]} h_u d\Lambda\left(\theta\right)\end{aligned} \tag{3}$$

where $\chi_{\left[\theta_p,\bar{\theta}\right]}$ is the indicator function over $\left[\theta_p, \bar{\theta}\right]$, i.e. $\chi_{\left[\theta_p,\bar{\theta}\right]} = 1$ for $\theta \in \left[\theta_p, \bar{\theta}\right]$ and zero otherwise.

The problem is convex, the Lagrangian is differentiable and the proposed allocation is continuous. It follows that $\underline{w}, u$ is optimal if and only if there exists some non-decreasing function Λ such that:

$$\partial L\left(\underline{w}, u; \underline{w}, u|\Lambda\right) = 0 \tag{4}$$

$$\partial L\left(\underline{w}, u; h_{\underline{w}}, h_u|\Lambda\right) \leq 0 \tag{5}$$

for all $h_{\underline{w}}$ and h_u that belongs to the convex cone given by $X = \{w, u : w \in R$ and u is a non-decreasing function $u : \Omega \to R\}$ (see Luenberger, Chapter 8).

Condition (5) requires that $\Lambda\left(\underline{\theta}\right) = 0$. Using this and integrating (3) by parts leads to

$$\partial L\left(\underline{w}, u; h_{\underline{w}}, h_u|\Lambda\right) = \gamma\left(\underline{\theta}\right) h_u\left(\underline{\theta}\right) + \int_{\underline{\theta}}^{\bar{\theta}} \gamma\left(\theta\right) dh_u\left(\theta\right) \tag{6}$$

where,

$$\gamma\left(\theta\right) \equiv \frac{1}{\beta}\int_{\theta}\left(\left(\beta-1\right)\theta f\left(\theta\right) - F\left(\theta\right) + \Lambda\left(\theta\right)\right) d\theta + \frac{\theta_p}{\beta}\int_{\theta}\left(\frac{\theta}{\theta_p} - 1\right)\chi_{\left[\theta_p,\bar{\theta}\right]} d\Lambda$$

It follows that condition (4) requires $\gamma\left(\theta\right) = 0$ for $\theta \in \left[\underline{\theta}, \theta_p\right]$, i.e. where u is strictly increasing. This implies,

$$\Lambda\left(\theta\right) = \left(1-\beta\right)\theta f\left(\theta\right) + F\left(\theta\right), \tag{7}$$

$\theta \in [\underline{\theta}, \theta_p)$. The proposed allocation thus determines a unique candidate multiplier Λ

in the separating region $[\underline{\theta}, \theta_p)$ and assumption A is necessary and sufficient for $\Lambda(\theta)$ to be non-decreasing. It follows that assumption A is necessary for the proposed solution to be optimal.

We now prove sufficiency by showing that there exists a non-decreasing multiplier Λ over the whole range Ω such that the proposed $\bar{w}, u$ satisfies (4) and (5). We've specified Λ for $[\underline{\theta}, \theta_p)$ so we only need to specify the value of Λ for $\left[\theta_p, \bar{\theta}\right]$ and we set $\Lambda(\theta) = 1$ in this interval.

The constructed Λ is not continuous, it has an upward jump at $\underline{\theta}$ and a jump at θ_p. To show that Λ is non-decreasing all that remains is to show that the jump at θ_p is upward,

$$1 - [(1-\beta)\theta_p f(\theta_p) + F(\theta_p)] \geq 0,$$

which follows from the definition of θ_p. To see this, note that if $\theta_p = \underline{\theta}$ the result is immediate since Λ jumps from 0 to 1 at $\underline{\theta}$. Otherwise, recall that θ_p is the lowest $\hat{\theta}$ such that $\gamma(\theta) \leq 0$ for all $\theta \geq \hat{\theta}$, which implies $\gamma'(\theta_p) = (1-\beta)\theta f(\theta) - (1 - F(\theta)) \leq 0$.

The proposed allocation, $\bar{w}$ and u, and the Lagrange multiplier, Λ, imply that $\gamma \leq 0$ and that $\gamma = 0$ wherever u is increasing. Using (6) it follows that (4) and (5) are satisfied. ■

The figure below illustrates the form of the multiplier $\Lambda(\theta)$ constructed in the proof of the proposition.

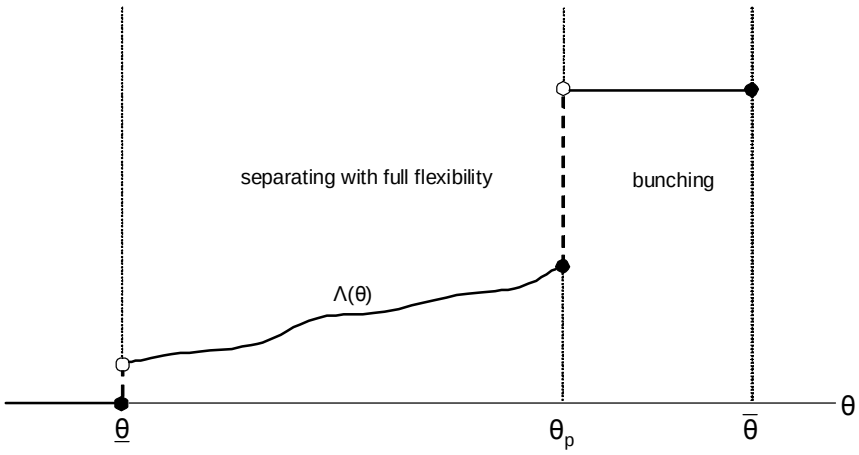

Figure 2: The Lagrange multiplier $\Lambda(\theta)$

Proposition 4 shows that under assumption A the optimal allocation is extremely simple. It can be implemented by imposing a maximum level of current consumption, or equivalently, a minimum level of savings. Such minimum saving policies are a pervasive part of social security systems around the world.

The next result shows the comparative statics of the optimal allocation with respect to temptation β. As the temptation increases, i.e. β decreases, more types are bunched (i.e. θ_p decreases). In terms of policies, as the disagreement increases the minimum savings requirement decreases so there is less flexibility in the allocation.

Proposition 5 *The bunching point θ_p increases with β. The minimum savings requirement, $s_{\min} = y - C\left(u\left(\theta_p\right)\right)$, decreases with β.*

Proof. That θ_p is weakly increasing follows directly from its definition. To see that $s_{\min}$ is decreasing note that $s_{\min}$ solves

$$\frac{\theta_p}{\beta}\frac{U'\left(y - s_{\min}\right)}{W'\left(s_{\min}\right)} = 1,$$

and that θ_p, when interior, solves,

$$\frac{\theta_p}{\beta} = E\left[\theta \,|\, \theta \geq \theta_p\right].$$

Combining these, we obtain $E\left[\theta|\theta \geq \theta_p\right] U'\left(y - s_{\min}\right)/W'\left(s_{\min}\right) = 1$. Since $E\left[\theta \,|\, \theta \geq \theta_p\right]$ is increasing in θ_p the result follows from concavity of U and W. ■

3.4 Drilling

In this subsection we study cases where assumption A does not hold and show that the allocation described in Proposition 4 can be improved upon by drilling holes in the separating section where the condition in assumption A is not satisfied.

Suppose we are offering the unconstrained optimum for a closed interval $[\theta_a, \theta_b]$ of agents and we consider removing the open interval (θ_a, θ_b). Agents that previously found their tangency within the interval will move to one of the two extremes, θ_a or θ_b. The critical issue in evaluating the change in welfare is counting how many agents moving to θ_a versus θ_b. For a small enough interval, welfare rises from those moving to θ_a and falls from those moving to θ_b.

Since the relative measure of agents moving to the right versus the left depends on the slope of the density function this explains its role in assumption A. For example, if $f' > 0$ then upon removing (θ_a, θ_b) more agents would move to the right than the left. As a consequence, such a change is undesirable. The proof of the next result formalizes these ideas.

Proposition 6 *Suppose an allocation* (u, w) *has* $(u(\theta), w(\theta)) = (u^*(\theta), w^*(\theta))$ *for* $\theta \in [\theta_a, \theta_b]$ *with* $\theta_b \leq \theta_p$. *If the condition in assumption A does not hold for* $\theta \in [\theta_a, \theta_b]$ *then removing the interior of the set* $\{(u^*(\theta), w^*(\theta)) \text{ for } \theta \in (\theta_a, \theta_b)\}$ *improves welfare.*

Proof. In the appendix. ■

Propositions 6 illustrates by construction why assumption A is necessary for a simple 'threshold rule' to be optimal and gives some insight into this assumption. Of course, Proposition 6 only identifies particular improvements whenever assumption A fails. We have not characterized the full optimum when assumption A does not hold. It seems likely that 'money burning' may be optimal in some cases.

4 Arbitrary Finite Horizons

We now show that our results extend to arbitrary finite horizons. We confine ourselves to finite horizons because with infinite horizons any mechanism may yield multiple equilibria in the resulting game. These equilibria may involve reputation in the sense that a good equilibrium is sustained by a threat of reverting to a bad equilibria upon a deviation. Some authors have questioned the credibility of such reputational equilibria in intrapersonal games (e.g. Gul and Pesendorfer, 2002a, and Kocherlakota, 1996). We avoid these issues by focusing on finite horizons.

Consider the problem with $N < \infty$ periods $t = 1, ..., N$ where the felicity function is $U(\cdot)$ in each period. Let $\theta^t = (\theta_1, \theta_2, ..., \theta_t)$ denote the history of shocks up to time t. A direct mechanism now requires that at time t the agent make reports on the history of shocks $r^t = (r_1^t, r_2^t, ..., r_t^t)$. The agent's consumption is allowed to depend on the whole history of reports: $c_t(r^t, r^{t-1}, ..., r^1)$. A strategy for self-t is a mapping from the history of shocks and past reports into current reports: $R^t(\theta^t, r^{t-1}, ..., r^1)$. Truth telling requires $R^t(\theta^t, \theta^{t-1}, ..., \theta^1) = \theta^t$ for all t and all histories θ^t.

We first argue that without loss in optimality we can restrict ourselves to mechanisms that at time t require only a report r_t on the current shock θ_t, and not of the whole history of shocks θ^t. This is the case in Atkeson and Lucas (1995) but in their setup since preferences are time-consistent there is a single player and the argument is straightforward.

In contrast, in the hyperbolic model we have N players and the difference in preferences between these selves can be exploited to punish past deviations. For example, an agent at time t that is indifferent between allocations can be asked to choose amongst them according to whether there has been a deviation in the past. In particular, she can 'punish' previous deviating agents by selecting the worst allocations from their point of view. Otherwise, if there have been no past deviations, she can 'reward' the truth-telling agents by selecting the allocation preferred by them. Such schemes may make deviations more costly, relaxing the incentive constraints, and are thus generally desirable.

One way to remove the possibility of these punishment schemes is to introduce the refinement that when agents are indifferent between several allocations choose the one that maximizes the utility of previous selves. Indeed, Gul and Pesendorfer's (2001,2002a,b) framework, discussed in Section 6, delivers, in the limit without self-control, the hyperbolic model with this added refinement.

However, with a continuous distribution for θ such a refinement is not necessary to rule out these punishment schemes. We show that for any mechanism the subset of Θ over which θ-agents are indifferent is at most countable. This implies that the probability that future selves will find themselves indifferent is zero so that the threat of using indifference to punish past deviations has no deterrent effect.

For any set A of pairs (u, w) define the optimal correspondence over $x \in X$

$$M(x; A) \equiv \arg \max_{(u,w) \in A} \{xu + w\}$$

(we allow the possibility that $M(x, A)$ is empty) then we have the following result.

Lemma (Indifference is countable). *For any A the subset $X^I \subset X$ for which $M(x; A)$ has two or more points (set of agents that are indifferent) is at most countable.*

Proof. The correspondence $M(x; A)$ is monotone in the sense that if $x_1 < x_2$ and

$(u_1, w_1) \in M(x_1; A)$ and $(u_2, w_2) \in M(x_2; A)$ then $u_1 \leq u_2$. Thus, in an obvious sense, points at which there are more than a single element in $M(x; A)$ represent upward 'jumps'. As with monotonic functions, it follows easily that $M(x; A)$ can have at most a countable number of such 'jumps'. ■

This result relies only on the single crossing property of preferences and not on the linearity in u and w. We make use of this lemma again in Section 5.

These considerations lead us to write the problem with $N \geq 3$ remaining periods recursively as follows.

N Period Problem

$$v_N(y) = \max_{c,\, y} \int \left[\theta U(c(\theta)) + v_{N-1}(y'(\theta))\right] dF(\theta)$$

$$\theta U(c(\theta)) + \beta v_{N-1}(y'(\theta)) \geq \theta_h U\left(c\left(\hat{\theta}\right)\right) + \beta v_{N-1}\left(y'\left(\hat{\theta}\right)\right) \text{ for all } \theta, \hat{\theta} \in \Theta$$

$$c(\theta) + y'(\theta) \leq y \text{ for all } \theta \in \Theta$$

where $v_2(\cdot)$ was defined in Section 2.

In the above formulation we assume, without loss in optimality, that the optimal mechanism is ex-post optimal given the resources available i.e. v_{N-1} shows up in the objective function and incentive constraint. This is without loss in optimality because any inefficient continuation utility $w'(\theta) < v_{N-1}(y'(\theta))$ can be achieved by 'money burning' with the same effect: setting $w'(\theta) = v_{N-1}(\tilde{y}'(\theta))$ for some $\tilde{y}'(\theta) < y'(\theta)$.

For the simple recursive representation to obtain it is critical that, although the principal and the agent disagree on the amount of discounting between the current and next period, they both agree on the utility obtained from the next period on, given by v_{N-1}. This is not true in the alternative setup where the principal and the agent both discount exponentially but with different discount factors. We treat this case separately in Section 7.

For any horizon N this problem has exactly the same structure as the two-period problem analyzed previously, with the exception that $v_{N-1}(\cdot)$ has substituted $W(\cdot)$. We only required $W(\cdot)$ to be increasing and concave and since $v_{N-1}(\cdot)$ has these properties all the previous results apply. We summarize this result in the next proposition.

Proposition 7 *Under assumption A the optimal allocation with a horizon of N periods can be implemented by imposing a minimum amount of saving $S_t(y_t)$ in period t.*

In Proposition 7 the minimum saving is a function of resources y_t. With CRRA preferences the optimal allocation is linearly homogenous in y, so that $c(\theta, y) = \tilde{c}(\theta) y$ and $y'(\theta, y) = \tilde{y}'(\theta) y$. It follows that the optimal mechanism imposes a minimum saving *rate* for each period that is independent of y_t.

Proposition 8 *Under assumption A and $U(c) = c^{1-\sigma}/(1-\sigma)$ the optimal mechanism for the N-period problem imposes a minimum saving rate s_t for each period t independent of y_t.*

4.1 Hidden Savings

Another property of the optimal allocation identified in Propositions 4 and 7 is worth mentioning. Suppose agents can save, but not borrow, privately behind the planner's back at the same rate of return as the planner, as in Cole and Kocherlakota (2001). The possibility of this 'hidden saving' reduces the set of allocations that are incentive compatible since the agent has a strictly larger set of possible deviations. Importantly, the mechanism described in Proposition 7 continues to implement the same allocation when we allow agents to save privately, and thus remains optimal.

To prove this claim we argue that confronted with the mechanism in Proposition 7 agents that currently have no private savings would never find it optimal to accumulate private savings. To see this, first note that by Proposition 7 the optimal mechanism imposes only a minimum on savings in each period. Thus agent-θ always have the option of saving more observably with the principal than what the allocation recommends, yet by incentive compatibility the agent chooses not to.

Next, note that saving privately on his own can be no better for the agent than increasing the amount of observable savings with the principal. This is true because the principal maximizes the agents utility given the resources at its disposal. Thus, from the point of view of the current self, future wealth accumulated by hidden savings is dominated by wealth accumulated with the principal.

It follows that agents never find it optimal to save privately and the mechanism implements the same allocation when agents can or cannot save privately.

Proposition 9 *Under assumption A the mechanism described in Proposition 7 implements the same allocation when agents can save privately.*

5 Heterogeneous Temptation

Consider now the case where the level of temptation, measured by β, is random. Heterogeneity in β captures the commonly held view that the temptation to overconsume is not uniform in the population and that it is the agents that save the least that are more likely to be 'undersaving' because of a higher temptation to consume (e.g. Diamond, 1977).

If the heterogeneity in temptation were due to permanent differences across individuals then the previous analysis would apply essentially unaltered. If agents knew their β at time time 0 they would truthfully report it so that their mechanism could be tailored to their β as described above[4]. To explore other possibilities we assume the other extreme, that differences in temptation are purely idiosyncratic, so that β is i.i.d. across time and individuals. Thus, each period θ and β are realized together from a continuous distribution – we do not require independence of θ and β for our results. We continue to assume that $\beta \leq 1$ for simplicity.

For any set A of available pairs (u, w) agents with (θ, β) maximize their utility:

$$\arg \max_{(u,w)\in A} \left\{ \frac{\theta}{\beta} u + w \right\}.$$

Note that this arg max set is identical for all types with the same ratio $x \equiv \theta/\beta$ which implies that we can without loss in optimality assume the allocation depend only on x.

To see this note that the allocation may depend on θ and β independently for a given x only if the x-agent is indifferent amongst several pairs of u, w. However, the lemma in section 4 showed that the set of x for which agents are indifferent is

[4] Of course, if agents can only report at $t = 1$ then one cannot costlessly obtain truthfull reports on β. However, with large enough N it is likely that the cost of revealing β would be small.

of measure zero. As a consequence, allowing the allocation to depend on θ or β independently, in addition to x, cannot improve the objective function. Without loss in optimality we limit ourselves to allocations that are functions of x only[5].

The objective function is then:

$$E\left[\theta u\left(x\right)+w\left(x\right)\right]=E\left[E\left[\theta u\left(x\right)+w\left(x\right)\middle|\, x\right]\right]=\int\left[\alpha\left(x\right)u\left(x\right)+w\left(x\right)\right]\hat{f}\left(x\right)dx$$

where $\alpha\left(x\right)=E\left(\theta\middle|\, x\right)$ and $\hat{f}\left(x\right)$ is the density over x. Let $X=[\underline{x},\bar{x}]$ be the support of x and $\hat{F}\left(x\right)$ be its cumulative distribution.

Define x_p as the lowest value such that for $\hat{x}\geq x_p$

$$\frac{E\left[\alpha\left(x\right)\middle|\, x\geq\hat{x}\right]}{\hat{x}}\leq 1$$

We modify our previous assumption A in the following way.

Assumption $\tilde{\mathbf{A}}$. *For $x\in[\underline{x},x_p]$, we have that*

$$\frac{x\hat{f}'\left(x\right)}{\hat{f}\left(x\right)}\geq-\frac{2-\alpha'\left(x\right)}{1-\alpha\left(x\right)/x}$$

Note that without heterogeneity $\alpha\left(x\right)=\beta x$ so that assumptions $\tilde{A}$ and A are equivalent in this case.

Heterogenous Temptation Problem

$$\max_{u(\cdot),w(\cdot)}\int\left\{\alpha\left(x\right)u\left(x\right)+w\left(x\right)\right\}d\hat{F}\left(x\right)$$

subject to

$$xu\left(x\right)+w\left(x\right)=\underline{x}u\left(\underline{x}\right)+w\left(\underline{x}\right)+\int^{x}u\left(x\right)dx$$

$$c\left(u\left(x\right)\right)+k\left(w\left(x\right)\right)\leq 1$$

$$u\left(x\right)\geq u\left(y\right)\text{ for all }x\geq y$$

[5]Given $\beta\leq 1$ a simpler argument is available. The planner can simply instruct agents with given x to choose the element of the arg max with the lowest u, since the planner has a strict preference for the lowest u element. The argument in the text is similar to the one used to extend the analysis to arbitrary horizons and can be applied to the case where $\beta>1$ is allowed.

The proof of the next result closely follows the proof of proposition 4.

Proposition 10 *Under assumption $\tilde{A}$ agents with $x < x_p$ are offered their unconstrained optimum and agents with $x \geq x_p$ are bunched at the unconstrained optimum for agent x_p.*

6 Commitment with Self Control

Gul and Pesendorfer (2001,2002a,b) introduced an axiomatic foundation for preferences for commitment. We review their setup and representation result briefly in general terms and then describe how we apply it to our framework.

In their static formulation the primitive is a preferences ordering over sets of choices, with utility function $P(A)$ over choice sets A. In the classical case $P(A) = \max_{a \in A} p(a)$ for some utility function p defined directly over actions. Note that in this case if a set A is reduced to A' without removing the best element, a^* from A, then P is not altered. In this sense, committment, a preference for smaller sets, is not valued.

To model a preference for commitment they assume a consumer may strictly prefer a set A' that is a strict subset A, i.e. $P(A') > P(A)$ and $A' \subset A$. They show that such preferences can be represented by two utility functions $\tilde{U}$ and $\tilde{V}$ over choices a by the relation:

$$P(A) = \max_{a \in A} \{p(a) + t(a)\} - \max_{a \in A} t(a)$$

One can think of $t(a) - \max_{a \in A} t(a)$ as the cost of self-control suffered by an agent when choosing a instead of $\arg\max_{a \in A} t(a)$. In a dynamic setting recursive preferences with temptation can be represented similarly (Gul and Pessendorfer (2002a,b).

In our framework the action is a choice for current consumption and savings, c and k. In order to nest the hyperbolic preferences model we follow Krusell, Kuruscu and Smith (2001) and use:

$$p(c,k) = \theta u(c) + w(k)$$

$$t(c,k) = \phi(\theta u(c) + \beta w(k))$$

where the parameter $\phi > 0$ captures the lack of self control. As $\phi \to \infty$ the agent has no self-control and yields fully to his temptation. His preferences essentially converge to those implied by the hyperbolic model. The only difference is that in the limit as $\phi \to \infty$ we obtain a tie-breaking criteria that whenever an agent is indifferent he selects whatever is best for his previous 'selves' (i.e. maximizes t).

The objective function is:

$$\int_{\underline{\theta}}^{\bar{\theta}} \left(\theta \hat{u}\left(\theta\right) + \hat{w}\left(\theta\right)\right) f\left(\theta\right) d\theta + \phi \int_{\underline{\theta}}^{\bar{\theta}} \left(\theta \hat{u}\left(\theta\right) + \beta \hat{w}\left(\theta\right)\right) f\left(\theta\right) d\theta \\ - \phi \int_{\underline{\theta}}^{\bar{\theta}} \left(\theta u\left(\theta\right) + \beta w\left(\theta\right)\right) f\left(\theta\right) d\theta$$

where (u, w) is the allocation for the "temptation agent" and $(\hat{u}, \hat{w})$ is the allocation of that is chosen by the "self-control agent". It is convenient to define everything for a support larger than Θ given by $\hat{\Theta} \equiv [\underline{\hat{\theta}}, \bar{\theta}]$ where $\underline{\hat{\theta}} = \underline{\theta}\beta/\hat{\beta} < \underline{\theta}$.

Given a set of offered (u, w) pairs, the "self-control agent" will choose an allocation that maximizes $\theta\hat{u} + \hat{\beta}\hat{w}$ with $\hat{\beta} \equiv (1 + \phi\beta) / (1 + \phi)$, while the "temptation agent" will choose an allocation that maximizes $\theta u + \beta w$. Since both the "temptation" and the "self-control" agents choose from the same set it follows that,

$$\hat{u}\left(\theta\right) = u(\theta\beta/\hat{\beta}), \tag{8}$$

so that the "self-control agent" θ acts as a "temptation-agent" with a lower taste shock.

Substituting (8) into the objective function we obtain,

$$(1 + \phi) \int_{\underline{\theta}}^{\bar{\theta}} \left(\theta u\left(\theta\beta/\hat{\beta}\right) + \hat{\beta} w\left(\theta\beta/\hat{\beta}\right)\right) f\left(\theta\right) d\theta - \phi \int_{\underline{\theta}}^{\bar{\theta}} \left(\theta u\left(\theta\right) + \beta w\left(\theta\right)\right) f\left(\theta\right) d\theta.$$

The first term can be shown to equal,

$$(1 + \phi)\, \hat{\beta}/\beta \int_{\underline{\theta}\beta/\hat{\beta}}^{\bar{\theta}\beta/\hat{\beta}} \left(\theta u\left(\theta\right) + \beta w\left(\theta\right)\right) f\left(\theta\hat{\beta}/\beta\right) \hat{\beta}/\beta d\theta.$$

Note that $h\left(\theta\right) \equiv f\left(\theta\hat{\beta}/\beta\right) \hat{\beta}/\beta$ is the density of the random variable $\theta\beta/\hat{\beta}$; let $H\left(\theta\right)$ represent its corresponding distribution function.

The objective function can be written as,

$$(1+\phi)\frac{\hat{\beta}}{\beta}\int_{\underline{\hat{\theta}}}^{\frac{\beta}{\hat{\beta}}\bar{\theta}}\left(\theta u\left(\theta\right)+\beta w\left(\theta\right)\right)h\left(\theta\right)d\theta-\phi\int_{\underline{\theta}}^{\bar{\theta}}\left(\theta u\left(\theta\right)+\beta w\left(\theta\right)\right)f\left(\theta\right)d\theta.$$

Substituting in the incentive constraint,

$$\theta u\left(\theta\right)+\beta w\left(\theta\right)=\int_{\underline{\hat{\theta}}}^{\theta}u(\tilde{\theta})d\tilde{\theta}+v_0,$$

where $v_0=(\underline{\theta}\beta/\hat{\beta})u(\underline{\hat{\theta}})+\beta w(\underline{\hat{\theta}})$ and integrating by parts, we obtain:

$$(1+\phi)\frac{\hat{\beta}}{\beta}\int_{\underline{\hat{\theta}}}^{\frac{\beta}{\hat{\beta}}\bar{\theta}}\left(1-H\left(\theta\right)\right)u\left(\theta\right)d\theta-\phi\int_{\underline{\theta}}^{\bar{\theta}}\left(1-F\left(\theta\right)\right)u\left(\theta\right)d\theta+\left(\left(1+\phi\right)\hat{\beta}/\beta-\phi\right)v_0.$$

where we are taking both intervals of integration as being from $\underline{\hat{\theta}}$ to $\bar{\theta}$ by letting $h\left(\theta\right)=0$, for all $\theta>\bar{\theta}\beta/\hat{\beta}$ and $f\left(\theta\right)=0$ for all $\theta<\underline{\theta}$. In addition we require u to be non-decreasing and the budget constraint:

$$v_0+\int_{\underline{\hat{\theta}}}^{\theta}u\left(\tilde{\theta}\right)d\tilde{\theta}-\theta u\left(\theta\right)\leq\beta K^{-1}\left(y-C\left(u\left(\theta\right)\right)\right).$$

Definition. *Let $\hat{\theta}_p$ be the lowest value such that for $\hat{\theta}\geq\hat{\theta}_p$,*

$$\int_{\hat{\theta}}^{\bar{\theta}}\left((1+\phi)\frac{\hat{\beta}}{\beta}\left(1-H\left(\theta\right)\right)-\phi\left(1-F\left(\theta\right)\right)\right)d\theta\leq0$$

Assumption $\hat{\mathbf{A}}$. *For $\theta\in[\underline{\theta},\hat{\theta}_p]$, we have that*

$$(1+\phi)\left(\hat{\beta}/\beta\right)^2 f\left(\theta\hat{\beta}/\beta\right)-\phi f\left(\theta\right)\geq0$$

We now show that the optimal allocation is to offer all types below $\hat{\theta}_p$ their unconstrained optimum and to bunch types higher than $\hat{\theta}_p$ at the unconstrained optimum for $\hat{\theta}_p$. Denote this allocation by $(\underline{w}^*,u)$ as before.

The Lagrangian is

$$L \equiv \int_{\frac{\beta}{\hat{\beta}}\underline{\theta}}^{\bar{\theta}} \left((1+\phi)\frac{\hat{\beta}}{\beta}(1-H(\theta)) - \phi(1-F(\theta)) - (1-\Lambda(\theta)) \right) u(\theta)\, d\theta$$
$$+ \int_{\frac{\beta}{\hat{\beta}}\underline{\theta}}^{\bar{\theta}} \left(\beta K^{-1}(y - C(u(\theta))) + \theta u \right) d\Lambda$$
$$+ \left((1+\phi)\frac{\hat{\beta}}{\beta} - \phi - (1-\Lambda(\underline{\theta})) \right) v_0$$

where Λ is a non-decreasing Lagrange multiplier for the budget constraint, normalized so that $\Lambda(\bar{\theta}) = 1$.

Proposition 11 *The allocation $(\underline{w}^*, u)$ is optimal if and only if assumption $\hat{A}$ holds.*

Proof. We follow the proof of Proposition 4 as closely as possible. At the proposed allocation we have:

$$\partial L(\underline{w}, u; h_{\underline{w}}, h_u|\Lambda) \equiv \int_{\frac{\beta}{\hat{\beta}}\underline{\theta}}^{\bar{\theta}} \left((1+\phi)\frac{\hat{\beta}}{\beta}(1-H(\theta)) - \phi(1-F(\theta)) - (1-\Lambda(\theta)) \right) h_u(\theta)\, d\theta$$
$$+ \theta_p \int_{\frac{\beta}{\hat{\beta}}\underline{\theta}}^{\bar{\theta}} \left(\frac{\theta}{\theta_p} - 1 \right) \chi_{[\theta_p, \bar{\theta}]} h_u d\Lambda + \left[(1+\phi)\frac{\hat{\beta}}{\beta} - \phi - (1-\Lambda(\underline{\theta})) \right] h_{v_0}$$

Equation (5) requires $\Lambda(\underline{\theta}) = (1+\phi)(\hat{\beta}/\beta - 1)$. Using this and integrating (3) by parts leads to

$$\partial L(\underline{w}, u; h_{\underline{w}}, h_u|\Lambda) = \gamma(\underline{\theta}) h_u(\underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \gamma(\theta)\, dh_u(\theta) \tag{9}$$

where,

$$\gamma(\theta) \equiv \int_{\theta}^{\bar{\theta}} \left((1+\phi)\frac{\hat{\beta}}{\beta}(1-H(\theta)) - \phi(1-F(\theta)) - (1-\Lambda(\theta)) \right) d\theta + \theta_p \int_{\theta}^{\bar{\theta}} \left(\frac{\theta}{\theta_p} - 1 \right) \chi_{[\theta_p, \bar{\theta}]} d\Lambda$$

It follows that (4) requires $\gamma(\theta) = 0$ for $\theta \in \left[\underline{\hat{\theta}}, \theta_p\right]$, i.e. where u is strictly increasing.

This in turn requires,

$$\Lambda(\theta) = 1 - (1+\phi)\frac{\hat{\beta}}{\beta}(1 - H(\theta)) + \phi(1 - F(\theta)) \tag{10}$$

$\theta \in [\underline{\hat{\theta}}, \theta_p)$. Given the proposed allocation, this defines a unique multiplier Λ in the separating region $[\underline{\hat{\theta}}, \theta_p)$ and assumption $\hat{A}$ is necessary and sufficient for $\Lambda(\theta)$ to be non-decreasing. It follows that assumption $\hat{A}$ is necessary for the proposed solution to be optimal.

We now prove sufficiency by showing that there exists a non-decreasing multiplier Λ over the whole range Ω such that the proposed $\bar{w}, u$ satisfies (4) and (5). We've specified Λ for $[\underline{\hat{\theta}}, \theta_p)$ so we only need to specify the value of Λ for $\left[\theta_p, \overline{\theta}\right]$ and we set $\Lambda(\theta) = 1$ in this interval.

Note that the constructed Λ is not continuous at θ_p. To show that Λ is non-decreasing all that remains is to show that the jump at θ_p is upward which requires:

$$-(1+\phi)\frac{\hat{\beta}}{\beta}(1 - H(\theta)) + \phi(1 - F(\theta)) \leq 0$$

This follows from the definition of θ_p for $\theta_p > \underline{\theta}$: the lowest $\hat{\theta}$ such that $\gamma(\theta) \leq 0$ for all $\theta \geq \hat{\theta}$, which implies $\gamma'(\theta_p) = -(1+\phi)\frac{\hat{\beta}}{\beta}(1 - H(\theta)) + \phi(1 - F(\theta)) \leq 0$. If $\theta_p = \underline{\hat{\theta}}$ then the result is immediate.

The proposed allocation, $\bar{w}$ and u, and multiplier, Λ, imply that $\gamma \leq 0$ and that $\gamma = 0$ whenever u is increasing. Using (6) it follows that (4) and (5) are satisfied. ■

The next result establishes a connection between assumptions A and $\hat{A}$ showing that $\hat{A}$ is a weaker requirement.

Proposition 12 *If the condition for assumption A holds for $[\underline{\theta}, \hat{\theta}_p\hat{\beta}/\beta]$, then the condition for assumption $\hat{A}$ holds for $[\underline{\theta}, \hat{\theta}_p]$.*

Proof. Let $\phi = \frac{1}{\varepsilon} > 0$ then assumption $\hat{A}$ is equivalent to

$$\Phi(\theta, \varepsilon) \equiv (1+\varepsilon)\left(\hat{\beta}(\varepsilon)/\beta\right)^2 f\left(\theta\hat{\beta}(\varepsilon)/\beta\right) - f(\theta) \geq 0$$

with $\hat{\beta}(\varepsilon) = (\beta + \varepsilon)/(1 + \varepsilon)$. Note that $\Phi(\theta, 0) = 0$,

$$\Phi_\varepsilon(\theta, \varepsilon) = \frac{\hat{\beta}^2}{\beta^2} f\left(\theta\hat{\beta}/\beta\right) + \frac{1+\varepsilon}{\beta^2}\left(2\hat{\beta} f\left(\theta\hat{\beta}/\beta\right) + \hat{\beta}^2 f'\left(\theta\hat{\beta}/\beta\right)\theta/\beta\right)\hat{\beta}'(\varepsilon),$$

and $\hat{\beta}'(\varepsilon) = (1-\beta)/(1+\varepsilon)^2$. Thus:

$$\Phi_\varepsilon(\theta, \varepsilon) = \frac{\hat{\beta}}{1+\varepsilon}\frac{1}{\beta^2}\left((2 - \beta + \varepsilon) f\left(\theta\hat{\beta}/\beta\right) + (1-\beta) f'\left(\theta\hat{\beta}/\beta\right)\theta\hat{\beta}/\beta\right)$$

assumption A holding at $\hat{\theta}$ implies that $(2-\beta) f(\hat{\theta}) + (1-\beta) f'(\hat{\theta})\hat{\theta} \geq 0$. This implies $(2 - \beta + \varepsilon) f(\hat{\theta}) + (1-\beta) f'(\hat{\theta})\hat{\theta} \geq 0$ for $\varepsilon \geq 0$. So if the condition in assumption A holds for $[\underline{\theta}, \hat{\theta}_p\hat{\beta}/\beta]$ then $\Phi_\varepsilon(\theta, \varepsilon) \geq 0$ for all $\varepsilon \geq 0$ and $\theta \in [\underline{\hat{\theta}}, \hat{\theta}_p]$. Given that $\Phi(\theta, 0) = 0$, we have that if A holds for $[\underline{\theta}, \hat{\theta}_p\hat{\beta}/\beta]$, then

$$\Phi(\theta, \varepsilon) = \Phi(\theta, 0) + \int_0^\varepsilon \Phi_\varepsilon(\theta, \tilde{\varepsilon})\, d\tilde{\varepsilon} \geq 0$$

for $\theta \in [\underline{\hat{\theta}}, \hat{\theta}_p]$ so that assumption $\hat{A}$ holds. ■

7 Disagreement on Exponential Discounting

This section departs a bit from our intra-personal temptation environment. We now consider the case where individuals discount the future exponentially but do so at a different rate than a 'social planner'. Caplin and Leahy (2001) and Phelan (2002) provide motivations for such an assumption. Here we simply explore the implications of such a difference in discount rates.

It is important to note that this modification constitutes more than just a departure on the form of discounting. Our previous analysis relied on the tensions within an individual due to temptation. In contrast, in the current case we require some paternalistic motivation for the social planner's disagreement with agents. As a consequence, some may view this case as more ad hoc and somehow less worthy of analysis. However, we believe that paternalistic motivations may be behind several government policies. In the next section we discuss other examples of paternalism.

For the two period case the analysis requires absolutely no change, only a different

interpretation for β. The difference in discounting in the incentive constraints versus the objective function now arises from an assumed difference in private and social discounting, it is no longer motivated by time inconsistent preferences. The relevant question that remains is whether we can extend the results to longer horizons.

A difficulty is that the planner and agent will disagree on more than just how much to discount future utility relative to present utility: now there is also disagreement on the evaluation of future lifetime utility itself. This makes a recursive formulation more difficult since the key simplification was that the same value function appeared in the objective function and in the incentive constraints.

Indeed, now we require two value functions, one for the planner, v, and one for the agent, v^A. Fortunately, in the case with logarithmic utility these two value functions are related in a simple way. This allows us to keep track of only one value function, v, rendering the analysis tractable. We can show that all our results go through in this case.

Consider first the situation with three periods. Let the exponential discount factor for the agent be given by β and for simplicity assume the discount factor for the social planner is unity.

The highest utility achievable by the agent in the last two periods is

$$v_2^A(y) = (1+\beta)\log(y) + B^A$$

for some constant B^A. For any homogenous mechanism the planner's value function for the last two periods takes the form:

$$v_2(y) = 2\log(y) + B^P$$

The important point is that these value functions differ only by constants and coefficients. As a consequence the correct incentive constraint for the first period can be written with either value function. That is,

$$\theta U(c(\theta)) + \beta v_2^A(y(\theta)) \geq \theta U\left(c\left(\tilde{\theta}\right)\right) + \beta v_2^A\left(y\left(\tilde{\theta}\right)\right),$$

is equivalent to,

$$\theta U\left(c\left(\theta\right)\right)+\hat{\beta}_3 v_2\left(y\left(\theta\right)\right)\geq\theta U\left(c\left(\tilde{\theta}\right)\right)+\hat{\beta}_3 v_2\left(y\left(\tilde{\theta}\right)\right), \tag{11}$$

where $\hat{\beta}_3=\beta(1+\beta)/2<1$ is a fictitious hyperbolic discount factor when there are three periods to go. Note that the incentive constraint (11) has all the features of the hyperbolic discounting case.

Thus, we can we can write the three period problem disagreement on exponential discounting in the same way as the hyperbolic discounting problem. The arguments generalizes to any finite horizon. With k remaining periods the discount factor that must be applied is

$$\hat{\beta}_k\equiv\beta\frac{1+\beta+...+\beta^{k-1}}{k}<1.$$

Note that β_k is decreasing in k and converges to zero (this last feature is special to the planner not discounting the future at all).

8 Other Interpretations

In this section we discuss how our model can be reinterpreted for other applications.

8.1 Paternalism

Another interesting application of the model to a paternalistic problem is the choice between schooling and leisure choice. In many cases the relevant agent is not yet an adult so that we can interpret paternalism literally as a struggle between the preferences of parents and child. Alternatively, other adults may be altruistically concerned about children without parents and support paternalistic legislation.

The child's preference are given by the utility function $\theta U\left(l\right)+\beta W\left(s\right)$where s represents schooling time and l represents other valuable uses of time. The taste parameter θ affects the relative value placed on schooling vs. other activities. The parent's preferences are given by $\theta U\left(s\right)+W\left(s\right)$, so that more weight is given to schooling time.

The allocation of time is constrained by the time endowment normalized to one, $s+l\leq1$. In this example no insurance is possible.

8.2 Externalities

Another origin for a divergence of preferences between the planner and the agents is when consumption of a good generates positive externalities. Agents do not internalize the effects of their consumption on other agents but the planner does.

To make this precise, suppose there are two goods, c and k, that the agent with taste shock θ obtains the following utility when the entire allocation is $(c(\theta), k(\theta))$

$$V(\theta; (c(\cdot), k(\cdot))) \equiv \theta U(c(\theta)) + \beta W(k(\theta)) + (1-\beta) \int W(k(\theta)) \, dF(\theta) \qquad (12)$$

The last term captures the externality from the consumption of k. The utilitarian welfare criterion is:

$$W = \int V(\theta, (c,k)) \, dF(\theta) = \int (\theta U(c(\theta)) + W(k(\theta))) \, dF(\theta).$$

Thus, we can represent $\theta U(c) + W(k)$ as the relevant utility function for agent-θ from the planner's point of view. Although this is not the utility actually attained by agent-θ, which is given by the expression in (12), it is an interpretation that leads to the same welfare functional.

9 Conclusion

This paper studied the optimal trade-off between commitment and flexibility in an intertemporal consumption/saving model without insurance. In our model, agents expect to receive relevant private information regarding their tastes which creates a demand for flexibility. But they also expect to suffer from temptations, and therefore value commitment. The model combined the representation theorems of preferences for flexibility introduced by Kreps (1979) with the preferences for commitment proposed by Gul and Pesendorfer (2002).

We solved for the optimal solution that trades-off commitment and flexibility by setting up a mechanism design problem. We showed that under certain conditions the optimal allocation takes the simple threshold form of a minimum savings requirement. We characterized the condition on the distribution of the shocks under which this result holds, and showed that if this condition is not satisfied, more complex

mechanisms might be optimal. Future work will focus on the case with insurance, with a hope of understanding how it may affect the results obtained here.

The model is open to a variety of other interpretations. A paternalistic principal who cares about an agent but believes the agent is biased on average in his choices would face a similar trade-off as long as the agent has some private information regarding his tastes that the planner also cares about. We discussed applications to schooling choices by teenagers and situations with externalities.

A Proof of Proposition 6

Suppose that we are offering a segment of the budget line between the tangency point for θ_L and that of θ_H, with associated allocation c_L and c_H. Define the θ^* that is indifferent from the allocation c_L and c_H then $\theta^* \in (\theta_L, \theta_H)$ for $\theta_H > \theta_L$. Upon removing the interval $\theta \in (\theta^*, \theta_H)$ types move to c_H and $\theta \in (\theta_L, \theta^*)$ types move to c_L allocation.

Let $\Delta(\theta_H, \theta_L)$ be the change in utility for the planner of such a move (normalizing income to $y = 1$ for simplicity)

$$\begin{aligned}\Delta(\theta_H, \theta_L) \equiv & \int_{\theta^*(\theta_H,\theta_L)}^{\theta_H} \{\theta U(c^*(\theta_H)) + W(y - c^*(\theta_H))\} f(\theta) d\theta \\ & + \int_{\theta_L}^{\theta^*(\theta_H,\theta_L)} \{\theta U(c^*(\theta_L)) + W(y - c^*(\theta_L))\} f(\theta) d\theta \\ & - \int_{\theta_L}^{\theta_H} \{\theta U(c^*(\theta)) + W(y - c^*(\theta))\} f(\theta) d\theta\end{aligned}$$

where the function $c^*(\theta)$ is defined implicitly by

$$\theta U'[c^*(\theta)] = \beta W'(y - c^*(\theta)) \tag{13}$$

and $\theta^*(\theta_H, \theta_L)$ is then defined by

$$\begin{aligned}&\theta^*(\theta_H, \theta_L) U(c^*(\theta_H)) + \beta W(y - c^*(\theta_H)) \\ &= \theta^*(\theta_H, \theta_L) U(c^*(\theta_L)) + \beta W(y - c^*(\theta_L))\end{aligned} \tag{14}$$

Notice that $\Delta(\theta_L, \theta_L) = 0$.

The following lemma regarding the partial derivative of $\Delta(\theta_H, \theta_L)$ is used below.

Lemma. The partial of $\Delta(\theta_H, \theta_L)$ with respect to θ_H can be expressed as:

$$\frac{\partial \Delta}{\partial \theta_H}(\theta_H, \theta_L) = S(\theta_H; \theta^*) \frac{U'(c^*(\theta_H))}{\beta} \frac{\partial c^*(\theta_H)}{\partial \theta_H}$$

where $S(\theta; \theta^*)$ is defined by,

$$S(\theta, \theta^*) \equiv (y - \beta)(\theta - \theta^*)\theta^* f(\theta^*) - \int_{\theta^*}^{\theta} \left(\theta - \beta\tilde{\theta}\right) f\left(\tilde{\theta}\right) d\tilde{\theta}$$

Since $U'(c^*(\theta_H)) > 0$ and $\frac{\partial c^*(\theta_H)}{\partial \theta_H} > 0$, then $sign(\Delta_1) = sign(S(\theta_H, \theta^*))$.

Proof. We have

$$\begin{aligned}
\Delta_1(\theta_H, \theta_L) = & \left[\theta_H U(c^*(\theta_H)) + W(y - c^*(\theta_H))\right] f(\theta_H) \\
& - \left[\theta^*(\theta_H, \theta_L) U(c^*(\theta_H)) + W(y - c^*(\theta_H))\right] f(\theta^*) \frac{\partial \theta^*}{\partial \theta_H} \\
& + \int_{\theta^*(\theta_H, \theta_L)}^{\theta_H} \left\{\theta U'(c^*(\theta_H)) - W'(y - c^*(\theta_H))\right\} f(\theta) \frac{\partial c^*(\theta_H)}{\partial \theta_H} d\theta \\
& + \left\{\theta^*(\theta_H, \theta_L) U(c^*(\theta_L)) + W(y - c^*(\theta_L))\right\} f(\theta^*) \frac{\partial \theta^*}{\partial \theta_H} \\
& - \left[\theta_H U(c^*(\theta_H)) + W(y - c^*(\theta_H)) f(\theta_H)\right]
\end{aligned}$$

Combining terms,

$\Delta_1(\theta_H, \theta_L) =$

$$\begin{aligned}
& \left(\int_{\theta^*(\theta_H, \theta_L)}^{\theta_H} \left\{\theta U'(c^*(\theta_H)) - W'(y - c^*(\theta_H))\right\} f(\theta) d\theta\right) \frac{\partial c^*(\theta_H)}{\partial \theta_H} \\
& + \left\{\theta^*(\theta_H, \theta_L) \left[U(c^*(\theta_L)) - U(c^*(\theta_H))\right] + W(y - c^*(\theta_L)) - W(y - c^*(\theta_H))\right\} f(\theta^*) \frac{\partial \theta^*}{\partial \theta_H}
\end{aligned}$$

Now, from (14) we have

$$\theta U'[c^*(\theta)] - W'(y - c^*(\theta)) = \left[\frac{\beta - 1}{\beta}\right] \theta U'[c^*(\theta)]$$

Substituting above

$$\Delta_1(\theta_H,\theta_L) =$$

$$\left(\int_{\theta^*(\theta_H,\theta_L)}^{\theta_H}\left(\theta-\frac{1}{\beta}\theta_H\right)f(\theta)\,d\theta\right)U'(c^*(\theta_H))\frac{\partial c^*(\theta_H)}{\partial\theta_H}$$
$$+\left\{\theta^*(\theta_H,\theta_L)\left[U(c^*(\theta_L))-U(c^*(\theta_H))\right]+W(y-c^*(\theta_L))-W(y-c^*(\theta_H))\right\}f(\theta^*)\frac{\partial\theta^*}{\partial\theta_H}$$

we also have that from (13)

$$-\frac{\theta^*(\theta_H,\theta_L)}{\beta}\left[U(c^*(\theta_L))-U(c^*(\theta_H))\right]=\left\{W(y-c^*(\theta_L))-W(y-c^*(\theta_H))\right\}$$

So,

$$\Delta_1(\theta_H,\theta_L)=\left\{\left[\frac{1}{\beta}-1\right]\theta^*f(\theta^*)\right\}\left[U(c^*(\theta_H))-U(c^*(\theta_L))\right]\frac{\partial\theta^*}{\partial\theta_H}$$
$$-\left(\int_{\theta^*}^{\theta_H}\left(\frac{1}{\beta}\theta_H-\theta\right)f(\theta)\,d\theta\right)U'(c^*(\theta_H))\frac{\partial c^*(\theta_H)}{\partial\theta_H}$$

Differentiating (14) we obtain:

$$\frac{\partial\theta^*}{\partial\theta_H}\left[U(c^*(\theta_H))-U(c^*(\theta_L))\right]=-\left[\theta^*U'(c^*(\theta_H))-\beta W'(y-c^*(\theta_H))\right]\frac{\partial c^*(\theta_H)}{\partial\theta_H}$$

Using the fact that $\theta U'[c^*(\theta)]-\beta W'(1-c^*(\theta))=0$ this implies

$$\frac{\partial\theta^*}{\partial\theta_H}\left[U(c^*(\theta_H))-U(c^*(\theta_L))\right]=\left[\theta_H-\theta^*\right]U'[c^*(\theta_H)]\frac{\partial c^*(\theta_H)}{\partial\theta_H}$$

Substituting back the result follows. ■

From the lemma we only need to sign $S(\theta_H,\theta^*)$. Clearly, $S(\theta^*,\theta^*)=0$. Taking derivatives we also get that

$$\frac{\partial S(\theta,\theta^*)}{\partial\theta}=[1-\beta]\theta^*f(\theta^*)-(1-\beta)\theta f(\theta)-\int_{\theta^*}^{\theta}f\left(\tilde{\theta}\right)d\tilde{\theta}$$

Notice that

$$\left.\frac{\partial S(\theta,\theta^*)}{\partial\theta}\right|_{\theta^*}=0$$

$$\frac{\partial^2S(\theta,\theta^*)}{(\partial\theta)^2}=-(2-\beta)f(\theta)-(1-\beta)\theta f'(\theta)$$

Note that $\partial^2 S(\theta,\theta^*)/(\partial\theta)^2$ does not depend on θ^*, just on θ. It follows that $sign\left(\frac{\partial^2 S(\theta,\theta^*)}{(\partial\theta)^2}\right) \leq 0$ if and only if

$$\frac{\theta f'(\theta)}{f(\theta)} \geq -\frac{2-\beta}{1-\beta} \tag{15}$$

That is, if A holds. Integrating $\partial^2 S(\theta,\theta^*)/(\partial\theta)^2$ twice:

$$S(\theta_H,\theta^*) = \int_{\theta^*}^{\theta_H}\int_{\theta^*}^{\theta} \frac{\partial^2 S\left(\tilde{\theta},\theta^*\right)}{\left(\partial\tilde{\theta}\right)^2} d\tilde{\theta}d\theta$$

Thus $S(\theta_H,\theta^*) \leq 0$ if A holds.

This implies then that $\Delta_1(\theta,\theta_L) \leq 0$ for all $\theta \geq \theta_L$ if assumption A holds; and

$$\Delta(\theta_H,\theta_L) = \int_{\theta_L}^{\theta_H} \Delta_1(\theta;\theta_L)\, d\theta$$

so that

$$\frac{\theta f'(\theta)}{f(\theta)} \geq -\frac{2-\beta}{1-\beta} \Rightarrow \Delta(\theta_H,\theta_L) \leq 0 \quad ; \text{for all } \theta_H \text{ and } \theta_L$$

and clearly $\theta_L \in \arg\max_{\theta_H \geq \theta_L} \Delta(\theta_H,\theta_L)$. In other words if assumption A holds then punching holes into any offered interval is not optimal.

The converse is also true: if A does not hold for some open interval $\theta \in (\theta_1,\theta_2)$ then the previous calculations show that it is optimal to remove the whole interval. In other words,

$$(\theta_1,\theta_2) \in \arg\max_{\theta_L,\theta_H} \Delta(\theta_H,\theta_L)$$
$$\text{s.t. } \theta_1 \leq \theta_L \leq \theta_H \leq \theta_2$$

This concludes the proof. ■

B Lemma on Differentiability

Definition. Given a function $T : \Omega \to Y$, where $\Omega \subset X$ and X and Y are normed spaces. If for $x, h \in \Omega$ the limit:

$$\lim_{\alpha \to 0} \frac{1}{\alpha} \left[T\left(x + \alpha h \right) - T\left(x \right) \right]$$

exists then it is called the *Gateaux differential* for $x, h \in \Omega$ and is denoted by $\partial T\left(x; h \right)$.

Lemma. Let

$$T\left(x \right) = \int_{\Theta} g\left(x\left(\theta \right), \theta \right) d\mu\left(\theta \right)$$

$(\Theta, \tilde{\Theta}, \mu)$ is any measure space (not necessarily R or a vector space) and $x : \Theta \to R^n$ in some space Ω for which T is defined (an arbitrary restriction or perhaps a required restriction to ensure measurability and integrability). Suppose $g\left(\cdot, \theta \right)$ is concave and $g_x\left(\cdot, \theta \right)$ exists and is continuous in x, for all θ. Then as long as $x + \alpha h \in \Omega$ for $\alpha \in [0, \varepsilon]$ for some $\varepsilon > 0$ (a minimum requirement for existence! of course) then:

$$\partial T\left(x; h \right) = \int_{\Theta} g_x\left(x\left(\theta \right), \theta \right) h\left(\theta \right) d\mu\left(\theta \right)$$

if this expression is well defined.

Proof. By definition

$$\begin{aligned}
\partial T\left(x; h \right) &= \lim_{\alpha \to 0} \frac{1}{\alpha} \left[T\left(x + \alpha h \right) - T\left(x \right) \right] \\
&= \lim_{\alpha \to 0} \int_{\Theta} \frac{1}{\alpha} \left[g\left(x\left(\theta \right) + \alpha h\left(\theta \right), \theta \right) - g\left(x\left(\theta \right), \theta \right) \right] d\mu\left(\theta \right) \\
&= \int_{\Theta} g_x\left(x\left(\theta \right), \theta \right) h\left(\theta \right) d\mu \\
&+ \lim_{\alpha \to 0} \int_{\Theta} \left[\frac{1}{\alpha} \left[g\left(x\left(\theta \right) + \alpha h\left(\theta \right), \theta \right) - g\left(x\left(\theta \right), \theta \right) \right] - g_x\left(x\left(\theta \right), \theta \right) h\left(\theta \right) \right] d\mu\left(\theta \right)
\end{aligned}$$

since for $\alpha < \varepsilon$ we have

$$\begin{aligned}
&\left| \frac{1}{\alpha} \left[g\left(x\left(\theta \right) + \alpha h\left(\theta \right), \theta \right) - g\left(x\left(\theta \right), \theta \right) \right] - g_x\left(x\left(\theta \right), \theta \right) h\left(\theta \right) \right| \\
&\leq \left| \frac{1}{\varepsilon} \left[g\left(x\left(\theta \right) + \varepsilon h\left(\theta \right), \theta \right) - g\left(x\left(\theta \right), \theta \right) \right] - g_x\left(x\left(\theta \right), \theta \right) h\left(\theta \right) \right|
\end{aligned}$$

by concavity of g.

Since $g(x(\theta)+\varepsilon h(\theta),\theta)$, $g(x(\theta),\theta)$ and $g_x(x(\theta),\theta)h(\theta)$ are integrable by assumption it follows that $\frac{1}{\varepsilon}[g(x(\theta)+\varepsilon h(\theta),\theta)-g(x(\theta),\theta)]-g_x(x(\theta),\theta)h(\theta)$ is also integrable. Since any function f is integrable if and only if $|f|$ is integrable [see Exercise 7.26, pg. 192, chapter 8, SLP] we have that $|\frac{1}{\varepsilon}[g(x(\theta)+\varepsilon h(\theta),\theta)-g(x(\theta),\theta)]-g_x(x(\theta),\theta)h(\theta)|$ is integrable. We then have the conditions for Lebesgue's Dominated Convergence Theorem.

It follows that:

$$\lim_{\alpha\to 0}\int_\Theta\left[\frac{1}{\alpha}[g(x(\theta)+\alpha h(\theta),\theta)-g(x(\theta),\theta)]-g_x(x(\theta),\theta)h(\theta)\right]d\mu$$
$$=\int_\Theta\left[\lim_{\alpha\to 0}\frac{1}{\alpha}[g(x(\theta)+\alpha h(\theta),\theta)-g(x(\theta),\theta)]-g_x(x(\theta),\theta)h(\theta)\right]d\mu=0$$

by continuity of g_x and its definition. It follows that $\partial T(x;h)=\int_\Theta g_x(x(\theta),\theta)h(\theta)\,d\mu$.
■

Remark: Suppose Ω is convex and that we are interested in $\delta T(x_0;h)$ then the obvious requirement that $x_0+\alpha h\in\Omega$ for $\alpha\in[0,\varepsilon]$ for some $\varepsilon>0$ is satisfied if and only if $h=x_1-x_0$ and $x_1\in\Omega$. Note that the case where Ω is the space of non-decreasing functions is convex and so is the space of measurable functions.

References

[1] **Athey, Susan; Andy Atkeson and Patrick Kehoe** (2003) "On the Optimality of Transparent Monetary Policy", Staff Report 326.

[2] **Atkeson, Andrew and Robert E. Lucas, Jr.** (1992) "On Efficient Distribution with Private Information." *Review of Economic Studies*, 59, no.3 (July): 427-53.

[3] **Caplin, Andrew and John Leahy** "The Social Discount Rate", mimeo, New York University.

[4] **Cole, Harold L. and Narayana R. Kocherlakota** (2001) "Efficient Allocations with Hidden Income and Hidden Storage", *Review of Economic Studies,* v68, n3 (July): 523-42.

[5] **Diamond, Peter** (1977) "A Framework for Social Security Analysis", *Journal of Public Economics*, v8, n3: 275-98.

[6] **Diamond, Peter and Botand Koszegi** (2002) "Quasi-Hyperbolic Discounting and Retirement", mimeo, MIT.

[7] **Feldstein, Martin S.** (1985)"The Optimal Level of Social Security Benefits", *Quarterly Journal of Economics,* v10, n2, May: 303-20

[8] **Fudenberg, D. and Tirole, J.** , "Game Theory". 1991. MIT Press.

[9] **Gul, Faruk and Wolfgang Pesendorfer**, 2001 "Temptation and Self-Control", *Econometrica*, Vol. 69 (6) pp. 1403-35

[10] **Gul, Faruk and Wolfgang Pesendorfer**, 2002a "Self-control and the theory of consumption", forthcoming in *Econometrica*.

[11] **Gul, Faruk and Wolfgang Pesendorfer**, 2002b. "Self-control, revealed preference and consumption choice", prepared for Society for Economic Dynamics.

[12] **Holmstrom, Bengt** (1984) "On the Theory of Delegation", in Bayesian Models in Economic Theory, Studies in Bayesian Econometrics, vol. 5, edited by Boyer, Marcel and Kihlstrom, Richard E.: 115-41.

[13] **Imrohoroglu, Ayse; Selahattin Imrohoroglu and Douglas H. Joines** (2000)"Time Inconsistent Preferences and Social Security", Federal Reserve of Minneapolis Discussion Paper Number 136

[14] **Kocherlakota, Narayana** 1996 "Reconsideration-Proofness: A Refinement for Infinite Horizon Time Inconsistency", *Games and Economic Behavior*, Vol. 15 (1), pp 33-54

[15] **Kreps, David**, 1979 "A Preference for Flexibility", *Econometrica*, 47: 5 65-576.

[16] **Krusell, Per; Burhanettin Kuruscu and Anthony Smith** (2001) "Temptation and Taxation", mimeo.

[17] **Laibson, David (**1994) "Self-Control and Saving", M.I.T. thesis.

[18] **Laibson, David** (1998)"Life-Cycle Consumption and Hyperbolic Discount Functions", *European Economic Review*, Vol. 42, nos. 3-5: 861-871.

[19] **Luenberger, David G.** (1969) "Optimization by Vector Space Methods", John Wiley & Sons, Inc.

[20] **Mirrlees, James** (1971) "An Exploration in the Theory of Optimum Income Taxation", The Review of Economic Studies, Vol. 38, No. 2., pp. 175-208.

[21] **O'Donaghue, Ted and Matthew Rabin** (2003) "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes", mimeo, UC Berkeley.

[22] **Phelan, Christopher** (2002) "Opportunity and Social Mobility", mimeo, Federal Reserve Bank of Minneapolis.

[23] **Phelps, Edmund S. and R.A. Pollack** (1968) "On Second Best National Savings and Game-Equilibrium Growth", *Review of Economic Studies*, Vol. 35, No. 2., pp. 185-199.

[24] **Ramsey, F.P.** (1928) "A Mathematical Theory of Savings" *Economic Journal*, vol. 38, 543-559.

[25] **Sheshinski, Eytan** (2002) "Bounded Rationality and Socially Optimal Limits on Choice in a Self-Selection Model", mimeo.

[26] **Weitzman, Martin L.**, (1974) "Prices vs. Quantities", *Review of Economic Studies*, v41, n4 : 477-91.