Commitment vs. Flexibility*

Manuel Amador†
Ivan Werning‡
George-Marios Angeletos§

February 23, 2004

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 with or without self-control – 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 (2001), or alternatively, 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. We also show that in these cases the optimal commitment device can be implemented sequentially by allowing the agent to manage a portfolio of liquid and illiquid assets.

---

*We'd like to thank comments and suggestions from Daron Acemoglu, Andy Atkeson, Francisco Buera, Peter Diamond and especially Pablo Werning. We also thank seminar participants at MIT, AEA, Rochester, UPenn, Maryland, Berkeley, NYU, Yale, Stanford SITE, Cleveland Fed, University of Texas at Austin, Boston Fed.

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

Introduction

A commonly articulated justification for government involvement in retirement income 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'.

More generally, if people suffer from temptation and self-control problems, what should be done to help them? Current models emphasizing these problems lead to a simple and extreme answer: it is optimal to completely remove all the individual's future choices. In particular, if the temptation is for present consumption, as is commonly assumed in the intertemporal framework, it is desirable to commit individuals to a particular consumption-savings path, removing all future consumption-savings choices. In these models the preference for commitment devices is simply overwhelming.

This paper studies the non-trival design of optimal commitment devices by modeling situations where removing all individual choices is not necessarily optimal. We introduce a value for flexibility and study the resulting trade-off with commitment. Let us briefly describe what we mean by these two concepts and how we formalize them.

Commitment entails an ex-ante reduction in the options available to an individual ex-post. Models with time-inconsistent preferences solved as a competitive game, as in Strotz (1956), were the first to formalize a value for commitment. In particular, in the consumption/saving framework the hyperbolic discounting model has proven useful for studying the effects of a temptation to 'over-consume' and the resulting desirability of certain commitment devices (e.g. Phelps and Pollack, 1968, Laibson, 1997).

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 possibly exert costly self-control. Commitment is valued because it avoids temptations that may either adversely affect choices or require exerting costly self-control.

On the opposite side of the spectrum, Kreps (1979) provided an axiomatic foundation for a preference for flexibility, preferences that place an ex-ante value on having a large set of options available for the individual ex-post. Kreps' representation theorem shows that such preferences can be represented by including taste shocks into an expected utility framework.

Our model incorporates a preference for flexibility and for commitment by combining Kreps' representation with both the Gul and Pesendorfer temptation model and the time-inconsistent preference framework.[1] The individual has preferences over random consumption streams. Agents suffer from the temptation for higher present consumption. In addition, each period a taste shock is realized that affects the individual's desire for current vs. future consumption. Importantly, taste shocks are assumed to be private information.

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. If the taste shocks were observable one could simply commit to a contingent consumption rule that depends on them. Because the information is private only the agent can act on it.

We require that any feasible consumption path satisfy the agent's budget constraint. By a commitment device we mean a further restriction on the set of feasible allocations. In a dynamic setting this requires the following arrangement. In period t the agent faces a period t decision problem: he must choose from a menu of current consumption choices and decision problems for the next period, $t+1$. In the last period the decision problem is simply a menu over the last period's consumption level.

To fix ideas, consider two extreme arrangements. The first offers no commitment and full flexibility: the agent in each period can choose consumption freely constrained only by his budget set. In this case past choices regarding consumption have an impact only through the remaining available resources. The other arrangement offers

[1] See Dekel, Lipman and Rustichini (2001) for axiomatic foundations and a representation theorem for preferences over choice sets that encompasses both Kreps and Gul and Pessendorfer's frameworks.

full commitment but no flexibility: each period the agent's menu is a singleton with a unique consumption choice from his budget set.

Of course, one can devise schemes where the menu available depends on past choices in richer ways. In this way we can offer both commitment and flexibility in varying degrees. We solve for the optimal commitment device, or decision problem, that the agent faces.

We show that it is convenient to map this problem into a mechanism design problem. Solving for the informationally constrained optimal allocation of this related problem then delivers the optimal commitment device.

We begin by considering a simple case without self-control 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 people 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.

Up to this point we consider the commitment device as being established in the initial period. In Section 6 we show that the optimal commitment mechanism can be implemented sequentially: each period the agent makes a consumption/savings choice subject to a minimum savings level and selects the minimum savings level for the next period. The choice regarding the minimum savings level for the next period thus increases the amount of choices made each period. Despite this difference we show that this arrangement continues to implement the same allocation and welfare as the optimal commitment device chosen in the initial period.

We then show that the optimal allocation and welfare can be obtained as the solution to a simple liquid-illiquid asset portfolio problem. In this implementation the only instruments available to the agent are a liquid asset, that can be used for current consumption, and an illiquid asset, that can only be consumed next period. This is the liquid-illiquid arrangement studied in Laibson (1997). Each period the investment choice in the illiquid asset mimics perfectly a choice of a minimum savings level for the next period. It follows that this arrangement also obtains the optimal allocation and welfare.

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 contains the more general case with temptation and self-control proposed by Gul and Pesendorfer (2001,2002a,b). Section 5 extends the analysis to arbitrary finite time horizons. Section 6 shows how the mechanism can be implemented sequentially and discusses an implementation with liquid and illiquid assets. 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.

Several papers involve a trade-off similar in ways to the one 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 agents that make choices over a discrete set of alternatives subject to random errors. Laibson (1994, Chapter 3) considers a moral-hazard model with a hyperbolic-discounting agent.

Dekel, Lipman and Rustichini (2001) provide a representation theorem for preferences that may value commitment and flexibility. Our time-consistent preference specification is motivated by combining Kreps (1979) with Gul and Pesendorfer (2001). However, for the analysis we find it useful to arrive at Dekel, Lipman and Rustichini's representation. Our paper is not choice theoretic in that we do not axiomatize preferences to derive representation theorems. Rather, we are users applying representation theorems derived in previous choice theoretic contributions.

1 Basic Setup

We begin by studying the allocation problem over two consumption periods for the case of a consumer without self-control. Section 4 extends the model to the more general framework with self-control.

There are two interpretations for the two period model without self-control: (i) a consumer with time-inconsistent hyperbolic-discounting preferences; (ii) a consumer with time-consistent preferences that suffers from temptations, but cannot exert self-control. In the two period model the differences between these two frameworks are mostly about interpretation. Section 5 extends the two-period model to more periods where the differences between the two frameworks are more important.

1.1 Time-Inconsistent Preferences

Here we follow Strotz (1956), Phelps and Pollack (1968), Laibson (1994, 1997, 1998) and many others by modeling the agent in each period as different *selves,* with different preferences. The approach of the time-inconsistent preference literature thus takes a game-theoretic perspective of the commitment desire.

Consider two periods of consumption: $t = 1, 2$. Each period individuals receive an i.i.d. taste shock $\theta \in \Theta$, normalized so that $\mathrm{E}\theta = 1$ which affects the marginal utility of current consumption: higher θ make current consumption more valuable. The taste shock is assumed to be private information. Although we model the shock as one to preferences, perhaps it is best thought of as a catch-all for the significant variation one observes in consumption and saving data after conditioning on available observable variables. Indeed, with exponential utility unobservable income shocks are equivalent to unobservable taste shocks.

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 then

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

where $U : \mathbb{R}^+ \to \mathbb{R}$ and $W : \mathbb{R}^+ \to \mathbb{R}$ are increasing, concave and continuously differentiable and $\beta \leq 1$. The notation allows $W(\cdot) \neq U(\cdot)$, this generality facilitates the extension to N periods in section 5.

Utility for *self-0* from periods $t = 1, 2$ is given by:

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

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. 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.

Without commitment self-1 enjoys full flexibility and the set of available choices that we consider is given by the resource constraint $B(y) \equiv \{(c, y) \mid c + k \leq y\}$, where we have normalized the interest rate between periods to zero. Facing $B(y)$ in equilibrium self-1 with taste shock θ solves:

$$\max_{(c,k)\in B(y)} \theta U(c) + W(k)$$

Denote the unique solution to this problem by, $(c^{flex}(\theta), k^{flex}(\theta))$. The ex-ante utility achieved by self-0 from $B(y)$ is thus given by,

$$\int [\theta U\left(c^{flex}(\theta)\right) + W\left(k^{flex}(\theta)\right)] dF(\theta)$$

In this two-period model providing commitment entails reducing the set of consumption bundles available. The optimal commitment problem is to choose the best subset $A \subset B(y)$ to maximize the expected utility of *self-0* given that choices are in the hands of self-1, that is, the allocation is the outcome of a subgame perfect equilibrium. Formally, we maximize $\int \theta U\left(c^A(\theta)\right) + W\left(k^A(\theta)\right) dF(\theta)$ subject to $c^A(\theta), k^A(\theta) \in \arg\max_{(c,k)\in A} \theta U(c) + W(k)$.

Indeed, one can by-pass the set A and view the problem as a mechanism design problem where we seek to maximize *self-0* utility subject to the budget constraint and the incentive constraint that θ is private information of *self-1*. A mechanism design approach that solves for the allocation preferred by *self-0* with resources y uses the revelation principle and sets up the optimal direct truth telling mechanism given y:

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

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

$$c(\theta) + k(\theta) \leq y \text{ for all } \theta \in \Theta \tag{2}$$

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.

The essential tension here is between tailoring consumption to the taste shock and *self-1*'s constant desire for higher current consumption. This tension generates a trade-off between commitment and flexibility from the point of view of *self-0*.

If self-0 is himself not subject to any commitment devices then his optimal allocation solves:

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

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.

For the ensuing games played between *selves* we consider subgame perfect equilibria as our solution concept. For dynamic games with finite horizons this is a very natural choice, and is the approach taken by Laibson (1994) and others. Subgame perfection imposes the desired backward induction reasoning originally emphasized by Strotz, termed 'consistent planning'. It does not however restrict the way selves that are indifferent between two or more choices resolve this indifference.

There are alternatives to subgame perfection that also capture the notion of a 'consistent plan'. These concepts use backward induction but restrict the way points of indifference can be resolved (e.g. they may be resolved in favor of previous selves or independently of past histories, etc.). Peleg and Yaari (1973) and Gul and Pesendorfer (2003) have shown that these alternatives pose problems: there are simple examples where a consistent plan does not exist.

To avoid these existence problems we adopt subgame perfection when studying the time-inconsistent framework. It turns out, however, that the subgame perfect equilibrium characterized in our main results can be implemented with the other notions of 'consistent plans'.

1.2 Time-Consistent Preference

Gul and Pesendorfer (2001, 2002a,b) introduced an axiomatic foundation for preferences for commitment. By introducing the notion of temptation they show that one can capture a desire for commitment without time inconsistent preferences. 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, commitment, 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 p and t over choices a by the relation:

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

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

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

$$
\begin{aligned}
p(c,k;\theta) &= \theta U(c) + W(k) \\
t(c,k;\theta) &= \phi(\theta U(c) + \beta W(k))
\end{aligned}
$$

where the parameter $\phi > 0$ affects the cost of self-control while β captures the degree of temptation towards current consumption.

Following Kreps' (1979) representation we have introduced a taste shock θ. The utility of a set A is then defined by taking the expectation over the different taste shocks:

$$
P(A) = \int [\max_{(c,k)\in A} (\theta U(c) + W(k) + \phi(\theta U(c) + \beta W(k))) - \phi \max_{(c,k)\in A} (\theta U(c) + \beta W(k))] dF(\theta)
$$

The commitment problem can then be stated as maximizing $P(A)$ by choosing a subset $A \subset B(y)$ where $B(y) = \{(c,k) \mid c + k \leq y\}$ is the budget constraint.

As $\phi \to \infty$ the agent has no self-control, yields fully to his temptation and we obtain:

$$
\begin{aligned}
P(A) \equiv & \int [\theta U(c(\theta)) + W(k(\theta))] dF(\theta) \\
& \text{s.t. } (c(\theta), k(\theta)) \in \arg\max_{c,k\in A} \{\theta U(c) + \beta W(k)\}
\end{aligned}
$$

Thus, the problem essentially converge to the time inconsistent hyperbolic model with the difference that we obtain a tie-breaking criteria: indifference is resolved to maximize $\theta U(c) + W(k)$, i.e. to benefit previous 'selves', which is not necessarily the case in the time-inconsistent hyperbolic preference model with the subgame perfect equilibrium concept. This difference turns out to be important with more than two periods. As a consequence, despite the similarities, the hyperbolic model demands special attention and separate analysis to that of the time-consistent with no self-control framework.

2 Optimal Commitment without Self-Control

To gain insight we begin by studying the optimal commitment with only two taste shocks. We then turn to the case with a continuum of taste shocks which is our main focus.

2.1 Two Types

Suppose $\theta_h > \theta_l$, occurring with probabilities p and $1-p$, respectively. Without temptation, $\beta = 1$, there is no disagreement between the principal and the agent and we can implement the ex-ante *first-best allocation* defined by the solution to $\theta U'(c_{fb}(\theta))/W'(k_{fb}(\theta)) = 1$ and $c_{fb}(\theta) + k_{fb}(\theta) = y$. 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. This result relies on the discrete difference in taste shocks and no longer holds when we study a continuum of shocks in Section 2.

For higher levels of temptation, *i.e.* $\beta < \beta^*$, 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.

Proposition 1 *If $\Theta = \{\theta_l, \theta_h\}$, with $\theta_l < \theta_h$, there exists a $\theta_l/\theta_h < \beta^* < 1$ such that for $\beta \in [\beta^*, 1]$ the first-best allocation is implementable. Otherwise,*
(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
In all cases, the optimum can always be attained with the budget constraint holding with equality: $c^*(\theta) + k^*(\theta) = y$ for $\theta = \theta_h, \theta_l$.

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 agent-θ_l holds with equality at the first best allocation. Then for $\beta > \beta^*$ both incentive constraints are slack at the first best allocation and $\beta^* > \theta_l/\theta_h$ follows since,

$$\beta^* \equiv \theta_l \frac{U\left(c_{fb}\left(\theta_h\right)\right) - U\left(c_{fb}\left(\theta_l\right)\right)}{W\left(y - c_{fb}\left(\theta_l\right)\right) - W\left(y - c_{fb}\left(\theta_h\right)\right)} \tag{3}$$

$$> \theta_l \frac{U'\left(c_{fb}\left(\theta_h\right)\right)\left(c_{fb}\left(\theta_h\right) - c_{fb}\left(\theta_l\right)\right)}{W'\left(y - c_{fb}\left(\theta_h\right)\right)\left(c_{fb}\left(\theta_h\right) - c_{fb}\left(\theta_l\right)\right)} = \theta_l \frac{U'\left(c_{fb}\left(\theta_h\right)\right)}{W'\left(y - c_{fb}\left(\theta_h\right)\right)} = \frac{\theta_l}{\theta_h}$$

Now, consider the case where $\beta > \theta_l/\theta_h$ and suppose that $c\left(\theta_h\right) + k\left(\theta_h\right) < y$. Then an increase in $c\left(\theta_h\right)$ and a decrease in $k\left(\theta_h\right)$ that holds $\left(\theta_l/\beta\right) U\left(c\left(\theta_h\right)\right) + U\left(k\left(\theta_h\right)\right)$ unchanged increases $c\left(\theta_h\right) + k\left(\theta_h\right)$ 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 $c\left(\theta_h\right) + k\left(\theta_h\right) = y$ at an optimum. This also shows that separating is optimal in this case, proving part (a). Analogous arguments establish parts (b) and (c).

Finally, $c\left(\theta_l\right) + k\left(\theta_l\right) < y$ cannot be optimal since lowering $c\left(\theta_l\right)$ and raising $k\left(\theta_l\right)$ holding $\theta_l U(c(\theta_l)) + \beta W(k(\theta_l))$ constant would then be feasible. Such a variation does not affect one of the incentive constraints and relaxes the other, yet it increases the objective function since $\theta_l U(c(\theta_l)) + W(k(\theta_l))$ increases. ■

Proposition 1 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. $\theta_l/\theta_h < \beta < \beta^*$, 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 $U = W$ this implies $c = k = y/2$. 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 1 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 $c(\theta) + k(\theta) < y$, is not required for optimality. As discuss below, with more than two types this is not a foregone conclusion.

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 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.

2.2 Continuous Distribution of Types

For the rest of the paper we assume that the distribution of types is represented by a continuous density $f(\theta)$ over the interval $\Theta \equiv [\underline{\theta}, \overline{\theta}]$. Define

$$G(\theta) \equiv F(\theta) + \theta(1-\beta) f(\theta),$$

an expression which will be used frequently below.

We find it convenient to change variables from $(c(\theta), k(\theta))$ to $(u(\theta), w(\theta))$ where $u(\theta) = U(c(\theta))$ and $w(\theta) = W(k(\theta))$ 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,

$$\frac{\theta}{\beta} u(\theta) + w(\theta) = \int_{\underline{\theta}}^{\theta} \frac{1}{\beta} u(\tilde{\theta}) d\tilde{\theta} + \frac{\underline{\theta}}{\beta} u(\underline{\theta}) + w(\underline{\theta}) \tag{4}$$

(see Milgrom and Segal, 2002). It is standard to see that incentive compatibility of (u, w) also requires u to be a non-decreasing function of θ – agents that are more eager for current consumption cannot consume less. Thus, condition (4) 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, 1991).

The principal's problem is thus,

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

subject to (4), $C(u(\theta)) + K(w(\theta)) \leq y$ 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 (4) 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 $w(\underline{\theta})$.

The problem to solve is the following,

$$v_2(y) \equiv \max_{\underline{w}, u(\cdot) \in \Phi} \left\{ \frac{\underline{\theta}}{\beta} u(\underline{\theta}) + \underline{w} + \frac{1}{\beta} \int_{\underline{\theta}}^{\bar{\theta}} (1 - G(\theta)) u(\theta) d\theta \right\} \tag{5}$$

$$K^{-1}(y - C(u(\theta))) + \frac{\theta}{\beta} u(\theta) - \frac{\underline{\theta}}{\beta} u(\underline{\theta}) - \underline{w} - \int_{\underline{\theta}}^{\theta} \frac{1}{\beta} u(\tilde{\theta}) d\tilde{\theta} \geq 0 \text{ for all } \theta \in \Theta \tag{6}$$

where

$$\Phi = \{\underline{w}, u \mid \underline{w} \in W\left(\mathbb{R}^{+}\right),\ u : \Theta \to U(\mathbb{R}^{+}) \text{ and } u \text{ is non-decreasing}\}$$

Note that both the objective function and the left hand side of the constraint are well defined for all $(\underline{w}, u) \in \Phi$. This follows because monotonic functions are integrable (Rudin, 1976, Theorem 6.9, pg. 126) and the product of two integrable functions, in this case $1 - G(\theta) = 1 - F(\theta) - \theta(1-\beta) f(\theta)$ and $u(\theta)$, is integrable (Rudin, 1976, Theorem 6.13, pg. 129).

Note that an allocation $(\underline{w}, u) \in \Phi$ (uniquely) determines an incentive compatible direct mechanism. If condition (6) holds, then this direct mechanism satisfies the budget constraint.

Definition. We say an allocation $(\underline{w}, u)$ is *feasible* if $(\underline{w}, u) \in \Phi$ and (6) holds.

2.3 Bunching at the Top

For any feasible allocation $(\underline{w}, u)$ it is always possible to modify the allocation so as to bunch an upper tail of agents, that is, give them the same bundle. Informally, this can be done by simply removing the bundles at the very top. Those agents whose bundle is removed will now choose the closest bundle available. The new allocation $(\underline{w}, \hat{u})$ is given by $\hat{u}(\theta) = u(\theta)$ for $\theta < \hat{\theta}$ and $\hat{u}(\theta) = u(\hat{\theta})$ for some type $\hat{\theta}$. Bunching the upper tail is always feasible, we now show that it is optimal.

To gain some intuition, note that agents with $\theta \leq \beta\bar{\theta}$ share the ordinal preferences of the principal 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 principal. 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 2 *Define θ_p as the lowest value in Θ such that for $\hat{\theta} \geq \theta_p$:*

$$\int_{\theta}^{\bar{\theta}} \left(1 - G(\tilde{\theta})\right) d\tilde{\theta} \leq 0$$

An optimal allocation $(\underline{w}, u^*)$ has $u^*(\theta) = u^*(\theta_p)$ for $\theta \geq \theta_p$ (i.e. it bunches all agents above θ_p)

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

$$\frac{1}{\beta} \int_{\theta_p}^{\bar{\theta}} (1 - G(\theta))\, u(\theta)\, d\theta.$$

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

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

Note that,

$$\int_{\theta}^{\bar{\theta}} \left(1 - G(\tilde{\theta})\right) d\tilde{\theta} \leq 0,$$

for all $\theta \geq \theta_p$. It follows that it is optimal to set $du = 0$, or equivalently $u(\theta) = u(\theta_p)$, for $\theta \geq \theta_p$. ■

Note that $\int_{\theta}^{\bar{\theta}} \left(1 - G(\tilde{\theta})\right) d\tilde{\theta} \leq 0$ is equivalent to $\mathrm{E}\left[\tilde{\theta} | \tilde{\theta} \geq \theta\right] / \theta \leq 1/\beta$ and that $\theta_p < \beta\bar{\theta}$. With two types Proposition 1 showed that bunching is strictly optimal whenever $\theta_h/\theta_l < 1/\beta$. Proposition 2 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 $\theta_p = \underline{\theta}$ then all agents are pooled at single point where $U(c)+W(k)$ is maximized subject to $c + k = y$. For the rest of the paper we study the case where $\theta_p > \underline{\theta}$.

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 might be optimal to bunch all agents depending on the Pareto parameter.

2.4 Minimum Saving Policies

To obtain a simple and full characterization of the optimal allocation for $\theta \leq \theta_p$ we impose the following condition on the distribution F and β.

Assumption A: $G(\theta) \equiv (1-\beta)\,\theta f(\theta) + F(\theta)$ *is increasing for all* $\theta \leq \theta_p$.

When the density f is differentiable assumption A is equivalent to,

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

which places a lower bound on the elasticity of the density f. The lower bound is negative and continuously 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.

Define $c^{flex}(\theta), k^{flex}(\theta)$ to be the unconstrained optimum for agent-θ, that is the allocation that is achieved when individuals are given full flexibility:

$$\left(c^{flex}(\theta), k^{flex}(\theta)\right) \equiv \arg\max_{c,k} \left\{\frac{\theta}{\beta} U(c) + W(k)\right\}$$
$$\text{s.t. } c + k \leq y$$

and let $u^{flex}(\theta) \equiv U\left(c^{flex}(\theta)\right)$ and $w^{flex}(\theta) \equiv W\left(k^{flex}(\theta)\right)$. 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 the point $\left(c^{flex}(\theta_p), k^{flex}(\theta_p)\right)$ given by the ex-post unconstrained optimum of the θ_p-agent.

Let the proposed allocation $(\underline{w}^*, u^*)$ be given by $\underline{w}^* = w^{flex}(\underline{\theta})$ and

$$u^*(\theta) = \begin{cases} u^{flex}(\theta) & \text{for } \theta < \theta_p \\ u^{flex}(\theta_p) & \text{for } \theta \geq \theta_p \end{cases}$$

This translates to $(c^*(\theta), k^*(\theta)) = \left(c^{flex}(\theta), k^{flex}(\theta)\right)$ for $\theta < \theta_p$ and $(c^*(\theta), k^*(\theta)) = \left(c^{flex}(\theta_p), k^{flex}(\theta_p)\right)$ for $\theta \geq \theta_p$.At this allocation, the agents have full flexibility for shocks smaller than θ_p and are bunched at θ_p for higher shocks. It is an allocation that corresponds to a minimum savings rule. We now proceed to show that this allocation is optimal.

Define the Lagrangian function as,

$$
\begin{aligned}
L(\underline{w}, u|\Lambda) \equiv & \frac{\underline{\theta}}{\beta} u(\underline{\theta}) + \underline{w} + \frac{1}{\beta}\int_{\underline{\theta}}^{\bar{\theta}} (1 - G(\theta))\, u(\theta)\, d\theta \\
& + \int_{\underline{\theta}}^{\bar{\theta}} \left(K^{-1}(y - C(u(\theta))) + \frac{\theta}{\beta} u(\theta) - \left(\frac{\underline{\theta}}{\beta} u(\underline{\theta}) + \underline{w}\right) - \int_{\underline{\theta}}^{\theta} \frac{1}{\beta} u(\tilde{\theta}) d\tilde{\theta} \right) d\Lambda(\theta)
\end{aligned}
$$

where the function Λ is the Lagrange multiplier associated with the incentive compatibility constraint. The function Λ is required to be left-continuous and non-decreasing, so that it defines a measure μ (on the Borel $\sigma-$algebra) by letting $\mu([a, b)) = \Lambda(b) - \Lambda(a)$ and the integral is a Lebesgue integral with respect to the measure μ. Without loss of generality set $\Lambda\left(\bar{\theta}\right) = 1$.

Intuitively, the Lagrange multiplier Λ can be thought of as a cumulative distribution function[2] that determines the importance of the resource constraints in the maximization. If Λ happens to be representable by a 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 such 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.

[2] 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.

Integrating the Lagrangian by parts yields:

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

Note that we do not need to incorporate the monotonicity condition explicitly. Instead, we work directly with Φ which incorporates the monotonicity condition. The next lemma shows that the appropriate first-order condition are necessary and sufficient for optimality.

Lemma of Optimality. *(a) If an allocation* $\left(\underline{w_0},u_0\right)\in\Phi$ *is optimal with* u_0 *is continuous then there exists a non-decreasing* Λ_0 *such that the following first-order conditions in terms of Gateaux differentials*[3]*:*

$$\partial L\left(\underline{w_0},u_0;\underline{w_0},u_0|\Lambda_0\right)=0 \tag{7}$$

$$\partial L\left(\underline{w_0},u_0;h_{\underline{w}},h_u|\Lambda_0\right)\leq 0 \tag{8}$$

hold for all $\left(h_{\underline{w}},h_u\right)\in\Phi$ *and* h_u *continuous. (b) Conversely, if first-order conditions (7) and (8) holds for some* Λ_0 *for all for all* $\left(h_{\underline{w}},h_u\right)\in\Phi$ *then* $\left(u_0,w_0\right)$ *is optimal.*
Proof. In the appendix. ■

The proof of the lemma proceeds by in three steps, drawing heavily on the methods developed in Luenberger (1969). We first verifying the conditions required to characterize optimality in terms of the maximization of the Lagrangian (cite Luenberger). We then show that because the Lagrangian is the integrals of concave functions it is sufficiently differentiable. Finally, because the Lagrangian is convex we show that it is maximized if and only the stated first-order conditions hold by modifying a result

[3] *Given a function* $T:\Omega\rightarrow Y$*, where* $\Omega\subset X$ *and* X *and* Y *are normed spaces. If for* $x\in\Omega$ *and* $h\in X$ *the limit*

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

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

in Luenberger.

The Gateaux differential at the proposed allocation $(\underline{w}^*, u^*)$ is given by:

$$\partial L\left(\underline{w}, u; h_{\underline{w}}, h_u | \Lambda\right) = \left(\frac{\underline{\theta}}{\beta} h_u\left(\underline{\theta}\right) + h_{\underline{w}}\right) \Lambda\left(\underline{\theta}\right) + \frac{1}{\beta} \int_{\underline{\theta}}^{\bar{\theta}} \left(\Lambda\left(\theta\right) - G\left(\theta\right)\right) h_u\left(\theta\right) d\theta + \frac{\theta_p}{\beta} \int_{\theta_p}^{\bar{\theta}} \left(\frac{\theta}{\theta_p} - 1\right) h_u d\Lambda\left(\theta\right) \tag{9}$$

for all $(h_{\underline{w}}, h_u) \in \Phi$.

Proposition 3 *The proposed allocation $(\underline{w}^*, u^*)$ is optimal if and only if assumption A holds.*

Proof. *Necessity.* Since $(\underline{w}^*, u^*)$ is optimal then there should exist a non-decreasing Λ^* such that first-order conditions (7) and (8) hold. We will show that if assumption A does not hold then the first-order conditions require a decreasing Λ^*, a contradiction.

Condition (8) with $h_u = 0$ requires that $\Lambda^*\left(\underline{\theta}\right) = 0$ since h_w is unrestricted. Using $\Lambda^*\left(\underline{\theta}\right) = 0$ and integrating (9) by parts leads to (Theorem 6.20 in Rudin, 1976, guarantees this step since h_u is continuous):

$$\partial L\left(\underline{w}^*, u^*; h_{\underline{w}}, h_u | \Lambda^*\right) = \gamma\left(\underline{\theta}\right) h_u\left(\underline{\theta}\right) + \int_{\underline{\theta}}^{\bar{\theta}} \gamma\left(\theta\right) dh_u\left(\theta\right), \tag{10}$$

where,

$$\gamma\left(\theta\right) \equiv \frac{1}{\beta} \int_{\theta}^{\bar{\theta}} [\Lambda^*(\tilde{\theta}) - G(\tilde{\theta})] d\tilde{\theta} + \frac{\theta_p}{\beta} \int_{\max\{\theta, \theta_p\}}^{\bar{\theta}} \left(\frac{\tilde{\theta}}{\theta_p} - 1\right) d\Lambda^*(\tilde{\theta}) \tag{11}$$

It follows that condition (8) implies that $\gamma\left(\theta\right) \leq 0$ for all $\theta \in \Theta$. Suppose that there is a θ_1 such that $\gamma\left(\theta_1\right) > 0$ then we argue that there is an interval $[\theta_0, \theta_1]$ such that $\gamma\left(\theta\right) > 0$ for all $\theta \in [\theta_0, \theta_1]$. This follows since in the definition of γ the first term is continuous and the second term is non-increasing. But such an interval leads to a contradiction with (8) for any continuous non-decreasing function h_u that is strictly increasing within $[\theta_0, \theta_1]$ and constant for $\Theta - [\theta_0, \theta_1]$.

Given $\gamma\left(\theta\right) \leq 0$ for all $\theta \in \Theta$, (7) implies that $\gamma\left(\theta\right) = 0$ for $\theta \in [\underline{\theta}, \theta_p]$, i.e. wherever u^* is strictly increasing. It follows that,

$$\Lambda^*\left(\theta\right) = G\left(\theta\right),$$

for all $\theta \in (\underline{\theta}, \theta_p]$. The proposed allocation $(\underline{w}^*, u^*)$ 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 in this region. It follows that assumption A is necessary for the proposed solution $(\underline{w}^*, u^*)$ to be optimal.

Sufficiency. We now prove sufficiency by showing that there exists a non-decreasing multiplier Λ^* such that the proposed $(\underline{w}^*, u^*)$ satisfies the first-order conditions (7) and (8) for all $(h_{\underline{w}}, h_u) \in \Phi$. We've specified Λ^* for $(\underline{\theta}, \theta_p]$ that is consistent with these first-order conditions so we only need to specify Λ^* for $(\theta_p, \bar{\theta}]$. We will show that $\Lambda^*(\theta) = 1$ for $\theta \in \left(\theta_p, \bar{\theta}\right]$ meets the requirements.

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,

$$\lim_{\theta \downarrow \theta_p} \Lambda^*(\theta) - \Lambda^*(\theta_p) = 1 - G(\theta_p) \geq 0,$$

which follows from the definition of θ_p. To see this, note that if $\theta_p = \underline{\theta}$ the result is immediate since then Λ^* would jump from 0 to 1 at $\underline{\theta}$. Otherwise, notice that, by definition, θ_p is the lowest $\hat{\theta}$ such that $\gamma(\theta) \leq 0$ for all $\theta \geq \hat{\theta}$, which implies that $\gamma'(\theta_p) = G(\theta_p) - 1 \leq 0$.

Given the proposed allocation $(\underline{w}^*, u^*)$ and the constructed Lagrange multiplier Λ^* imply that $\gamma \leq 0$ and that $\gamma = 0$ wherever u^* is increasing. The Gateaux differential is also given by (10), integration by parts is warranted for non-decreasing h_u given the particular Λ^* constructed. It follows that the first-order conditions (7) and (8) are satisfied. ■

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

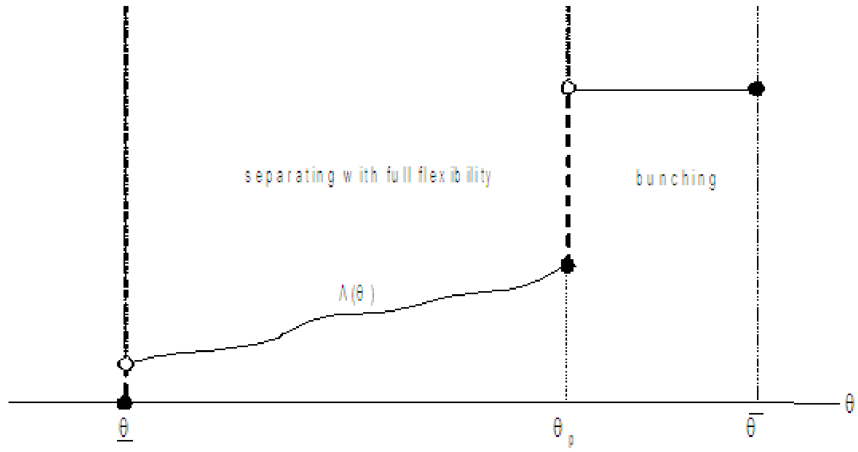

Figure 2: The Lagrange multiplier $\Lambda^*(\theta)$

Proposition 3 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 4 *The bunching point θ_p increases with β. The minimum savings requirement, $s_{\min} = y - C(u(\theta_p))$, 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

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

and that θ_p, when interior, solves,

$$\frac{\theta_p}{\beta} = \mathrm{E}[\theta \mid \theta \geq \theta_p].$$

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

If assumption A does not hold then no-minimum savings rules are optimal. Notice that this is not directly implied by the necessity part of proposition 3, given that this proposition analyzes the optimality of the proposed allocation which is a particular minimum savings rule. In the next proposition we show that there are no minimum savings rules that are optimal without assumption A.

Proposition 5 *If assumption A does not hold, then there are no optimal minimum savings rules.*

Proof. Let $[a, b] \subset [\underline{\theta}, \theta_p)$, with $a < b$, be an interval where G is strictly decreasing. Let $(\underline{\hat{w}}, \hat{u})\,|_{x_p}$ be a minimum savings allocation indexed by x_p:

$$\begin{aligned} \hat{u}(\theta) &= \begin{cases} u^{flex}(\theta) & \text{for } \theta < x_p \\ u^{flex}(x) & \text{for } \theta \geq x_p \end{cases} \\ \underline{\hat{w}} &= w^{flex}(\underline{\theta}) \end{aligned}$$

where x_p denotes the proposed bunching point. Towards a contradiction suppose that $(\underline{\hat{w}}, \hat{u})\,|_{x_p}$ is optimal for some x_p.

We now follow the proof of proposition 3. Let $\hat{\Lambda}$ be the associated Lagrange multiplier for budget constraint. And let $\hat{\gamma}$ be described in the same way as equation (11) but with x_p in place of θ_p:

$$\hat{\gamma}(\theta|x_p) = \frac{1}{\beta} \int_{\theta}^{\bar{\theta}} \left(\hat{\Lambda}\left(\tilde{\theta}\right) - G\left(\tilde{\theta}\right) \right) d\tilde{\theta} + \frac{x_p}{\beta} \int_{\max\{x, \theta\}}^{\bar{\theta}} \left(\frac{\tilde{\theta}}{x_p} - 1 \right) d\hat{\Lambda}\left(\tilde{\theta}\right) \tag{12}$$

where just as before, a necessary condition for optimality is $\hat{\gamma}(\theta|x_p) \leq 0$ for all $\theta \in \Theta$.

Then $x \leq a$, otherwise, the associated multiplier $\hat{\Lambda}(\theta)$, which is equal to G in the separating region, would be decreasing for $\theta \in [a, \min\{x, b\}]$. Integrating by parts the second term of equation (12) we obtain:

$$\hat{\gamma}(x_p|x_p) = \frac{1}{\beta} \int_{x_p}^{\bar{\theta}} [1 - G(\tilde{\theta})] d\tilde{\theta}$$

which is independent of the choice of the multiplier $\hat{\Lambda}$. But for any $x_p \leq a < \theta_p$, we have then that $\hat{\gamma}(x_p|x_p) > 0$ by the definition of θ_p. Hence no minimum savings rule is optimal. ■

2.5 Drilling

In this subsection we study cases where assumption A does not hold and show that the allocation described in Proposition 3 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 some 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.

Let $\theta_{ind} \in [\theta_a, \theta_b]$ be the agent type that obtains the same utility from reporting θ_a or θ_b. We find it more convenient to state the next result in terms of $c(\theta)$ and $k(\theta)$.

Proposition 6 *Suppose an allocation $(c(\theta), k(\theta))$ satisfies incentive compatibility (1) and the budget constraint (2) and has $c(\theta) = c^{flex}(\theta)$ and $k(\theta) = k^{flex}(\theta)$ for $\theta \in [\theta_a, \theta_b]$, where $\theta_b \leq \theta_p$. Then if $G(\theta)$ is decreasing on $[\theta_a, \theta_b]$ the allocation $(\tilde{c}(\theta), \tilde{k}(\theta))$ defined below increases the objective function, remains incentive compatible (1) and satisfies the budget constraint (2):*

$$\tilde{c}(\theta), \tilde{k}(\theta) = \begin{cases} c(\theta), k(\theta) & ;\text{ for } \theta \notin [\theta_a, \theta_b] \\ c(\theta_a), k(\theta_a) & ;\text{ for } \theta \in (\theta_a, \theta_{ind}) \\ c(\theta_b), k(\theta_b) & ;\text{ for } \theta \in [\theta_{ind}, \theta_b) \end{cases}$$

Proof. In the appendix. ■

Proposition 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.

3 Stochastic Temptation

We now consider now the case where the level of temptation, measured by β, is random. Variation 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 an agent knew his temptation level β at time 0, he would tailor the mechanism to this particular β, and the analysis would proceed as in the previous sections. This then reduces to a comparative static exercise on β. To explore other possibilities we assume that the temptation level for $t = 1$ is random and realized at $t = 1$ (together with θ) so that it is unknown to the time-0 self. We assume that both β and θ are realized from a continuous joint distribution. We do not require independence of θ and β for our results. We continue to assume that $\beta \leq 1$.

We will make use of a result that shows that indifference is rare. 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). 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 $u_1 \leq u_2$. Points at which there are more than a single element in $M(x; A)$ represent upward 'jumps'. As with monotonic functions, there can be at most a countable number of such 'jumps'. This is the logic behind the next result.

Lemma (Countable Indifference). *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. Let $M^u(x;A) = \{u : \exists w \ (u,w) \in M(x;A)\}$ and let D be the set of points in X where $M(x;A)$ has more than one element, so that $\inf M^u(x;A) \neq \sup M^u(x;A)$ for all $x \in D$. For each $x \in D$ we can choose a rational number r_x such that $\inf M^u(x;A) < r_x < \sup M^u(x;A)$. Given the stated monotonicity property for M^u it follows that r_x are strictly increasing in x, so that in particular for each $x \in D$ we have a distinct rational number r_x. Thus, there is a 1-1 correspondence between the set D and a sub-set of the rational numbers. Since the rational numbers are countable the result follows. ■

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 $x \equiv \theta/\beta$. The allocation however 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 Countable Indifference lemma shows that the set of x for which agents are indifferent is of measure zero. As a consequence, allowing the allocation to depend on θ or β independently, in addition to x, would not improve the objective function. Without loss in optimality, we focus then in allocations that are functions of x only.

The objective function can be written as

$$\mathrm{E}\left[\theta u(x) + w(x)\right] = \mathrm{E}\left[\mathrm{E}\left[\theta u(x) + w(x)|\, x\right]\right] = \int \left[n(x)\, u(x) + w(x)\right] \tilde{f}(x)\, dx$$

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

Let $\tilde{G}(x) = \tilde{F}(x) - (n(x) - x)\tilde{f}(x)$ and define $x_p \in X$ as the lowest value such that for $\hat{x} \geq x_p$, $\hat{x} \in X$,

$$\int_x^{\bar{x}} (1 - \tilde{G}(x))dx \leq 0$$

Let u^{flex}, w^{flex} be the values indexed by x that maximize $\{xu + w\}$ subject to the resource constraint $y \geq C(u) + K(w)$. Let the proposed allocation be given by

$\underline{w} = w^{flex}(\underline{x})$ and $u^*(x) = u^{flex}(x)$ if $x < x_p$ and $u^*(x) = u^{flex}(x_p)$ if $x \geq x_p$.

We modify our previous assumption A in the following way.

Assumption $\tilde{\mathbf{A}}$. $\tilde{G}(x)$ is increasing in x *for* $x \in [\underline{x}, x_p]$

The incentive compatibility constraint is thus

$$xu(x) + w(x) = \underline{x}u(\underline{x}) + w(\underline{x}) + \int_{\underline{x}}^{x} u(x)\,dx$$

plus the standard monotonicity restriction on $u(x)$. Substituting this last equation into the objective function, we can write the principal's problem as

$$\max_{\underline{w}, u(\cdot) \in \tilde{\Phi}} \left\{ \underline{x}u(\underline{x}) + w(\underline{x}) + \int_{\underline{x}}^{\bar{x}} (1 - \tilde{G}(x))u(x)\,dx \right\} \tag{13}$$

subject to

$$K^{-1}(y - C(u(x))) + xu(x) - \underline{x}u(\underline{x}) - w(\underline{x}) - \int_{\underline{x}}^{x} u(x)\,dx \geq 0 \text{ for all } x \in X$$

where

$$\tilde{\Phi} = \left\{\underline{w}, u(\cdot) \,|\underline{w} \in W\left(\mathbb{R}^+\right), u : X \to \mathbb{R} \text{ and } u \text{ is non-decreasing}\right\}$$

The next proposition states that minimum savings rules are optimal under assumption Ã.

Proposition 7 *The allocation $(\underline{w}^*, u^*)$ is optimal if and only if assumption $\tilde{A}$ holds.*

Proof. The proof proceeds along the same lines as the proof of Proposition 3. Note that problem (13) is equivalent to problem (5), the only thing to check is that the new Lagrange multiplier has a non-negative jump at zero. This jump is equal in sign to $(\underline{x} - n(\underline{x}))$, which non-negative given the assumption that $\beta \leq 1$. ■

So, when temptation is stochastic, under assumption $\tilde{A}$, minimum savings rule are optimal for an individual. Suppose now that there is a population of individuals such that (β, θ) is i.i.d. across agents and private information to each agent. Then, it is easy to see that the allocation described in the above proposition is the optimal allocation for every agent in the population. A principal will offer a unique minimum savings rule for the whole population.

4 Optimal Commitment with Self-Control

In section 1.2 we introduced a preference for commitment specification with self control for an agent subject to taste shocks. In this section, we study the optimal commitment device of an agent with such preferences. We will show now that minimum saving rules are optimal for this preferences specification as well, under similar conditions as before.

In keeping with the previous section we introduce self-control while consider the general case where the level of temptation, measured by β, and the degree of self-control ,ϕ, are random and realized together with θ, that is, at the time temptation is being experienced. We assume θ, β and ϕ are realized together from a continuous distribution – we do not require independence of θ, ϕ and β for our results. Assume that (θ, β, ϕ) has rectangular support $\left[\underline{\theta}, \bar{\theta}\right] \times [\underline{\beta}, \bar{\beta}] \times [\underline{\phi}, \bar{\phi}] \subset [0, \infty) \times (0, 1] \times [0, \infty)$.

After a few manipulations the objective function can then be written as:

$$E\left\{(1+\beta\phi) \max_{v,\omega \in A} \left[(\theta/\hat{\beta})v + \omega\right] - \phi\beta \max_{v,\omega \in A}[(\theta/\beta)v + \omega]\right\}$$

where $\hat{\beta} \equiv (1+\beta\phi) / (1+\phi)$. Define the random variables $\hat{z}$ and z by $\hat{z} \equiv \theta/\hat{\beta}$ and $z \equiv \theta/\beta$. Let $\hat{\Theta}$ be the union of the supports for z and $\hat{z}$, so that $\hat{\Theta} \equiv [\underline{x}, \bar{x}] \equiv [\underline{\theta}(1+\underline{\phi})/(1+\overline{\beta}\underline{\phi}),\ \bar{\theta}/\underline{\beta}]$. Consider an allocation over $\hat{\Theta}$ given by a pair of functions $u : \hat{\Theta} \to U\left(R_+\right)$ and $w : \hat{\Theta} \to W\left(R_+\right)$. Then the objective function is equivalent to

$$\mathrm{E}\left\{(1+\beta\phi)\left(\hat{z}u(\hat{z}) + w(\hat{z}) - \beta\phi(zu(z) + w(z))\right\}\right.$$

subject to,

$$(u(x), w(x)) \in \arg\max_{v,\omega \in A}[xv + \omega] \tag{14}$$

for all $x \in \hat{\Theta}$.

Using the law of iterated expectations, we can write the objective function as:

$$\int_{\hat{\Theta}} \alpha(\hat{z})(\hat{z}u(\hat{z}) + w(\hat{z}))h(\hat{z}) - \int_{\hat{\Theta}} \kappa(z)(zu(z) + w(z))f(z)$$

where $h(\hat{z})$ and $f(z)$ are the densities of $\hat{z}$ and z, respectively and $\alpha(x) \equiv \mathrm{E}[(1+\beta\phi) \mid \hat{z} = x]$ and $\kappa(x) \equiv \mathrm{E}[\beta\phi \mid z = x]$.

We thus obtain a representation of the objective function as,

$$\int_{\hat{\Theta}} (xu(x) + w(x))\hat{g}(x)\, dx$$

where the density $\hat{g}(x) \equiv \alpha(x) h(x) - \kappa(x) f(x)$, can be negative or positive, and defines a signed measure over the state space $\hat{\Theta}$. This is the representation of preferences derived axiomatically by Dekel, Lipman and Rustichini (2001). Their general framework allows preferences for commitment and flexibility.

Note that $h(z) = 0$ for $z > \overline{\theta}(1+\bar{\phi})/(1+\underline{\beta}\bar{\phi})$ and that $f(z) = 0$ for $z < \underline{\theta}/\overline{\beta}$, so $\hat{g}(x) < 0$ for $z \in (\overline{\theta}(1+\bar{\phi})/(1+\underline{\beta}\bar{\phi}), \bar{x}]$. Define the cumulative measure function $\hat{G}(x) = \int^x \hat{g}(z)\, dz$ where $\hat{G}(\bar{x}) = 1$.

From (14) we need to impose that:

$$xu(x) + w(x) \geq xu(x') + w(x') \text{ for all } x, x' \in \hat{\Theta} \tag{15}$$

or equivalently :

$$xu(x) + w(x) = \underline{x}u(\underline{x}) + \underline{w} + \int_{\underline{x}}^{x} u(x')\, dx' \tag{16}$$

with the associated monotonicity constraint.

We can now substitute (16) into the objective function and the resource constraints. Integrating by parts the objective function, and using $\hat{G}(\bar{x}) = 1$, we can rewrite the problem as

$$\max_{(\underline{w}, u(\cdot)) \in \hat{\Phi}} \left\{ \underline{x}u(\underline{x}) + \underline{w} + \int_{\hat{\Theta}} [1 - \hat{G}(x)]u(x)\, dx \right\} \tag{17}$$

subject to the resource constraints:

$$K^{-1}\left(y - C\left(u\left(x\right)\right)\right) + xu\left(x\right) - \underline{x}u\left(\underline{x}\right) - \underline{w} - \int_{\underline{x}}^{x} u(x')dx' \geq 0 \tag{18}$$

and where $\hat{\Phi} \equiv \left\{\underline{w}, u \mid \underline{w} \in W\left(\mathbb{R}^{+}\right), u : \hat{\Theta} \to \mathbb{R} \text{ and } u \text{ is non-decreasing}\right\}$

Note that the problem stated in (17) is equivalent to the problem (5) stated in Section 2. The following propositions follow directly from the results in that section.

Proposition 8 *Let x_p be the lowest value in $\hat{\Theta}$ such that for all $\hat{x} \geq x_p$*

$$\mathrm{E}\left[1 - \hat{G}\left(x\right) \ \mid x \geq \hat{x}\right] \leq 0$$

An optimal allocation, $\left(\underline{w}^{*}, u^{*}\right)$ has $u^{*}\left(x\right) = u^{*}\left(x_p\right)$ for $x \geq x_p$.

Proof. Equivalent to proof of Proposition 2. ■

Define u^{flex}, w^{flex} be the values that maximize $\{xu + w\}$ subject to the resource constraint $y \geq C\left(u\right) + K\left(w\right)$. Let the proposed allocation be given by $\underline{w} = w^{flex}\left(\underline{x}\right)$ and $u^{*}\left(x\right) = u^{flex}\left(x\right)$ if $x < x_p$ and $u^{*}\left(x\right) = u^{flex}\left(x_p\right)$ if $x \geq x_p$. We introduce the following assumption analogous to that of assumption A

Assumption B: *$\hat{G}\left(x\right)$ is increasing for all $x \leq x_p$.*

The next proposition states that minimum savings rules are optimal under assumption B.

Proposition 9 *The allocation $\left(\underline{w}^{*}, u^{*}\right)$ is optimal if and only if assumption B holds.*

Proof. The proof proceeds along the same lines as the proof of Proposition 3, except that Λ does not jump at $\underline{x}$ (because $\hat{G}\left(x\right)$ is zero at $\underline{x}$). ■

In the case where β and ϕ are non-stochastic, we can establish a connection between assumptions A and B by showing that B is a weaker requirement.

Lemma. If β and ϕ are deterministic and the condition for assumption A holds for $[\underline{\theta}, \theta'_p\hat{\beta}/\beta]$, then the condition for assumption B holds for $[\underline{\theta}, \theta'_p]$.
Proof. In the appendix. ■

5 Arbitrary Finite Horizons

We now show that our two-period analysis extends to arbitrary finite horizons. Our choice of confining our attention to finite horizons is motivated by the fact that with infinite horizons the time-inconsistent framework may otherwise yield reputational equilibria. The equilibria 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.

Indeed, Krusell, Kuruscu and Smith (2001) have shown that the time-consistent temptation framework we are working with may also be problematic with an infinite horizon. Even with full flexibility, with infinite horizons, multiple solutions may exist to the recursive dynamic program that describes the agent's problem. To avoid having deciding which solution to focus on we work with finite horizons where these problems do not arise.

5.1 Time-Consistent Preferences

Extending the two-period results with time-consistent preferences is straightforward and leads naturally to a recursive dynamic formulation. We treat separately the case with and without self-control and show that, in both cases, the resulting Bellman equation has the same structure as the two-period problem studied in previous sections, implying that the same analysis can be obtained.

We begin with the case without self-control. The problem with $T \geq 2$ periods remaining can be written as:

$$v_T(y) = \max \int \theta U(c(\theta)) + v_{T-1}(k(\theta))\, dF(\theta)$$

$$\theta U(c(\theta)) + \beta v_{T-1}(k(\theta)) \geq \theta U(c(\theta')) + \beta v_{T-1}(k(\theta')) \text{ for all } \theta, \theta' \in \Theta$$

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

where $v_1 \equiv W$. Obviously, the case with random β can be treated similarly.

With self-control we start by writing the problem as,

$$v_T(y) = \max_{A \subset B(y)} \mathrm{E}\{(1+\phi)\max_{c,k \in A}[zU(c) + \hat{\beta} v_{T-1}(k)] - \phi \max_{c,k \in A}[zU(c) + \beta v_{T-1}(k)]\}$$

where $B(y) = \{(c,k) \mid c + k \leq y\}$. This dynamic program maximizes over the subset A of the budget constraint. Given this set the agent can be seen as maximizing $zU(c) + \hat{\beta} v_{T-1}(k)$, where $\hat{\beta} \equiv (1+\beta\phi)/(1+\phi)$, and suffers from not maximizing his temptation utility $zU(c) + \beta v_{T-1}(k)$. In both cases $v_{T-1}(k)$ adequately summarizes the value attached to the continuation game with resources k. Since $\hat{\beta} < 1$ the agent can be seen as yielding somewhat to the temptation of higher current consumption. Since $\hat{\beta} > \beta$ the agent is exerting some self-control.

Using the same arguments as in section 4 it is easy to re-write this problem in terms of an allocation $c, k : \hat{\Theta} \to \mathbb{R}$

$$v_T(y_T) = \max \int_{\hat{\Theta}} [\theta U(c(\theta)) + v_{T-1}(k(\theta))] \, n(\theta) \, d\theta$$

$$\theta U(c(\theta)) + v_{T-1}(k(\theta)) \geq \theta U(c(\theta')) + v_{T-1}(k(\theta')) \text{ for all } \theta, \theta' \in \hat{\Theta}$$

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

It is immediate that v_T is increasing, concave and continuously differentiable if v_{T-1} has these properties. Since $v_1 = W$ has these properties by assumption it follows by assumption that v_T has them, for all T. It follows that the analysis from previous sections immediately applies with v_{t-1} in the role of W.

For any horizon T these problems has exactly the same structure as their two-period counterparts analyzed previously, with the function v_{T-1} playing the role of the utility function W. We only required W to be increasing and concave and since v_{T-1} has these properties all the previous analysis apply immediately implying the following proposition.

Proposition 10 *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.*

Another property of the optimal allocation identified in Proposition 3 is worth mentioning. Suppose agents can save, but not borrow, privately behind the principal's back at the same rate of return as the principal, 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 10 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 10 agents that currently have no private savings would never find it optimal to accumulate private savings. To see this, first note that by Proposition 10 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.

In Proposition 10 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 $k(\theta, y) = \tilde{k}(\theta) y$. It follows that the optimal mechanism imposes a minimum saving *rate* for each period that is independent of y_t.

Proposition 11 *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.*

5.2 Time-Inconsistent Preferences

It is tempting to conclude that in the in the time-inconsistent hyperbolic model one can simply study the multi-period dynamic problem with no-self control time-consistent preferences described above. After all, the two models were equivalent in the case of two periods.

This is misleading for the following reason. In the hyperbolic model we have T 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 4, 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.

Our aim is to take a general mechanism that generates some equilibrium and show that in the case of a continuos distribution of the taste shock we can use a simpler direct mechanism that has a truth telling equilibrium with the same outcome.

We first describe the general mechanism, define an equilibrium there and derive a properties of equilibria with a continuous distribution of shocks. We then describe the simpler direct mechanism and prove the result that these can be used without loss in generality.

5.2.1 General Mechanism

Time runs from $t = 0, 1, ..., T$ and for each t there is a player which we label "self-t". After observing the current taste shock self-t sends a message m_t to the principal from a set of allowable messages $M_t(m^{t-1})$, which may depend on the history of past messages summarized by $m^{t-1} = (m_0, m_1, ..., m_{t-1})$. Let M^t be the set of all allowable histories m^t, i.e. $m_s \in M_s(m^{s-1})$ for $s \leq t$. The principal pays out utility at time

t as a function of the history of reports: $u_t(m^t) : M^t \to U(\mathbb{R}_+)$. Let $M = \{M_t\}_{t=1}^T$ and $u = \{u_t\}_{t=1}^T$ collect the message and payoffs, then a mechanism is summarized by (M, u) and defines a game between the selves.

Note however that, in this game, self-t observes the history of past shocks and the history of messages. This is only natural since indeed, the various selves are different time incarnations of the same person.

An application of Revelation Principle ideas in this context would lead us to focus on mechanisms that have $M_t = \Theta^t$, self-t reports on the history of taste shocks, since this is the primitive information (i.e. type) that self-t has. The Revelation Principle would not justify focusing on $M_t = \Theta$, self-t reports his current taste shock only, as it does when there is a single agent.

Indeed, with finite taste shocks it is easy to construct examples where having the message space Θ^t is strictly preferred to having Θ. However, with a continuous distribution of taste shocks we will show that these advantages disappear and $M_t = \Theta$ as a message space can be used without loss in optimality. It turns out that making the argument for Θ over an arbitrary space M_t is just as simple as making it against Θ^t. Thus, we do not as a preliminary step appeal informally to the revelation principle and continue to work below with general mechanisms (u, M).

Definition. A reporting strategy r_t for self-t is a mapping from the history of shocks and messages to the current message, i.e. for each $m^{t-1} \in M^{t-1}$ we have $r_t(\cdot, m^{t-1}) : \Theta^t \to M_t(m^{t-1})$.

Let $r = \{r_t\}_{t=0}^T$ collect the strategy profile of all selves.

We introduce notation to present the equilibrium definition recursively. Given any history of messages m^s up to time s a strategy profile r determines the message reports that follow as a function of the shocks. Let $r^{t,s}(\theta^t, m^s)$ represent the messages reported in periods $s, s+1, ..., t$ if the realization of shocks up to t is given by θ^t defined by the recursion:

$$r^{t+1,s}\left(\theta^{t+1}, m^{s-1}\right) = \left(r^{t,s}\left(\theta^t, m^{s-1}\right), r_{t+1}\left(\theta^{t+1}, r^{t,s}\left(\theta^t, m^{s-1}\right)\right)\right)$$

with $r^{s,s}(\theta^s, m^{s-1}) = r_s(\theta^s, m^{s-1})$.

Define the *expected continuation utility* in period t by:

$$w_t\left(\theta^t, m^t; r\right) \equiv \sum_{s=1}^{T-t} \int_{\Theta^s} \tilde{\theta}_{t+s} u_{t+s}\left(m^t, r^{t+s,t+1}((\theta^t, \tilde{\theta}_{t+1}^{t+s}), m^t)\right) f(\tilde{\theta}_{t+1}) \cdots f(\tilde{\theta}_{t+s}) d\tilde{\theta}_{t+1}^{t+s}$$

or in terms of the recursion,

$$\begin{aligned} w_t\left(\theta^t, m^t; r\right) \;&=\; \int_{\Theta} [\; \tilde{\theta}_{t+1} u_{t+1}\left(m^t, r_{t+1}((\theta^t, \tilde{\theta}_{t+1}), m^t)\right) \\ &\qquad + w_{t+1}\left((\theta^t, \tilde{\theta}_{t+1}), (m^t, r_{t+1}((\tilde{\theta}^t, \tilde{\theta}_{t+1}), m^t))\right)] f(\tilde{\theta}_{t+1}) d\tilde{\theta}_{t+1} \end{aligned}$$

with $w_T\left(\theta^T, m^T; r\right) = 0$.

Note that continuation utility $w_t\left(\theta^t, m^t; r\right)$ may depend on the sequence of past values of the taste shocks θ^t. Since taste shocks are independently distributed over time, the dependence on θ^t is result of future reporting strategies, r_{t+s} for $s \geq 1$, dependence on θ^t. Thus, future reports may provide information on past shocks. This dependence of future reports has the potential to induce good behavior from previous selves.

For example, imagine a mechanism where that requests a message on the whole sequence of shocks (or simply on the current shock and on whether or not all previous selves have reported truthfully previous shocks). If future selves could be induced to "tell-on" previous selves then we could punish the mis-reporting and induce truth-telling more easily.

Given a mechanism (M, u) a strategy profile r is a sub-game perfect equilibrium of the game if and only if:

$$\begin{aligned} &\theta_t u_t\left(m^{t-1}, r_t\left(\theta^t, m^{t-1}\right)\right) + \beta w_t\left(\theta^t, (m^{t-1}, r_t\left(\theta^t, m^{t-1}\right)); r\right) \\ &\qquad\qquad \geq \theta_t u_t(m^{t-1}, m_t) + \beta w_t\left(\theta^t, (m^{t-1}, m_t); r\right) \end{aligned}$$

for all $m_t \in M_t\left(m^{t-1}\right)$, all $m^{t-1} \in M^{t-1}$, $t = 0, 1, 2, ..., T$.

Our argument will be to show that for equilibrium strategies the dependence of future reports on the true history of shocks is limited enough so that $w_t\left(\theta^t, m^t; r\right)$ is actually independent of θ^t and we write $w_t\left(\theta^t, m^t; r\right) = w_t^*\left(m^t; r\right)$ for some function w_t^*.

The idea is that in order for self-t to tell on his previous selves he must be indifferent to doing so. This is true because the history of shocks does not affect his payoff directly. With a continuous distribution of shocks indifference is rare (Countable Indifference Lemma, Section 3) and thus the expectation the continuation is not affected by these tattle tales.

Proposition 12 *Assume a continuous distribution over* Θ. *If* r *is an equilibrium strategy given the mechanism* (u, M) *then for all* m^{t-1} *the report* $r_t\left((\theta^{t-1}, \theta_t), m^{t-1}\right)$ *is independent of* θ^{t-1} *for almost all* θ_t *and the continuation utilities* $w_t\left(\theta^t, (m^{t-1}, m_t); r\right)$ *are independent of* θ^t.

Proof. The proof is by induction and applies the Countable Indifference Lemma.

Suppose that given equilibrium strategies r the continuation value for self-t is only a function of m^t and not of θ^t, so that $w_t(\theta^t, m^t; r) = w_t^*(m^t; r)$. We show below that this implies (i) reports at t are independent of θ^{t-1} for almost all θ_t, i.e. $r_t\left((\theta^{t-1}, \theta_t), m^{t-1}; r\right) = r_t^*(\theta_t, m^{t-1}; r)$ for almost all θ; (ii) this, in turn, implies that the continuation value at $t-1$ is independent of θ^{t-1}, i.e. $w_{t-1}\left(\theta^{t-1}, m^{t-1}; r\right) = w^*(m^{t-1}; r)$. Finally, in the last period $w_T\left(\theta^t, m^t; r\right) \equiv 0$ by definition so the result follows by induction using (i) and (ii).

To establish (i) note that since r is an equilibrium the report self-t makes $r_t\left(\theta^t, m^{t-1}; r\right)$ must satisfy

$$r_t\left((\theta^{t-1}, \theta_t), m^{t-1}\right) \in R_t^*\left(\theta_t | m^{t-1}\right)$$

where the correspondence R_t^* is defined by

$$R_t^*\left(\theta | m^{t-1}; r\right) \equiv \arg \max_{\hat{m} \in M_t(m^{t-1})} \left\{\theta u_t(m^{t-1}, \hat{m}) + \beta w_t^*\left(m^{t-1}, \hat{m}; r\right)\right\}$$

By the Countable Indifference Lemma, $R_t^*(\cdot | m^{t-1}, r)$ is single valued except possibly for a countable subset of Θ (it cannot be empty by the definition of an equilibrium). It follows that for all θ where $R_t^*(\cdot | m^{t-1}, r)$ is single valued $r_t\left((\theta^{t-1}, \theta), m^{t-1}\right)$ cannot depend on θ^{t-1}. Thus, we can write

$$r_t\left((\theta^{t-1}, \theta), m^{t-1}\right) = r_t^*\left(\theta, m^{t-1}\right)$$

except for a countable subset of Θ. In short: $r_t = r_t^*$ almost everywhere.

To establish (ii) note that the continuation value w_{t-1} satisfies:

$$\begin{aligned}
w_{t-1}\left(\theta^{t-1}, m^{t-1}; r\right) &= \int_{\Theta} [\ \tilde{\theta}_t u_t \left(m^{t-1}, r_t((\theta^{t-1}, \tilde{\theta}_t), m^{t-1})\right) \\
&\qquad + w_t^* \left(m^{t-1}, r_t((\theta^{t-1}, \tilde{\theta}_t), m^{t-1}); r\right)\ ] f(\tilde{\theta}_t) d\tilde{\theta}_t \\
&= \int_{\Theta} [\ \tilde{\theta}_t u_t \left(m^{t-1}, r_t^*(\tilde{\theta}_t, m^{t-1})\right) \\
&\qquad + w_t^* \left(m^{t-1}, r_t^*(\tilde{\theta}_t, m^{t-1}); r\right)\ ] f(\tilde{\theta}_t) d\tilde{\theta}_t \\
&\equiv w_{t-1}^*(m^{t-1}; r)
\end{aligned}$$

The first equality follows from the definition of w_{t-1} and the hypothesis regarding w_t. The second follows from the result above that $r_t = r_t^*$ almost everywhere. ■

Below we use this result to argue that we can, without loss in generality, consider a simpler mechanism that only requests messages on the current taste shock and focus on the truth-telling equilibrium.

5.2.2 Simple Direct Mechanism

For any mechanism (u, M) with equilibrium r we define an associated direct mechanism with $\tilde{M}_t = \Theta$ and payoff,

$$\tilde{u}_t\left(\theta^t\right) \equiv u_t\left(r^{t,0}\left(\theta^t\right)\right).$$

This is the payoff that is obtained along the equilibrium path of (u, M) with r after a history of shocks θ^t. For this direct mechanism $(\tilde{u}, \tilde{M})$ consider the truth-telling strategy $\tilde{r}\left(\theta^t\right) = \theta_t$. The continuation implied by $\tilde{u}, \tilde{M}$ with strategy $\tilde{r}$ satisfies:

$$\tilde{w}_t(\theta^t, \hat{\theta}^t; \tilde{r}) = w_t(\hat{\theta}^t, r^{t,0}(\hat{\theta}^t); r).$$

that is, $\tilde{w}_t$ is the continuation value that is obtained in the original mechanism (u, M) with equilibrium r along the equilibrium path if the realized history of shocks equal $\hat{\theta}^t$.

Finally, the equilibrium condition for $\tilde{r}$ is,

$$\theta_t \tilde{u}_t(\hat{\theta}^{t-1}, \theta_t) + \beta \tilde{w}_t(\theta^t, (\hat{\theta}^{t-1}, \theta_t); \tilde{r}) \geq \theta_t \tilde{u}_t(\hat{\theta}^{t-1}, \hat{\theta}_t) + \beta \tilde{w}_t(\theta^t, (\hat{\theta}^{t-1}, \hat{\theta}_t); \tilde{r}) \qquad (19)$$

for all $\theta^t, \hat{\theta}^t \in \Theta^t$, which we term the incentive compatibility constraints.

The next proposition shows that this new strategy profile $\tilde{r}$ is a subgame perfect equilibrium of the simple mechanism $\left(\tilde{u}, \tilde{M}\right)$ and it achieves, by construction, the same principal's utility level than mechanism (u, M).

Proposition 13 *The truth telling strategy $\tilde{r}$ is an equilibrium of the direct mechanism $(\tilde{u}, \tilde{M})$ defined above.*

Proof. We need to check the incentive compatibility constraints (19). By definition of $\tilde{u}$ in terms of u we have,

$$\theta_t \tilde{u}_t(\hat{\theta}^{t-1}, \theta_t) + \beta \tilde{w}_t(\theta^t, (\hat{\theta}^{t-1}, \theta_t); \tilde{r}) = \theta_t u_t(r^{t,0}(\hat{\theta}^{t-1}, \theta_t)) + \beta w_t((\hat{\theta}^{t-1}, \theta_t), r^{t,0}(\hat{\theta}^{t-1}, \theta_t); r)$$

And since the initial mechanism induced r as an equilibrium we have that,

$$\begin{aligned} \theta_t u_t(r^{t,0}(\hat{\theta}^{t-1}, \theta_t)) + \beta w_t((\hat{\theta}^{t-1}, \theta_t), r^{t,0}(\hat{\theta}^{t-1}, \theta_t); r) \\ \geq \theta_t u_t(r^{t,0}(\hat{\theta}^t)) + \beta w_t((\hat{\theta}^{t-1}, \theta_t), r^{t,0}(\hat{\theta}^t); r) \end{aligned}$$

since $r_t(\hat{\theta}^t) \in M_t(r^{t-1,0}(\hat{\theta}^{t-1}))$ is a feasible message. Finally, note that by definition $u_t(r^{t,0}(\hat{\theta}^t)) = \tilde{u}_t(\hat{\theta}^t)$ and that

$$w_t((\hat{\theta}^{t-1}, \theta_t), r^{t,0}(\hat{\theta}^t); r) = w_t^*(r^{t,0}(\hat{\theta}^t); r) = w_t(\theta^t, r^{t,0}(\hat{\theta}^t); r) = \tilde{w}_t(\theta^t, \hat{\theta}^t; r)$$

Combining the first equality, the two inequalities and these last relations leads to the incentive compatibility constraint. ■

So without loss in optimality, we can focus in simple direct mechanisms where the agent is called to report only is current shock. This consideration lead us to write the problem with $T \geq 3$ remaining periods and income y_T recursively as follows.

$$v_T(y_T) = \max_{c_T,\, k_T} \int \left[\theta U(c_T(\theta)) + v_{T-1}(k_T(\theta))\right] dF(\theta)$$

$$\theta U(c_T(\theta)) + \beta v_{T-1}(k_T(\theta)) \geq \theta_h U(c_T(\theta')) + \beta v_{T-1}(k_T(\theta')) \text{ for all } \theta, \theta' \in \Theta$$

$$c_T(\theta) + k_T(\theta) \leq y_T \text{ for all } \theta \in \Theta$$

where $v_2(\cdot)$ was defined in Section 1.

In above formulation, we make of use of the following: any feasible continuation utility profile $\tilde{w}_{T-1}(\theta) \in [(T-1)U(0), v_{T-1}(k_T(\theta))]$ can be achieved by 'money burning': setting $\tilde{w}_{T-1}(\theta) = v_{T-1}(\tilde{k}_T(\theta))$ for some $\tilde{k}_T(\theta) \leq k_T(\theta)$. Consequently, without loss of generality we can impose ex-post optimality, that is given the resources available the continuation utility is $v_{T-1}(\tilde{k}_T(\theta))$.

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_{T-1}. This is not true in the alternative setup where the principal and the agent both discount exponentially but with different discount factors.

6 Liquid and Illiquid Assets: Sequential Implementation

Illiquid assets have been emphasized as a commitment device that helps mitigate the time inconsistency problem. In particular, Laibson (1997) shows that illiquid assets are used by agents and improve welfare. Angeletos, Laibson, Repetto, Tobacman, Weinburg (2001) numerically simulate a hyperbolic consumer's problem that faces shocks to income and can manage a portfolio of illiquid and liquid risk free assets. They report that illiquid assets are an important component of wealth.

Although the commitment property of illiquid assets has been stressed, the optimal commitment devices have not been studied and consequently the virtue of illiquid assets as a commitment device has not been fully assessed. Surprisingly, we now show that in our model, if assumption A holds, a combination of liquid and illiquid assets under control of the agent can implement the optimal allocation and achieve the optimal welfare. This result is also true with time-consistent preferences, with and without self-control.

To establish this result we first establish a sequential implementation result that is interesting in its own right: the optimal minimum savings mechanism can be modified so that the agent freely selects in each period the minimum savings requirement for the next period (subject only to the budget constraint). This added sequential choice turns out not to affect the allocation outcome nor the welfare obtained by it (recall that with self-control the first statement does not necessarily imply the second).[4]

This follows from the fact that the disagreement on discounting – between selves in the time-inconsistent framework or between the regular and temptation preferences in the time consistent framework – is limited to the current period. There is no disagreement regarding future periods and in particular no disagreement on the optimal minimum savings for the next period, given the saving level chosen in the current period. Thus, the entire tension remains on the current choice between consumption and savings. Thus, there is no need to commit in the initial period to future commitment devices.

Consider the case with time-consistent self-control preferences. The dynamic program summarizing the sequential choice problem described above is simply:

$$\begin{aligned} w_T\left(y,s\right) = \mathrm{E}\Big\{ & \left(1+\phi\right) \max_{(c,k,s')\in\tilde{B}(y,s)} \left(zU\left(c\right)+\hat{\beta}w_{T-1}\left(k,s'\right)\right) \\ & - \phi \max_{(c,k,s')\in\tilde{B}(y,s)} \left(zU\left(c\right)+\beta w_{T-1}\left(k,s'\right)\right)\Big\} \end{aligned}$$

where $\tilde{B}\left(y,s\right) \equiv \left\{\ (c,k,s') : c+k \leq y,\ k \geq s,\ k \geq s'\ \right\}$. Note that a choice problem at time t is indexed now by the minimum savings requirement s for the current period and the total amount of resources y. The choices available in the period are greater than previously: agents choose over c and k as before, subject to the current minimum savings requirement s and in addition select a minimum savings requirement s' for the next period.

In the last period, provided assumption A holds, we have that,

$$v_2(y) = \max_{s\in[y,\infty)} w_2\left(y,s\right)$$

by Proposition 10. That is, the optimal mechanism is the best minimum savings rule, so the welfare values are equivalent. We now proceed by backward induction: suppose that $v_{T-1}\left(k\right) = \max_s w_{T-1}\left(y,s\right)$, we will prove that this implies that $v_T\left(k\right) = \max_s w_T\left(y,s\right)$.

[4] Indeed, this is true regardless of whether or not the optimal commitment device is a minimum savings rule and is implicit in the dynamic programming formulation of previous sections. However, here we introduce notation that allows us to stress this point for the minimum savings rule.

Maximizing over s on both sides of (**??**) and performing the maximization over s' and maximize over s to write:

$$\begin{aligned}\max_{s\in[y,\infty)} w_T(y,s) &= \max_{s\in[y,\infty)} \mathrm{E}\Big\{ (1+\phi) \max_{(c,k)\in B(y)} \big(zU(c)+\hat{\beta}v_{T-1}(k)\big) \\ &\qquad - \phi \max_{(c,k)\in B(y)} \big(zU(c)+\beta v_{T-1}(k)\big)\Big\} \\ &= v_T(y,s)\end{aligned}$$

which proves that the sequential choice problem attains the same welfare as the optimal commitment device.

Now consider the following alternative setup where the agent has access each period to a liquid asset, L, and an illiquid asset, I, with the following characteristics. Both assets have the same rate of return, the only difference is that the liquid asset can be used immediately for consumption within the period. In contrast the illiquid asset must be 'sold' one period in advance and converted into liquid assets before its value can be consumed. The budget constraints for the consumer are therefore:

$$\begin{aligned} c_t &\leq L_t \\ c_t + k_{t+1} &\leq k_t \\ 0 &\leq L_{t+1}, I_{t+1}\end{aligned}$$

Where $k_t = L_t + I_t$ represents total assets; while L_t and I_t represent the portfolio split between liquid and illiquid assets, respectively. The consumer's consumption, c_t, is constrained by his availability of liquid resources L_t. Thus, illiquid assets $I_t = k_t - L_t$ can only be consumed by converting them first into liquid assets and consuming them in the following period.

Clearly the consumer's budget constraint $c_t + k_{t+1} \leq k_t$ will always hold with equality. Thus, for the consumer the constraints above are equivalent to:

$$\begin{aligned} k_{t+1} &\geq I_t \\ c_t + k_{t+1} &= k_t \\ 0 &\leq I_{t+1}\end{aligned}$$

so that the choice of illiquid assets at time t effectively imposes a minimum savings constraint for $t+1$. This shows that a time-consistent consumer with self-control will use the assets to implement the optimal allocation and achieve the optimal welfare. Similar arguments establish the case without self control or time-inconsistent preferences. We summarize these results in the following proposition.

Proposition 14 *If assumption A holds, then in time consistent and time-inconsistent models with liquid and illiquid assets the allocation and welfare are equivalent to those with the optimal mechanism.*

7 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) and the hyperbolic preferences.

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.

We showed that this optimal mechanism could be implemented sequentially with an illiquid asset. The agent each period consumes out of his liquid holdings and decides the portfolio shares of his savings allocated to liquid and illiquid funds for next period. We showed that with these assets, the agent implements the optimal allocation and obtains the optimal welfare.

Our model is open to other interpretations, discussed in Amador, Werning and Angeletos (2003). For example, a paternalistic principal who cares about an informed agent but believes the agent is biased on average in his choices faces a similar trade-off. We also discuss two other applications, externalities and schooling choices by teenagers.

The problem we studied imposes a standard budget constraint, thus we abstracted from insurance. This choice was motivated by several considerations. First, the case without insurance is of direct relevance if insurance is not possible because of other considerations outside the scope of our model. Second, this assumption is in keeping with work that has studied the role of illiquid assets as commitment devices making our work comparable to this literature.

Finally, even without temptation or time inconsistent preferences, i.e. $\beta = 1$, constrained optimum insurance problems with private information, such as Mirrlees (1971) and Atkeson and Lucas (1995), are nontrivial and the resulting optimal allocations have avoided sharp characterizations. Thus, with insurance, comparing situations with temptation or time inconsistency to those without would be difficult. In contrast, with no insurance, the optimal allocation without temptation is straightforward – every agent chooses their tangency point on the budget line – making the comparison with temptation clear.

In any case, we hope that studying the case without insurance may yield insights into the case with insurance which we leave for future research.

A Lemma of Optimality and First Order Conditions

The Lemma of Optimality of Section 2 characterizes an allocation as optimal if and only if appropriate first order conditions on the Lagrangian hold. This was a fundamental step for the results of Proposition 3. In this appendix we provide the proof of this important Lemma.

To do this, we first show that the maximization of the Lagrangian is a necessary and sufficient condition for optimality of an allocation. This is stated in the following two results,

Result (a'). *Necessity.* An allocation $\left(\underline{w_0}, u_0\right) \in \Phi$ with u_0 continuous is optimal if there exists a non-decreasing Λ_0 such the Lagrangian is maximized at $(\underline{w_0}, u_0) \in \Phi$:

$$L\left(\underline{w_0}, u_0; \underline{h_w}, h_u|\Lambda_0\right) \leq L\left(\underline{w_0}, u_0; \underline{w_0}, u_0|\Lambda_0\right) \tag{20}$$

for all $(\underline{h_w}, h_u) \in \Phi$ and h_u continuous

Result (b'). *Sufficiency.* If there exists a non-decreasing Λ_0 such that for some $(\underline{w_0}, u_0) \in \Phi$

$$L\left(\underline{w_0}, u_0; \underline{h_w}, h_u|\Lambda_0\right) \leq L\left(\underline{w_0}, u_0; \underline{w_0}, u_0|\Lambda_0\right) \text{ and all } \left(\underline{h_w}, h_u\right) \in \Phi \tag{21}$$

then the allocation $(\underline{w_0}, u_0)$ is optimal.

Proof of Results (a') and (b'). Our optimization problem maps into the general problem studied in Sections 8.3-8.4 by Luenberger (1969): $\max_{x \in X} Q(x)$ subject to $x \in \Omega$ and $G(x) \in P$, where Ω is a subset of the vector space X, $Q : \Omega \to \mathbb{R}$ and $G : \Omega \to Z$, where Z is a normed vector space and P is a positive non-empty convex cone in Z.

For Result (b'), set:

$$\begin{aligned}
&X = \{\underline{w}, u \mid \underline{w} \in W\left(\mathbb{R}^+\right) \text{ and } u : \Theta \to \mathbb{R}\} \\
&\Omega = \{\underline{w}, u|\ \underline{w} \in W\left(\mathbb{R}^+\right),\ u : \Theta \to U\left(\mathbb{R}^+\right) \text{ and } u \text{ is non-decreasing}\} \equiv \Phi \\
&Z = \left\{z \mid z : \Theta \to \mathbb{R} \text{ with } \sup_{\theta \in \Theta} |z(\theta)| < \infty\right\} \text{ with the norm } \|z\| = \sup_{\theta \in \Theta} |z(\theta)| \\
&P = \{z \mid z \in Z \text{ and } z(\theta) \geq 0 \text{ for all } \theta \in \Theta\}
\end{aligned}$$

We also define the objective function Q and the left hand side of the resource constraint G by,

$$\begin{aligned}
Q(\underline{w}, u) &= \frac{\theta}{\beta} u(\underline{\theta}) + \underline{w} + \frac{1}{\beta} \int_{\underline{\theta}}^{\bar{\theta}} \left[(1 - F(\theta)) - \theta(1 - \beta) f(\theta)\right] u(\theta)\, d\theta \\
G(\underline{w}, u) &= K^{-1}(y - C(u(\theta))) + \frac{\theta}{\beta} u(\theta) - \frac{\theta}{\beta} u(\underline{\theta}) - \underline{w} - \int_{\underline{\theta}}^{\theta} \frac{1}{\beta} u(\tilde{\theta}) d\tilde{\theta}
\end{aligned}$$

Result (b') then follows immediately since the hypothesis of Theorem 1, pg. 220 in

Luenberger (1969) are met.

For Result (a'), modify Ω and Z to require continuity of u:

$$\Omega = \{\underline{w}, u |\ \underline{w} \in W\left(\mathbb{R}^+\right),\ u : \Theta \to U\left(\mathbb{R}^+\right) \text{ and } u \text{ is continuous and non-decreasing}\}$$
$$Z = \{z \mid z : \Theta \to \mathbb{R} \text{ and } z \text{ is continuous}\} \text{ with the norm } \|z\| = \sup_{\theta \in \Theta} |z(\theta)|$$

with X, P, Q and G as before. Note that Q and G are concave, Ω is convex, P contains an interior point (e.g. $z(\theta) = 1$, for all $\theta \in \Theta$, is interior) and that the positive dual of Z is isomorphic to the space of non-decreasing functions on Θ by the Riesz Representation Theorem (see Chapter 5, pg. 113 in Luenberger (1969)). Finally, if $\underline{w_0}, u_0$ is optimal within Φ and $\underline{w_0}, u_0 \in \Phi \cap \{u \textit{ is continuous}\}$ then $\underline{w_0}, u_0$ is optimal within the subset $\Phi \cap \{u \textit{ is continuous}\} \equiv \Omega$. Result (a') then follows since the hypothesis of Theorem 1, pg. 217 in Luenberger (1969) are met.■

Once obtained results (a') and (b'), to prove the Lemma of Optimality, we need to show that the maximization conditions in (20) and (21) are equivalent to the appropriate first order conditions. We first show that these first order conditions can indeed be computed. The following Lemma helps in this.

Lemma A.1. (Differentiability of integral functionals with convex integrands). *Given a measure space (Θ, Θ, μ) and a function $\varphi : X \times \Theta \to R$, where $X \subset R^n$, suppose the functional $T : \Omega \to R$, where Ω is some subset of the space of all functions mapping Θ into X, is given by,*

$$T(x) = \int_\Theta \varphi(x(\theta), \theta)\, \mu(d\theta)$$

Suppose that (i) for each $\theta \in \Theta$, $\varphi(\cdot, \theta) : X \to R$ is concave; (ii) that the derivative φ_x exists and is a continuous function of (x, θ); and that (iii) $x + \alpha h \in \Omega$ for $\alpha \in [0, \varepsilon]$ for some $\varepsilon > 0$.

Then the h-directional Gateaux differential, $\partial T(x; h)$ exists and is given by

$$\partial T(x; h) = \int_\Theta \varphi_x(x(\theta), \theta)\, h(\theta)\, \mu(d\theta),$$

if the right hand side expression is well defined.

Proof. By definition of Gateaux differential,

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

We seek to show that the last term is well defined and vanishes.

For $\alpha < \varepsilon$ one can show that,

$$\begin{aligned}&\left|\frac{1}{\alpha}\left[\varphi\left(x(\theta)+\alpha h(\theta),\theta\right)-\varphi\left(x(\theta),\theta\right)\right]-\varphi_x\left(x(\theta),\theta\right)h(\theta)\right|\\ &\leq\left|\frac{1}{\varepsilon}\left[\varphi\left(x(\theta)+\varepsilon h(\theta),\theta\right)-\varphi\left(x(\theta),\theta\right)\right]-\varphi_x\left(x(\theta),\theta\right)h(\theta)\right|,\end{aligned}\tag{22}$$

by concavity of $\varphi(\cdot,\theta)$. Given that $\varphi\left(x(\theta)+\varepsilon h(\theta),\theta\right)$, $\varphi\left(x(\theta),\theta\right)$ and $\varphi_x\left(x(\theta),\theta\right)h(\theta)$ are all integrable by hypothesis then it follows that

$$\frac{1}{\varepsilon}\left[\varphi\left(x(\theta)+\varepsilon h(\theta),\theta\right)-\varphi\left(x(\theta),\theta\right)\right]-\varphi_x\left(x(\theta),\theta\right)h(\theta)$$

is also integrable. Since a function is integrable if and only if its absolute value is integrable it follows that (22) provides the required integrable bound to apply Lebesgue's Dominated Convergence Theorem (see Theorem 7.10, pg. 192, Stokey and Lucas with Prescott, 1989) implying:

$$\begin{aligned}&\lim_{\alpha\downarrow 0}\int_\Theta\left[\frac{1}{\alpha}\left[\varphi\left(x(\theta)+\alpha h(\theta),\theta\right)-\varphi\left(x(\theta),\theta\right)\right]-\varphi_x\left(x(\theta),\theta\right)h(\theta)\right]\mu(d\theta)\\ &=\int_\Theta\left[\lim_{\alpha\downarrow 0}\frac{1}{\alpha}\left[\varphi\left(x(\theta)+\alpha h(\theta),\theta\right)-\varphi\left(x(\theta),\theta\right)\right]-\varphi_x\left(x(\theta),\theta\right)h(\theta)\right]\mu(d\theta)=0\end{aligned}$$

where the second equality follows by definition of φ_x. It follows that $\partial T(x;h) = \int_\Theta \varphi_x\left(x(\theta),\theta\right)h(\theta)\,\mu(d\theta)$. ■

In our case, we can apply the lemma A.1. because the Lagrangian functional is the sum of three terms expressible as integrals with concave differentiable integrands. The differentiability of the first two linear terms is trivial. For the last term one applies the lemma to the function $\varphi(u,\theta) = K^{-1}\left(y - C(u)\right) + \frac{\theta}{\beta}u$, whose derivative

φ_u is clearly continuous in (u, θ) and hence (Borel) measurable. Since the Lagrangian functional is defined over a convex cone Φ the hypothesis (iii) of the lemma is met with any $\varepsilon \leq 1$ for any $x \in \Phi$ and $h = y - x$, for $y \in \Phi$.

Furthermore, in our case $\int \varphi_u (u(\theta), \theta) h_u(\theta) d\Lambda(\theta)$ is well defined for any u and h_u such that $(\underline{w}, u) \in \Phi$ and $(h_{\underline{w}}, h_u) \in \Phi$, for some $\underline{w}, h_{\underline{w}} \in \mathbb{R}$. Since u and h_u must be non-decreasing on Θ they are measurable and bounded. It follows that the composition $\varphi_u (u(\theta), \theta)$ is bounded and measurable and finally that the product $\varphi_u (u(\theta), \theta) h_u(\theta)$ is measurable and bounded. Finally, all measurable and bounded functions are integrable.

These arguments establish that we can write the Gateaux differential of the Lagrangian for $(\underline{w}, u), (h_{\underline{w}}, h_u) \in \Phi$ as

$$\begin{aligned} \partial L(\underline{w}, u; h_{\underline{w}}, h_u | \Lambda) = & \left(\frac{\underline{\theta}}{\beta} h_u(\underline{\theta}) + h_{\underline{w}}\right) \Lambda(\underline{\theta}) + \frac{1}{\beta} \int_{\underline{\theta}}^{\bar{\theta}} (\Lambda(\theta) - G(\theta)) h_u(\theta) d\theta \\ & + \int_{\underline{\theta}}^{\bar{\theta}} \left[\frac{\theta}{\beta} - (K^{-1})' (y - C(u(\theta))) C'(u(\theta))\right] h_u d\Lambda(\theta) \end{aligned}$$

which collapses to (9) at the proposed allocation.

Finally, the following Lemma, which is a simple extension of a result in Luenberger (Lemma 1, pg. 227, 1969), allows us to characterize the maximization conditions of the Lagrangian (obtained in results (a') and (b')) by appropriate first-order conditions.

Lemma A.2. (Optimality and first-order conditions) *Let f be a concave functional on P, a convex cone in X. Take $x_0 \in P$ and define $H(x_0) \equiv \{h : h = x - x_0 \text{ and } x \in P\}$. Then $\delta f(x_0, h)$ exists for $h \in H(x_0)$. Assume that for $h_1, h_2 \in H(x_0)$ we have that $\delta f(x_0, \alpha_1 h_1 + \alpha_2 h_2)$ exists and $\delta f(x_0, \alpha_1 h_1 + \alpha_2 h_2) = \alpha_1 \delta f(x_0, h_1) + \alpha_2 \delta f(x_0, h_2)$ for all $\alpha_1, \alpha_2 \in R$.*

A necessary and sufficient condition that $x_0 \in P$ maximizes f is that

$$\begin{aligned} \delta f(x_0, x) &\leq 0 \text{ for all } x \in P \\ \delta f(x_0, x_0) &= 0 \end{aligned}$$

Proof. *Necessity.* If x_0 maximizes f then for any $x \in P$ then $((1-\alpha)x_0 + \alpha x) \in P$ for $\alpha \in (0,1)$ since P is convex and therefore,

$$0 \geq f(x_0 + \alpha(x - x_0)) - f(x_0)$$

dividing by α and taking limits, which exists due to the concavity of f, we obtain:

$$\lim_{\alpha \downarrow 0} \frac{f(x_0 + \alpha(x - x_0)) - f(x_0)}{\alpha} = \delta f(x_0; x - x_0) \leq 0$$

Setting $x = 2x_0 \in P$ (since P is a cone) then $\delta f(x_0; x_0)$ exists and

$$\delta f(x_0; x_0) \leq 0 \tag{23}$$

Since whenever $\delta f(x_0; h)$ exists then $\delta f(x_0; \alpha h)$ exists for all $\alpha \in R$ and $\delta f(x_0; \alpha h) = \alpha \delta f(x_0; h)$. Now take $x = x_0/2 \in P$ (since P is a cone) then $\delta f(x_0; -x_0/2)$ exists and using $\alpha = -2$ it follows that $\delta f(x_0; x_0) = -2\delta f(x_0; -x_0/2)$, implying:

$$\delta f(x_0; -x_0/2) = -\delta f(x_0; x_0)/2 \leq 0$$

Implying:

$$\delta f(x_0; x_0) \geq 0. \tag{24}$$

Combining (23) and (24) we obtain:

$$\delta f(x_0; x_0) = 0. \tag{25}$$

From (23) and (25) we obtain that:

$$\begin{aligned} \delta f(x_0; x - x_0) &= \delta f(x_0; x) + \delta f(x_0; -x_0) \\ &= \delta f(x_0; x) - \delta f(x_0; x_0) \\ &= \delta f(x_0; x) \leq 0 \end{aligned}$$

The first step is warranted since we assume that if $h_1 \in P$ and $-h_2 \in P$ then $\delta f(x_0; h_2)$ exists and $\delta f(x_0; h_1 + h_2) = \delta f(x_0; h_1) + \delta f(x_0; h_2)$.

Sufficiency. For $x_0, x \in P$ and $\alpha \in (0,1)$ concavity of f implies that:

$$f\left(x_0 + \alpha\left(x - x_0\right)\right) \geq f\left(x_0\right) + \alpha(f\left(x\right) - f\left(x_0\right))$$

or

$$f\left(x\right) - f\left(x_0\right) \leq \frac{1}{\alpha}\left[f\left(x_0 + \alpha\left(x - x_0\right)\right) - f\left(x_0\right)\right]$$

as $\alpha \downarrow 0$ then:

$$f\left(x\right) - f\left(x_0\right) \leq \delta f\left(x_0, x - x_0\right) \tag{26}$$

It follows that if

$$\begin{aligned} \delta f\left(x_0, x\right) &\leq 0 \text{ for all } x \in P \\ \delta f\left(x_0, x_0\right) &= 0 \end{aligned}$$

then using (26):

$$\begin{aligned} f\left(x\right) - f\left(x_0\right) &\leq \delta f\left(x_0, x - x_0\right) \\ &= \delta f\left(x_0, x\right) - \delta f\left(x_0, x_0\right) \leq 0 \end{aligned}$$

where the equality follows from the hypothesis of linearity of δf. ■

All the hypothesis of Lemma A.2 are met for the Lagrangian in our case because it is a convex functional over a convex cone and Lemma A.1 verifies the differentiability requirement, as discussed above. Thus, we obtain that a necessary and sufficient condition for the Lagrangian to be maximized at $\left(u_0, \underline{w_0}\right)$ over Φ is that

$$\partial L\left(\underline{w_0}, u_0; \underline{w_0}, u_0 | \Lambda_0\right) = 0$$

$$\partial L\left(\underline{w_0}, u_0; h_{\underline{w}}, h_u | \Lambda_0\right) \leq 0$$

for all $(h_{\underline{w}}, h_u) \in \Phi$.

Given results (a') and (b'), the proof of the Lemma of Optimality follows.

B 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 principal 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

$$\theta U'[c^*(\theta)] = \beta W'(y - c^*(\theta)) \tag{27}$$

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

$$\begin{aligned}& \theta^*(\theta_H, \theta_L) U(c^*(\theta_H)) + \beta W(y - c^*(\theta_H)) \\ & = \theta^*(\theta_H, \theta_L) U(c^*(\theta_L)) + \beta W(y - c^*(\theta_L))\end{aligned} \tag{28}$$

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

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

Combining terms,

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

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

Now, from (28) we have

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

Substituting above

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

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

we also have that from (27)

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

So,

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

Differentiating (28) we obtain:

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

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

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

Substituting back the result follows. ■

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

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

Notice that

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

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

Note that $\partial^2 S\left(\theta,\theta^*\right)/\left(\partial\theta\right)^2$ 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

$$\frac{\theta f'\left(\theta\right)}{f\left(\theta\right)}\geq-\frac{2-\beta}{1-\beta}\tag{29}$$

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

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

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 an improvement.

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 an improvement to remove the whole interval. In other words,

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

This concludes the proof. ■

C Proof of Lemma A implies B

Let $\phi = \frac{1}{\varepsilon} > 0$ then assumption B 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 $[\underline{\theta}, \theta_p\hat{\beta}/\beta]$ then $\Phi_\varepsilon(\theta, \varepsilon) \geq 0$ for all $\varepsilon \geq 0$ and $\theta \in [\hat{\underline{\theta}}, \theta_p]$. Given that $\Phi(\theta, 0) = 0$, we have that if A holds for $[\underline{\theta}, \theta_p\hat{\beta}/\beta]$, then

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

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

References

[1] **Amador, Manuel; Ivan Werning; George-Marios Angeletos** (2003) "Optimal Commitment", NBER Working Paper #10151.

[2] **Angeletos, George-Marios; David Laibson; Jeremy Tobacman; Andrea Repetto and Stephen Weinberg** (2001) "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation", Journal of Economic Perspectives v15, n3: 47-68.

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

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

[5] **Caplin, Andrew and John Leahy** (2001) "The Social Discount Rate", Discussion Paper 137.

[6] **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.

[7] **Dekel, Eddie; Barton L. Lipman and Aldo Rustichini** (2001) "Representing Preferences with a Unique Subjective State Space" *Econometrica,* v69, n4: 891-934

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

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

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

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

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

[13] ———————— (2002a) "Self-control and the theory of consumption", forthcoming in *Econometrica*.

[14] ———————— (2002b) "Self-control, revealed preference and consumption choice", prepared for Society for Economic Dynamics.

[15] **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.

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

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

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

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

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

[21] ———————— (1997) "Golden Eggs and Hyperbolic Discounting" *Quarterly Journal of Economics,* v112, n2: 443-77.

[22] ———————— (1998) "Life-Cycle Consumption and Hyperbolic Discount Functions", *European Economic Review*, Vol. 42, nos. 3-5: 861-871.

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

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

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

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

[27] **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.

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

[29] **Rudin, Walter** (1976) "Principles of Mathematical Analysis", 3rd Edition, McGraw-Hill.

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

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