Optimal Retirement Policies With Time-Inconsistent Agents *

Pei Cheng Yu †
April 14, 2017

Abstract

This paper develops a general theory for the design of retirement policies, like social security and retirement accounts, within a Mirrlees taxation framework with time-inconsistent agents. The paper shows how the design of off-equilibrium path policies utilize the time inconsistency of agents to improve welfare. Despite the presence of asymmetric information, the full information efficient outcome is implementable, regardless of the degree of sophistication or temptation. In particular, in an environment with both time-consistent and time-inconsistent agents, welfare increases monotonically with the population of time-inconsistent agents. For implementation, the paper focuses on the design of social security and retirement accounts. The optimal policy has social security benefits decreasing in progressivity with the initial withdrawal age. It also allows early withdrawals from retirement accounts only when there are large income discrepancies. These proposals outperform traditional policies, like linear savings subsidies or mandatory savings, by raising welfare above the constrained efficient optimum. (JEL Codes: D03, D62, D82, D84, D86, D91, H21)

Keywords: Retirement policy, Time inconsistency, Optimal taxation

---

*This was my job market paper and Chapter 1 of my PhD dissertation. I thank Manuel Amador, Anmol Bhandari, V.V. Chari, Kim-Sau Chung, Emmanuel Farhi, Xavier Gabaix, Simone Galperti, Mike Golosov, Weiwen Leung, Erzo Luttmer, Radek Paluszyński, Aldo Rustichini, Martin Szydlowski, Jan Werner, Zoe Leiyu Xie, Pierre Yared and Hsin-Jung Yu for invaluable comments and suggestions. I am especially grateful to David Rahman for his guidance and endless encouragements. I also acknowledge the support of the University of Minnesota Doctoral Dissertation Fellowship.

†University of New South Wales (e-mail: pei-cheng.yu@unsw.edu.au)
1 Introduction

Policymakers and researchers have been concerned with the issue of inadequate retirement savings\(^1\). Empirical evidence shows that models with time-inconsistent preference can explain the consumption and savings patterns observed in the data (Angeletos et al. (2001) and Laibson et al. (2007)). In response, strengthening social security and increasing participation in retirement accounts are mentioned as core issues by the last US administration\(^2\). On the other hand, policymakers are also concerned about social insurance and the sustainability of these programs. The top two recommendations of the National Commission on Fiscal Responsibility and Reform (2010) (henceforth NCFRR) are to make social security benefits more progressive with income, and to enhance the minimum benefits for low-wage workers. It also recommends increasing the maximum amount of taxable income for social security.

For many, this reform will raise taxes and cut their social security benefits, which could introduce additional distortions to the labor supply\(^3\). As a result, the trade-off between increasing retirement welfare and minimizing the cost of its provision is an urgent issue.

This paper addresses this by extending the Mirrlees taxation framework (agents privately observe productivity) to include time-inconsistent agents. It provides a general theoretical framework to designing retirement policies, which sheds light on features in social security and retirement accounts that could raise welfare\(^4\). The key is in utilizing the agent’s time inconsistency, so the policies go beyond mitigating the present bias. Traditional policy suggestions, such as linear savings subsidies or mandatory savings, increase savings by offsetting the bias independent of the asymmetric information. Using traditional policies, the government is able to guarantee the constrained efficient optimum, but not better. I consider more sophisticated policy

---

\(^1\)The National Research Council (2012) finds that up to \(\frac{2}{3}\) of the US population is saving inadequately for retirement. Scholz et al. (2006), one of the more conservative studies, estimates at least 20% of the US population are not saving enough for retirement.

\(^2\)See President Obama’s core issues (https://www.whitehouse.gov/issues/seniors-and-social-security). It also aims to utilize lessons learned from behavioral economics to attain its goals (see Executive Order No. 13707 (2015)).

\(^3\)Auerbach et al. (2016) show the current fiscal system may encourage the elderly to retire early. It demonstrates how the optimal retirement policy needs to be considered in tandem with the design of the tax system.

\(^4\)By retirement policies, I am referring to social security, policies regarding retirement accounts, and any other programs or policies related to retirement welfare. For the paper, I will be focusing on social security and retirement accounts.
instruments.

With time-inconsistent agents, the design of off-equilibrium path policies is important (Esteban and Miyagawa (2005), Eliaz and Spiegler (2006) and Galperti (2015)). It exploits the disagreement between present and future-selves. This is the first paper to use this technique in public finance and to explore its implications on the design of retirement policies. Recent taxation papers with behavioral biased agents, Farhi and Gabaix (2015), Lockwood (2015) and Moser and de Souza e Silva (2015), do not take advantage of the bias using the methods employed in this paper. Therefore, this paper provides a new perspective on how time inconsistency affects welfare and the optimal design of policies.

To see how policies can be designed to exploit time inconsistency, note that naïve agents are unaware of their time inconsistency. The government can design policies that induce efficient output by promising larger benefits in the future, which unknown to the agent, their future-selves would reject in favor of higher immediate benefits. Sophisticated agents are aware of their bias and thus demand commitment. The government can provide commitment in exchange for efficient output, which is sustained by off-equilibrium path policies that tempt the agents’ future-selves and unravel commitment if they worked inefficiently. In essence, the optimal policy has naïve agents exchanging information rents for empty promises, while sophisticated agents exchange information rent for commitment. In an economy with only time-inconsistent agents, the government can implement the full information efficient allocation (efficient allocation) despite the presence of information asymmetry. This is true regardless of sophistication levels.

This result provides new insights on the design of existing policies. It has been argued that people in the US are claiming social security benefits too early in life, and should instead retire later and delay benefits claiming. The NCFRR recommends the use of behavioral economics, more specifically choice architecture, to nudge people to retire and claim benefits later. Contrary to this perspective, this paper suggests that social security benefits should decrease in progressivity with the initial age of claiming benefits. Non-sophisticated agents would plan to claim at a later age and

---

5Efficient allocation satisfies consumption smoothing, full insurance and efficient production (See Proposition 1).
7Benefits are progressive if the ratio of lifetime benefits to lifetime payroll taxes is higher for low income individuals than it is for people with higher average income. Benefits decrease in progressivity
the decrease in progressivity would encourage them to work efficiently, which helps the sustainability of the social security system by increasing taxable income. However, they would claim earlier than planned and the more progressive benefits for early claimants improve social insurance.

Sophisticated agents are concerned that they would withdraw early from their retirement accounts due to present bias, and therefore prefer illiquid accounts (Beshears et al. (2015a) and Beshears et al. (2015b)). I show that increasing the liquidity of retirement accounts to allow for early withdrawals helps implement the efficient allocation. The idea is to allow for early withdrawals only if the agent’s present income is significantly higher than past income. Agents who work efficiently would face illiquid retirement accounts in the future. Agents who work inefficiently would face a liquid account that tempts them to work more to withdraw early in the future, resulting in low savings for retirement. Thus, agents work efficiently for commitment.

I also discuss how the main insights can be extended to an environment with hidden sophistication level and hidden present bias. The multi-dimensional screening problem does not change the results. A special case is when some agents in the economy are time-consistent. The presence of time-consistent agents limits the effectiveness of off-equilibrium path policies. I find that welfare increases with the population of time-inconsistent agents. This is because total output increases with the proportion of agents with bias, so the government can provide more information rent per time-consistent agent without causing additional distortions.

1.1 Example: Story of Lear

Consider the story of Lear, who lives for only two days, today and tomorrow. Lear works in the mornings and consumes and saves at night, though he is also able to work overtime and simultaneously make consumption-savings decisions. He has two options: work today and retire tomorrow, or retire today. Lear is also extremely present-biased: if left to his own accord, he would not have any savings for retirement ($\beta = 0$ in a quasi-hyperbolic model). The government prefers Lear to work today if he was productive, while it would prefer him to retire early if he was
not (efficient production). The government wants to redistribute from the productive to the unproductive, but productivity is not observed by the government. To the government, Lear is productive or unproductive with equal probability. To prevent a tragedy, the government would also like to smooth Lear’s consumption. This story captures the challenge in designing retirement policies.

The following table summarizes Lear’s output:

<table>
<thead>
<tr>
<th></th>
<th>Unproductive</th>
<th>Productive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Work</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Work</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 1: Lear’s Output

Production is strictly positive (4 units) only when Lear is productive and chooses to work. Disutility from work is $-\infty$ if Lear was unproductive, and $-1$ if he was productive. Lear has period utility $u(c) = \ln(c)$ and discount factor $\delta = 1$.

The government likes Lear to consume one unit each day (consumption smoothing), regardless of whether he was productive or not (full insurance). Traditional policies are unable to implement consumption smoothing and full insurance without efficiency loss. More sophisticated instruments are needed.

If Lear were fully naïve, the following social security system can implement the efficient allocation. Suppose Lear has the option of claiming social security today or tomorrow. If he claims today, he would receive a unit of benefit for each day regardless of whether he worked or not. If he works today and claims benefits tomorrow, his consumption today would be taxed but he would received a large benefit tomorrow: his after-tax consumption today would be $1 - \epsilon$ and the benefit he receives tomorrow would be $\exp\left(\frac{1}{1-\epsilon}\right)$, where $\epsilon \in (0, 1)$.

Given this policy, a productive Lear would weakly prefer to work because he expects to claim the large benefit tomorrow: $\ln (1 - \epsilon) - 1 + \ln \left(\frac{\epsilon}{1-\epsilon}\right) \geq \ln (1) - 0 + \ln (1) = 0$. However, after work, he would claim benefits early because he is tempted to consume more today: $\ln (1) + 0 \times \ln (1) > \ln (1 - \epsilon) + 0 \times \ln \left(\frac{\exp\left(\frac{1}{1-\epsilon}\right)}{1-\epsilon}\right)$. The benefits are designed to take advantage of Lear’s incorrect beliefs about his future behavior. Lear believes that by working and sacrificing a bit of consumption today, he is able to consume more during retirement. However, he does not realize that his future-self completely ignores his retirement welfare. This is accomplished by a higher ratio of benefits to taxes for a productive Lear who claims tomorrow rather than today. While
the ratio for an unproductive Lear does not vary with the date of benefits claiming.

If Lear were sophisticated, the policy introduced above fails. However, the government can still implement the efficient allocation with the following retirement account. The government endows Lear with a unit of retirement savings in his account. If Lear works in the morning, then he is unable to withdraw from his retirement account at night. He would receive a unit today and withdraw a unit from the account tomorrow. However, if Lear does not work in the morning, then he can withdraw \(1 - \epsilon\), with \(\epsilon \in (0, 1)\), from his retirement account and receive a subsidy of \(\exp - (1 - \epsilon)\) if he works at night.

Given this policy, if Lear is productive and does not work in the morning, then at night he would weakly prefer withdrawing early and work: \(\ln (\exp) - 1 + 0 \times \ln (\epsilon) \geq \ln (1) + 0 \times \ln (1)\). This would leave him with low retirement savings. For a sufficiently small \(\epsilon\), by backward induction, Lear would prefer to work in the morning and face an illiquid account: \(\ln (1) - 1 + \ln (1) > \ln (\exp) - 1 + \ln (\epsilon)\). Notice that if Lear were unproductive, then he would not be able to withdraw early because working would be prohibitively costly for him. In essence, the retirement account is designed to take advantage of a productive Lear’s fear of depleting his savings for retirement. It is also engineered so that Lear’s productivity can be identified. Only if Lear were productive would he be able to increase his output in exchange for early withdrawal. Hence, a productive Lear would prefer to work efficiently to prevent his future-self from withdrawing early. If Lear were fully naïve this policy would not work, because he would make the incorrect forecast about his future withdrawal decision.

The story of Lear suggests how making social security benefits more progressive with respect to later withdrawal dates (if Lear was non-sophisticated) and setting the liquidity of retirement accounts to depend on income history (if Lear was sophisticated) can help implement the efficient allocation. In the paper, I generalize Lear’s story to a three period quasi-hyperbolic discounting model, which includes varying degrees of naiveté and multiple levels of productivity.

1.2 Related Literature

There have been several papers examining the design of policy with behavioral agents. Farhi and Gabaix (2015) study optimal taxation (Ramsey, Pigou and Mirrlees) with behavioral agents by using sparse maximization (Gabaix 2014). They
are able to derive general results without specifying the bias, so it could potentially be applied to environments with agents who suffer from a wide array of behavioral biases. Lockwood (2015) extends Farhi and Gabaix (2015) to present biased agents, and finds that the optimal marginal tax rate could be negative, which is consistent with the EITC. Guo and Krause (2015) study a dynamic Mirrlees environment with sophisticated time-inconsistent agents where the government does not have full commitment. Bassi (2010) considers a dynamic Mirrlees environment with time-inconsistent agents where the quasi-hyperbolic discount factor is non-observable. Moser and de Souza e Silva (2015) consider a similar two-dimensional screening setup and decentralizes the optimum using social security and retirement accounts. These papers address similar issues as this one, but are unable to achieve the full information optimum. That is because, unlike the others, this paper takes advantage of time-inconsistent agents by expanding the set of contractible behavior to include off-equilibrium threats and promises. As a result, this paper delivers a different implementation in social security and retirement accounts.

In other related work, Diamond and Spinnewijn (2011) discuss a model with heterogeneity in both productivity and time preference (agents are time-consistent). Krusell et al. (2010) study the optimal taxation of consumers who suffer from temptation in a complete information environment. Amador et al. (2006) examine government policies for agents who suffer from temptation and are subject to future taste shocks. Halac and Yared (2014) apply a repeated model of Amador et al. (2006) with persistent shocks. Similar to this paper, Halac and Yared (2014) choose more relaxed policies that tempt the future-self to threaten and discipline present behavior. This can deter misreporting, but at the expense of exacerbating the present bias of certain types. In my setting, I construct off-equilibrium path policies that cause no distortions.

This paper is also related to several behavioral contracting papers. In particular, Esteban and Miyagawa (2005) examine optimal pricing schemes with time-inconsistent agents and find that distortions from information asymmetry can be averted when agents are tempted to over-consume. Bond and Sigurdsson (2015) demonstrate how off-equilibrium path options in commitment contracts can help time-inconsistent agents follow through with an ex-ante plan that accommodates their flexible needs. Eliaz and Spiegler (2006) examine a model with diversely naïve agents and found that firms can screen beliefs by bisecting the population into relatively
sophisticated and relatively naïve agents. Similar to this paper, they find relatively sophisticated agents exert no informational externality on the relatively naïve agents. Galperti (2015) extends Amador et al. (2006) to a sequential screening model where a mechanism designer first screens time consistency and then the taste shock. In this paper, the government screens both simultaneously.

The paper is organized as follows. Section 2 presents the model. Section 4 discusses a reform of the social security and retirement accounts. Section 3 presents the general mechanism. Section 5 considers the effects of time-consistent agents. Section 6 discusses some extensions and impediments to the mechanism and Section 7 concludes. The proofs are in Appendix A.

2 The Model

A continuum of agents of measure one live for three periods: \( t \in \{0, 1, 2\} \). There are \( |M| \geq 2 \) types of agents denoted by the set of productivity \( \Theta = \{\theta_1, \theta_2, \ldots, \theta_M\} \), with \( \theta_{m+1} > \theta_m \). The types are distributed according to \( \Pr(\theta = \theta_m) = \pi_m > 0 \), for all \( \theta_m \in \Theta \) with \( \sum_{m=1}^{M} \pi_m = 1 \).

The production technology is linear and depends only on labor input \( l_t \) and the productivity of the agent: \( y_t = \theta l_t \). Agents have access to a storage technology that transfers one unit of good in period \( t \) to one unit of period \( t+1 \) good. The government does not observe \( \theta \) and \( l_t \), but it observes \( y_t \).

The period utilities \( u_t : \mathbb{R}_+ \mapsto \mathbb{R} \) are continuously differentiable and \( u_t', -u_t'' > 0 \). The dis-utility from labor \( h_t : \mathbb{R}_+ \mapsto \mathbb{R} \) satisfies \( h_t', h_t'' > 0 \). Furthermore, \( u_t \) has full range (\( u_t(\mathbb{R}_+) = \mathbb{R} \)), which implies \( u_t \) is unbounded below and above. Single crossing is automatically satisfied.

The agents have the following utility before consumption at \( t = 0 \)

\[
U_0(c, y; \theta) = \left[ u_0(c_0) - h_0 \left( \frac{y_0}{\theta} \right) \right] + \delta \left[ u_1(c_1) - h_1 \left( \frac{y_1}{\theta} \right) \right] + \delta^2 u_2(c_2),
\]

where \( c_t \) is consumption in period \( t \). \( U_0 \) is the ex-ante utility for \( t = 0 \). I call the incarnation of the agent with the ex-ante utility as the planner. This is the utility agents use to evaluate their consumption plans. Analogously, the ex-ante utility for

---

9The word ‘planner’ refers to the farsighted incarnation of the agent, and should not be confused with the government.
\( t = 1 \) is \( U_1(c, y; \theta) = u_1(c_1) - h_1 \left( \frac{y}{\theta} \right) + \delta u_2(c_2) \). The utility changes when agents are consuming in period 0 to

\[
V_0(c, y; \theta) = \left[ u_0(c_0) - h_0 \left( \frac{y_0}{\theta} \right) \right] + \beta \left[ \delta \left( u_1(c_1) - h_1 \left( \frac{y_1}{\theta} \right) \right) \right] + \delta^2 u_2(c_2),
\]

à la Laibson (1997). \( V_0 \) is the ex-post utility for \( t = 0 \). The incarnation of the agent with the ex-post utility is called the doer. The ex-post utility for \( t = 1 \) is \( V_1(c, y; \theta) = u_1(c_1) - h_1 \left( \frac{y}{\theta} \right) + \beta \delta u_2(c_2) \).

I will focus on the case with present bias, where \( \beta \in (0, 1) \). \( \beta \) measures the degree of temptation the agents suffer from. Following O'Donoghue and Rabin (2001), partially naïve agents perceive their degree of present bias to be \( \hat{\beta} \in (\beta, 1) \). Let \( W_0(c, y; \theta) \) denote the non-sophisticated agents’ perceived ex-post utility in \( t = 0 \):

\[
W_0(c, y; \theta) = \left[ u_0(c_0) - h_0 \left( \frac{y_0}{\theta} \right) \right] + \hat{\beta} \left[ \delta \left( u_1(c_1) - h_1 \left( \frac{y_1}{\theta} \right) \right) \right] + \delta^2 u_2(c_2).
\]

\( W_1 \) is defined analogously. If \( \hat{\beta} = 1 \), the planner is fully naïve and unaware of the doer’s present bias. If \( \hat{\beta} = \beta \), the planner is sophisticated and aware of the doer’s bias. Partially naïve agents know they have present bias, \( \hat{\beta} < 1 \), but \( \hat{\beta} > \beta \). In other words, the planner underestimates the severity of the doer’s temptation problem. I will refer to \( \hat{\beta} \) as the sophistication level. Let \( \delta = 1 \), which does not affect the results of the paper.

The timing is as follows: At the beginning of \( t = 0 \), the government designs the tax system, and has full commitment. The agents then learn about their productivity. They proceed to work first, and then consume and save in \( t < 2 \). The agents retire in \( t = 2 \). In essence, the planner works and the doer makes consumption-savings decisions. The doer can also produce, which is a crucial assumption for the mechanism (Lear could work at night to withdraw early).

The government tries to help the agents commit to the ex-ante utility. Implicitly, I assume the non-sophisticated agents do not draw any inferences from the policies the government enacts. This is because they do not share the same prior as the government and are dogmatic in their beliefs. The government maximizes the following

---

10. The main idea of utilizing time inconsistency to raise welfare does not change if \( \beta > 1 \).
11. If \( \hat{\beta} < \beta \), it is still possible for the government to take advantage of the incorrect belief.
welfare criterion at \( t = 0 \)

\[
    \sum_{m=1}^{M} \pi_m [\kappa U_0 (c_m, y_m; \theta_m) + (1 - \kappa) V_0 (c_m, y_m; \theta_m)],
\]

(1)

where \((c_m, y_m)\) denotes the vector of allocations a \( \theta_m \) agent consumes and \( \kappa \in (0, 1] \) represents the welfare weight the government places on the planner\(^{12}\). Since both \( U \) and \( V \) are strictly increasing and concave in \((c, k)\), the government has a desire to insure agents against the realization of \( \theta \).

Finally, there are no private markets to insure against productivity shocks and no markets for illiquid assets or other commitment devices\(^{13}\).

2.1 The Benchmarks

2.1.1 No Private Information

In the benchmark, no private information case, the government maximizes social welfare \([1]\) subject to the feasibility constraint,

\[
    \sum_{t=0}^{2} \sum_{m=1}^{M} \pi_m (y_{m,t} - c_{m,t}) = 0,
\]

(2)

with \( y_{m,2} = 0 \) for all \( \theta_m \in \Theta \).

**Proposition 1** The efficient allocation \( \{(c^*_m, y^*_m)\}_{\theta_m \in \Theta} \) satisfies \( (2) \) and for any \( \theta_m \in \Theta \):

- (i.) full insurance: for any \( t, c^*_{m,t} = c^*_{t} \),
- (ii.) consumption smoothing: for any \( t > 0 \), \( u'_0 (c^*_m, 0) = [\kappa + (1 - \kappa) \beta] u'_t (c^*_m, t) \), and
- (iii.) efficient output: for any \( t < 2 \), \( u'_t (c^*_m, t) = \frac{1}{\bar{\theta}_m} h'_t (\frac{y^*_{m,t}}{\bar{\theta}_m}) \).

With complete information, the government achieves full insurance regardless of the agents’ time inconsistency or their sophistication level. This is because the agents work according to their productivity type. The government then chooses an appropriate linear savings subsidy to correct the distortion caused by the present bias.

\(^{12}\)Much of the literature on dynamically inconsistent preferences have evaluated welfare with \( \kappa = 1 \). With a Pareto criterion, as long as \( \kappa \in (0,1] \), the government wants to increase future consumption.

\(^{13}\)Prescott and Townsend (1984) show that an efficient insurance market can make distortionary taxes redundant. A market for commitment minimizes the role of paternalism.
2.1.2 Private Information without Time Inconsistency

With private information, the implementable allocations must be incentive compatible. The government maximizes social welfare \( U_0(c_m, y_m; \theta_m) \) subject to the feasibility constraint (2) and the incentive compatibility constraints,

\[
\forall \theta_m, \theta_m' \in \Theta, \\
U_0(c_m, y_m; \theta_m) \geq U_0(c_m', y_m'; \theta_m).
\] (3)

Due to (3), the government implements the constrained efficient optimum.

**Proposition 2** The constrained efficient allocation \( \{(c^{**}_m, y^{**}_m)\}_{\theta_m \in \Theta} \) satisfies (2), (3) and: (i.) partial insurance: for any \( t \) and \( \theta_m, \theta_m' \in \Theta \), with \( \theta_m > \theta_m' \), \( c^{**}_{m,t} > c^{**}_{m',t} \), (ii.) consumption smoothing: for any \( t \) and \( \theta_m \in \Theta \), \( u'_t(c^{**}_{m,t}) = u'_{t+1}(c^{**}_{m,t+1}) \), and (iii.) output distortions: for any \( t < 2 \) and \( \theta_m < \theta_M \), \( u'_t(c^{**}_{m,t}) > \frac{1}{\theta_m} h'_t(y^{**}_{m,t}/\theta_m) \).

The constrained efficient allocation distorts the labor decisions of all agents except for the most productive agents \( \theta_M \). This distortion relaxes the incentive compatibility constraint, which allows the government to provide partial insurance. Hence, Proposition 2 characterizes the optimal trade-off between efficiency and equity.

With sophisticated time-inconsistent agents, if the government uses a linear savings subsidy to correct the present bias, the constrained efficient optimum can still be obtained. The linear savings subsidy would correct the doer’s present bias without changing the set of incentive compatible allocations.

3 The General Mechanism

In this section, I will present a blueprint on designing policies that take advantage of time inconsistency.

The revelation principle holds, so the analysis will be in terms of a truth-telling direct mechanism. The government presents a menu \( C_m \) for type \( \theta_m \) agents defined as

\[ C_m = \{(c_m, y_m), (c'_m, y'_m), \ldots\} \].

An agent is assigned a menu \( C_m \) after reporting \( \theta_m \). Enlarged menus have been used to exploit time-inconsistent agents in the literature, see Esteban and Miyagawa (2005), Eliaz and Spiegler (2006) and Galperti (2015). Let \( (c^R_m, y^R_m) \in C_m \) be the real allocation, which is the optimal allocation the government can implement. The government posts \( C = \{C_1, \ldots, C_M\} \). The agents then choose a menu from \( C \) after learning \( \theta \). After temptation sets in, the agents choose an allocation
from the menu they selected. Non-sophisticated agents could potentially learn their
time inconsistency, as in Ali (2011). To prevent this, for all \( \theta_m \in \Theta \), set all allocations
in \( t = 0 \) equal to \( (c^R_m, y^R_m) \), so agents start facing an enlarged menu at \( t = 1 \).

Incentive compatibility is characterized by what the planner perceives the doer
will choose when he reports truthfully, and when he misreports. Let

\[
C^\hat{\beta}_m = \left\{ (c_m, y_m) \in C_m \mid (c_m, y_m) \in \arg\max_{(c_m', y_m') \in C_m} W_1(c_m', y_m'; \theta_m) \right\}.
\]

\( C^\hat{\beta}_m \) denotes the set of allocations a truthful \( \theta_m \) agent with sophistication level \( \hat{\beta} \)
predicts the doer would choose at \( t = 1 \). Let

\[
C^\hat{\beta}_{m|m'} = \left\{ (c_{m'}, y_{m'}) \in C_{m'} \mid (c_{m'}, y_{m'}) \in \arg\max_{(c_{m'}, y_{m'}) \in C_{m'}} W_1(c_{m'}, y_{m'}; \theta_m) \right\}.
\]

\( C^\hat{\beta}_{m|m'} \) denotes the set of allocations a type \( \theta_m \) agent with sophistication level \( \hat{\beta} \)
predicts the doer would choose after he misreports to be a type \( \theta_{m'} \) agent. Incentive
compatibility is thus expressed as, \( \forall \theta_m, \theta_{m'} \in \Theta \),

\[
\max_{(c_m, y_m) \in C^\hat{\beta}_m} U_0(c_m, y_m; \theta_m) \geq \max_{(c_{m'}, y_{m'}) \in C^\hat{\beta}_{m|m'}} U_0(c_{m'}, y_{m'}; \theta_m) \tag{4}
\]

The incentive compatibility constraints (4) make sure the agents choose the menu
that is intended for their productivity given their sophistication level \( \hat{\beta} \). Additional
constraints are needed to make sure the real allocations are implemented at \( t = 1 \).

The executability constraints are, \( \forall \theta_m \in \Theta \),

\[
(c^R_m, y^R_m) \in \arg\max_{(c_m, y_m) \in C_m} V_1(c_m, y_m; \theta_m) \tag{5}
\]

The executability constraints (5) make sure the doer would choose the real allocations.
With non-common priors, the mechanism has to consider the agents’ beliefs in the
incentive compatibility constraints and the government’s beliefs in the executability
constraints. Other allocations besides the real allocations are off-equilibrium path.
3.1 Savings with Non-sophisticated Agents

Non-sophisticated agents mispredict their future behavior. Hence, their reporting strategies reflect the expected choices of a fictitious doer, and not the real doer. The government can exploit this by using a betting mechanism.

**Definition 1** A direct betting mechanism for non-sophisticated agents ($\hat{\beta} \in (\beta, 1]$) has a menu $C = \{C_m\}_{\theta_m \in \Theta}$ with $C_m = \{(c^R_m, y^R_m), (c^I_m, y^I_m)\}$ and $(c^R_m, y^R_m) \neq (c^I_m, y^I_m)$ for some $\theta_m$ satisfying the fooling constraints: $(c^I_m, y^I_m) \in C^\hat{\beta}_m$ and $(c^I_{m'}, y^I_{m'}) \in C^\hat{\beta}_{m'}$, $\forall \theta_m, \theta_{m'} \in \Theta$.

Non-sophisticated agents of type $\theta_m$ predict they would choose $(c^I_m, y^I_m)$, which, following [Eliaz and Spiegler (2006)] and [Heidhues and Koszegi (2010)], will be referred to as the imaginary allocations. However, the government intends the agents to choose allocation $(c^R_m, y^R_m)$. I will set $y^R_m = y^I_m = y_m$, and show that $c^R_m \neq c^I_m$ is sufficient.

Planners make their reporting decisions based on $U(c, l)$ while anticipating an ex-post utility of $W(c, l)$. The government requires the imaginary allocations to be more appealing than the real allocations under $W(c, l)$, and truth-telling. To implement the real allocation, it is required to be more appealing than the imaginary allocation for the doer.

**Definition 2** An allocation $\{(c^R_m, y^R_m)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct betting mechanism if there exists $\{(c^I_m, y^I_m)\}_{\theta_m \in \Theta}$ such that the following are satisfied: (i.) incentive compatibility, and (ii.) executability.

The imaginary allocations are not required to satisfy the feasibility constraint. The government is certain about the degree of the naïveté and present bias of the agents, so it places no weight on a future where it may need to honor the delivery of imaginary allocations. Another concern is that the agents may realize that the aggregate imaginary allocation violates feasibility and doubt the validity of the government’s promise. However, each agent is infinitesimally small, and though an agent believes he would consume the imaginary allocation, he does not consider the belief and behavior of others. It can be shown the government achieves the efficient allocation by using a betting mechanism.

**Theorem 1** If $\hat{\beta} \in (\beta, 1]$, then the efficient allocation $\{(c^*_m, y^*_m)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct betting mechanism.
With the imaginary allocations, the government is able to provide the information rents necessary for truth-telling. However, these rents are fictional. After preference reversal, the government implements the efficient allocation without paying information rents. The key is to load the rents on retirement savings, $c_{R_{m}} \geq c_{m,2}$ and $c_{R_{m}} \geq c_{m,1}$, which the planner values more than the doer. The government can promise a higher return on savings for the imaginary allocations to elicit truthful reports as long as $\hat{\beta} \neq \beta$. In essence, non-sophisticated agents trade information rents for empty promises or for a sucker bet.

There is a discontinuity in the optimal welfare with respect to sophistication level. The government achieves the efficient welfare level for any $\hat{\beta} \in (\beta, 1]$. However, when $\hat{\beta} = \beta$, the best the betting mechanism can implement is the constrained efficient allocation which requires information rent for the productive types. This is because the sophisticated agents would perfectly foresee their doers’ present bias at the consumption stage and would thus be immune to the bet. Such discontinuities with respect to sophistication can also be found in Heidhues and Koszegi (2010).

3.2 Savings with Sophisticated Agents

For sophisticated agents, the government can design an off-equilibrium path option that exacerbates the temptation, which will be chosen only if an agent misreports. The off-equilibrium option disciplines the agent’s reports. This type of mechanism will be called a conditional commitment mechanism, since commitment is provided conditional on truth-telling, which is similar to the mechanism in Esteban and Miyagawa (2005).

Definition 3 A direct conditional commitment mechanism for sophisticated agents ($\hat{\beta} = \beta$) has $C = \{C_{m}\}_{\theta_m \in \Theta}$ with $C_m = \{(c_{R_{m}}, y_{R_{m}}), (c_{m}^T, y_{m}^T)\}$ and $(c_{m}^R, y_{m}^R) \neq (c_{m}^T, y_{m}^T)$ for some $\theta_m$ satisfying the threat constraints: $\forall \theta_m, \theta_{m'} \in \Theta, (c_{m}^T, y_{m}^T) \in C_{m' \mid m}$.

I will refer to $(c_{m}^T, y_{m}^T)$ as the threat allocation. Sophisticated agents have the correct belief about $\beta$, so the executability constraints imply that truth-telling is evaluated at the real allocation and the threat constraints imply that misreports are evaluated at the threat.

Definition 4 An allocation $\{(c_{m}^R, y_{m}^R)\}_{\theta_m \in \Theta}$ is truthfully implementable by a direct conditional commitment mechanism if there exists $\{(c_{m}^T, y_{m}^T)\}_{\theta_m \in \Theta}$ such that the following are satisfied: (i.) incentive compatibility, and (ii.) executability.
The threat, \((c_{m}^{T}, y_{m}^{T})\), for type \(\theta_{m}\) are designed such that, after preference reversal, a type \(\theta_{m}\) planner who reports truthfully would never choose it (by the executability constraint), but a planner who misreports as type \(\theta_{m}\) would (by the threat constraint). The threat allocation deters the agents from misreporting. A threat that satisfies the threat and executability constraints will be referred to as satisfying the credible threat constraints\(^{14}\).

Note that the threat does not work if \(y_{m}^{R} = y_{m}^{T}\). This is because all agents share the same preference for goods consumption, so a misreporting agent would make the same consumption choice as a truthful agent if \(y_{m}^{R} = y_{m}^{T}\). Hence, agents will not be deterred from misreporting. As a result, unlike a betting mechanism, the real allocations and threat allocations have to be different in both consumption and output. In an environment with sophisticated agents, the government can use a conditional commitment mechanism and attain the efficient optimum.

**Theorem 2** If \(\hat{\beta} = \beta\), then the efficient allocation \(\{(c_{m}^{*}, y_{m}^{*})\}_{\theta_{m} \in \Theta}\) is truthfully implementable by a direct conditional commitment mechanism.

Theorem 2 holds because sophisticated agents are aware that their doers would distort their savings plan, and would desire a commitment device that deters them from doing so. The government provides this commitment device only if the sophisticated agents report truthfully. If not, the threat allocation caters to the temptations of the doers. In essence, the government holds the agents’ doers hostage and threatens to distort the savings plan unless agents report truthfully. This also helps screening because the threat allocations can be designed such that it separates productivity by using the single crossing property without introducing any additional distortions. In essence, sophisticated agents trade information rents for commitment.

### 3.3 Savings with Partially Naïve Agents

Fully naïve agents have to be fooled, since they do not respond to threats. While sophisticated agents have to be threatened, because they can never be fooled. Since partially naïve agents also have demand for commitment, they are also susceptible to threats.

\(^{14}\)The government has full commitment, so the credible threat constraints are not to ensure subgame perfection. They address the need for the government to find tempting allocations that deter misreports and to avoid truthful agents from being tempted.
Definition 5 A direct conditional commitment mechanism has a menu \( C = \{ C_m \}_{\theta_m \in \Theta} \) with \( C_m = \{ (c_{Rm}^R, y_{Rm}^R), (c_{Tm}^T, y_{Tm}^T) \} \) and \( (c_{Rm}^R, y_{Rm}^R) \neq (c_{Tm}^T, y_{Tm}^T) \) for some \( \theta_m \) satisfying credible threat constraints: \( (c_{Rm}^R, y_{Rm}^R) \in C^\beta_m \) and \( (c_{Tm}^T, y_{Tm}^T) \in C^\beta_{m'} \), \( \forall \theta_m, \theta_{m'} \in \Theta \). The real allocation \( \{ (c_{Rm}^R, y_{Rm}^R) \}_{\theta_m \in \Theta} \) is truthfully implementable by a direct conditional commitment mechanism if there exists \( \{ (c_{Tm}^T, y_{Tm}^T) \}_{\theta_m \in \Theta} \) such that incentive compatibility (4) and executability constraints (5) are satisfied.

For \( \hat{\beta} \in (\beta, 1) \), the threats are evaluated using the erroneous \( \hat{\beta} \). Therefore, the credible threat constraints are defined to make sure the planner perceives his doer choosing the real allocation when report is truthful and the threat allocation if he mis-reported. The government also needs to make sure that after preference reversal, the truthful agents will indeed choose the real allocation, so the executability constraint is needed. In contrast to Definition 3, the constraint, \( (c_{Rm}^R, y_{Rm}^R) \in C^\beta_m \), is redundant when agents are sophisticated. The following corollary shows partially naive agents can be screened using a conditional commitment mechanism and Table 2 summarizes the applicability of the mechanisms to each sophistication level.

Corollary 1 It is optimal to use the conditional commitment mechanism when \( \hat{\beta} \in [\beta, 1) \).

\[
\begin{array}{ccc}
\text{Fully Naïve} & \text{Partially Naïve} & \text{Sophisticated} \\
\hline
\text{Betting} & \checkmark & \checkmark & \times \\
\text{Conditional Commitment} & \times & \checkmark & \checkmark \\
\end{array}
\]

Table 2: Summary of Mechanism for Different Sophistication Levels

3.4 Hidden Sophistication and Temptation with \( \beta < 1 \)

Consider an economy where all agents are time-inconsistent, but vary in temptation \( \beta \) and sophistication \( \hat{\beta} \), so types are represented by \( (\theta_m, \beta, \hat{\beta}) \), where \( \hat{\beta} \in [\beta, 1] \). Let \( B = [\beta, \bar{\beta}] \subset [0, 1] \) be the set of temptations, with \( \bar{\beta} > \beta \).

Lemma 1 For any \( \beta \in B \), betting mechanisms implementing the efficient allocation for \( \hat{\beta} \in (\beta, 1) \) also implement it for any \( \hat{\beta}' \geq \hat{\beta} \). Similarly, for any \( \beta \in B \), conditional commitment mechanisms implementing the efficient allocation for \( \hat{\beta} \in (\beta, 1) \) also implement it for any \( \hat{\beta}' \leq \hat{\beta} \).
Lemma 1 shows, for a known $\beta$, a betting (conditional commitment) mechanism designed for sophistication level $\hat{\beta}$ agents can also fool (threaten) agents who are more naïve (sophisticated). This is because fooling (threatening) the less naïve (sophisticated) agents for truth-telling is the more difficult, so any incentives that could separate the productivity of less naïve (sophisticated) agents will also induce truth-telling for more naïve (sophisticated) agents.

If $\beta$ is hidden, the government is unable to ascertain whether an agent with reported belief $\hat{\beta}$ is sophisticated or non-sophisticated. However, the government can choose $\hat{\beta} \in (\beta, 1)$ and design both conditional commitment and betting mechanisms for sophistication $\hat{\beta}$. Lemma 1 can be applied, because all agents have $\beta \leq \hat{\beta}$ and the temptation of the agents can be ignored. The subtlety lies in how the threats and bets are constructed to satisfy the executability constraints. For the imaginary (threat) allocations, the government needs to ensure $\beta = \tilde{\beta}$ ($\beta = \beta$) would prefer the real allocations, then all agents with stronger temptation $\beta < \tilde{\beta}$ (weaker temptation $\beta > \tilde{\beta}$) would strictly prefer the real allocation. The government chooses a fixed target sophistication level at $\hat{\beta} \in [\beta, 1)$ and introduce $C_m = \{(c^R_m, y^R_m), (c^I_m, y^I_m), (c^T_m, y^T_m)\}$, and menu $C = \{C_m\}_{\theta_m \in \Theta}$. The imaginary and threat allocations are chosen such that agents with $\hat{\beta}$ are indifferent between the betting and conditional commitment mechanisms. I will refer to this mechanism as a *hybrid* mechanism.

**Theorem 3** *For the environment where all agents are time-inconsistent with hidden temptation and sophistication, if $\kappa = 1$, the efficient allocation $\{(c^*_m, y^*_m)\}_{\theta_m \in \Theta}$ is truthfully implementable with a hybrid mechanism.*

Theorem 3 is robust to changes in the joint distribution of $(\theta_m, \beta, \hat{\beta})$. In addition to the primitives introduced in Section 2, the government does not need to know more than the support of $B$. Theorem 3 also demonstrates how the efficient allocation is implementable as long as all agents are time-inconsistent and the welfare criterion only takes the planner’s utility into consideration, regardless of sophistication or temptation. If $\kappa \in (0, 1)$, then the government cares about the doer’s utility and the government would need to screen the agents’ temptation, so Theorem 3 would no longer hold.
4 Social Security and Retirement Plans

Based on the mechanisms in Section 3, this section provides new perspectives on the social security and retirement plans reforms.

4.1 The Timing of Claiming Social Security Benefits

A majority of the US population relies on social security benefits as their primary source of income during retirement. Also, while the US population is living longer, the average retirement age has remained steady for the past decade. This increases the duration of relying on social security benefits. Consequently, discussions on social security reforms to improve retirement welfare while maintaining its sustainability is an important policy issue.

A retiree in the US can choose when they wish to start claiming social security benefits. The earliest age possible for receiving benefits is 62. A person can delay claiming and receive higher monthly benefits. Several papers have shown it is optimal for most people to delay benefits claiming, and that average Americans are receiving benefits too early (see Footnote 6). Knoll and Olsen (2014) find that the age of 62 is the most frequent enrollment age, and the age of 70 to be the least frequent. The NCFRR suggested the Social Security Administration to adopt behavioral economic approaches to encourage the US population to retire later and save more.

Knoll et al. (2015) show that people expect to retire and claim benefits later, but many end up retiring and claiming benefits earlier than they have initially planned. This suggests that time inconsistency with present-bias could explain the tendency to claim early. It also suggests that people are non-sophisticated. They devise effective choice architectures to delay claiming. I propose a new approach to this issue.

---

15 According to the Social Security Administration, nine out of ten individuals aged 65 or older receive social security benefits. Also, among the elderly beneficiaries, over half of the households receive over 50% or more of their income from social security.

16 See Munnell (2015).

17 For example, according to the Social Security Administration, the average monthly social security benefit for a beneficiary who started claiming at the age of 62 in 2014 is $1,098. If the same beneficiary waited till the age of 70 to start claiming benefits (the oldest enrollment age possible), then the monthly benefits would increase to $1,932.

18 Only 2% of the population choose to delay benefits till 70.

19 They showed process intervention (asking people to consider the benefits of delaying before considering the benefits of claiming early) can postpone enrollment by 9.4 months.
Given the benefits structure, the labor decisions of the agents are made according to the benefits received later. However, agents claim their benefits earlier than planned due to present bias. In other words, whether the government knows it or not, the benefits for claiming late affects the pre-retirement labor decision of agents, while the benefits for claiming early affects the retirement consumption of agents. Consequently, optimal social security reforms should take time-inconsistent behavior and non-sophistication as given when designing the benefits.

With less progressive benefits for agents who claim late, non-sophisticated agents can be encouraged to work efficiently at a younger age. Unknowingly, non-sophisticated agents would want to claim benefits earlier than expected. Therefore, with more progressive benefits for early claimants, redistribution can be increased without distorting labor supply. In essence, the imaginary allocations are the benefits for claiming late, and the real allocations are the benefits for claiming early. The government and the agents bet on when the agents would claim their benefits and the progressivity is the wager.

To be more concrete, consider the following social security policy: the agents work in $t = 0$, and decide whether to claim benefits $b_1(y_0, y_1)$ in $t = 1$ or to claim $b_2(y_0, y_1)$ when they retire in $t = 2$. Social security as a policy with tax $T_t$ is defined as $P^{ss} = (b_1(y_0, y_1), b_2(y_0, y_1), T_0(y_0), T_1(y_0, y_1, k_1))$. The budget constraint in $t = 0$ is standard: $c_0 + k_1 \leq y_0 - T_0(y_0)$, where $k$ denotes savings. In $t = 1$, agents choose to claim early at $t = 1$ or delay and claim at $t = 2$:

$$c_1 + k_2 \leq y_1 + k_1 + 1_1 b_1(y_0, y_1) - T_1(y_0, y_1, k_1),$$

where $1_t$ is an indicator function that is equal to 1 if benefits are claimed at $t$ and zero otherwise. In $t = 2$, the agents face the following budget constraint:

$$c_2 \leq k_2 + 1_1 b_1(y_0, y_1) + 1_2 b_2(y_0, y_1),$$

---

20 This paper does not encourage early retirement. Instead, it is proposing a social security program that achieves the optimum despite the fact that agents start claiming earlier than they had expected. It is possible to engineer the program such that agents retire and start claiming at the optimal age.

21 People can start claiming social security benefits at any age between 62 and 70. The model simplifies the decision by modeling it as a choice between early or late enrollment.

22 Though most people choose to claim benefits after they retire, it is still possible to claim benefits while working. For implementation, it is possible to include a penalty for retiring early. A key recommendation of the NCFRR is to raise the full retirement age.
so consumption in retirement depends on savings and benefits. In each $t$, the doer maximizes $V_t$ and the planner maximizes $U_t$ with the perception that his doer maximizes $W_t$.

**Proposition 3**  If $\hat{\beta} \in (\beta, 1]$, then the efficient allocation $\{(c^*_m, y^*_m)\}_{\theta_m \in \Theta}$ can be implemented by $P^{ss}$, where $b_2(y_0, y_1)$ is increasing and less progressive in income $y_0$ and $y_1$ than $b_1(y_0, y_1)$.

Proposition 3 shows how social security can implement the efficient allocation. The social security benefits $b_2$ is regressive in income to incentivize productive agents to produce efficiently. The benefits for early enrollment $b_1$ is a lump-sum transfer that provides full insurance and consumption smoothing for early retirees. Since the agents are non-sophisticated, they imagine claiming $b_2$ and would thus work efficiently in $t = 0$. However, the present-biased agents would claim $b_1$.

The current US system has benefits that are equally progressive in income for both early and late claimants. A reform along the lines proposed in Proposition 3 would make the benefits even more progressive for early claimants but less so for late claimants. Since the US population is already claiming earlier than planned, such a reform can help increase output efficiency, which would help increase taxable income and raise sustainability, while simultaneously improve social insurance. Both of which are goals set forth by the NCFRR.

### 4.1.1 An Example for Social Security

Consider the following example. Agents have $u_t(c_t) = \ln(c_t)$ and $h_t(l_t) = \frac{1}{2}l^2_t$, with $\beta = \frac{1}{2}$ and $\hat{\beta} = 1$. There are two productivity types, $\Theta = \{\theta_L, \theta_H\}$, with $\theta_L = \frac{1}{4}$ and $\theta_H = \frac{1}{3}$. The efficient allocation in this example is

$$c^*_{H,t} = c^*_{L,t} = c^* = \frac{5}{12\sqrt{3}} \approx 0.2406; \quad y^*_{L,t} = y^*_L = \frac{3\sqrt{3}}{20}; \quad y^*_{H,t} = y^*_H = \frac{4\sqrt{3}}{15}.$$  

Let $Y^* = \{y^*_L, y^*_H\}$. If $y_t \in Y^*$, let $y^{-1}_t \in \Theta$ denote the corresponding type, so $y^{-1}_m,t = \theta_m$. Let $EO = \{y = (y_0, y_1) | y_0, y_1 \in Y^*$ and $y^{-1}_0 = y^{-1}_1\}$ denote set of efficient

---

23Here, I do not consider the possibility the agent learns about his incorrect beliefs through the difference in planned savings and actual savings. Though in this setup, it is possible to shut down learning, which is shown in the example and the proof.
output history. Construct $P^{SS}$ with $T_0(y_0) = y_0 - c^*$, and $T_1(y_0, y_1, k_1) = y_1 + k_1 + 1_b(y_0, y_1)$ and $T_2(y_0, y_1, c^* - c^*)$, and benefits:

$$b_1(y_0, y_1) = \begin{cases} c^* & \text{if } y \in EO \\ 0 & \text{otherwise} \end{cases}, \quad b_2(y_0, y_1) = \begin{cases} 0.7415 & \text{if } y_t = y^*_H, \text{ for all } t = 0, 1 \\ c^* & \text{if } y_t = y^*_L, \text{ for all } t=0, 1 \\ 0 & \text{otherwise} \end{cases}$$

with $\alpha(y_0, y_1) = c^* - 0.137$, if $y_t^{-1} = \theta_H$ for all $t < 2$, otherwise $\alpha(y_0, y_1) = c^*$.

First, agents always choose $y \in EO$, otherwise they do not receive benefits in $t = 2$. Also, all agents choose $k_t = 0$ and consume $c^*$ at $t = 0$, which shuts down the possibility of learning about their bias before $t = 1$. Agents who produce $y^*_L$ for both periods would prefer claiming at $t = 1$. Next, $\theta_H$ agents would work efficiently in $t = 0$. This is because they expect to claim higher benefits at $t = 2$:

$$\ln(0.137) - \frac{1}{2} \left(y^*_H \theta_H\right)^2 + \hat{\beta} \ln(0.7415) > \ln(c^*) - \frac{1}{2} \left(y^*_H \theta_H\right)^2 + \hat{\beta} \ln(c^*),$$

and the benefits are constructed such that incentive compatibility is satisfied:

$$\ln(c^*) - \frac{1}{2} \left(y^*_H \theta_H\right)^2 + \ln(0.137) - \frac{1}{2} \left(y^*_H \theta_H\right)^2 + \ln(0.7415)$$

$$> \ln(c^*) - \frac{1}{2} \left(y^*_L \theta_H\right)^2 + \ln(c^*) - \frac{1}{2} \left(y^*_L \theta_H\right)^2 + \ln(c^*).$$

The caveat is the $\theta_H$ agents would never claim at $t = 2$, and would instead claim at $t = 1$, because the future-self prefers higher present consumption,

$$\ln(c^*) - \frac{1}{2} \left(y^*_H \theta_H\right)^2 + \beta \ln(c^*) > \ln(0.137) - \frac{1}{2} \left(y^*_H \theta_H\right)^2 + \beta \ln(0.7415).$$

Recall that $c^* \approx 0.2406 > 0.137$. Finally, it is easy to check that $\theta_L$ agents would never produce $y^*_H$ in exchange for claiming higher benefits at $t = 2$. Hence, this social security reform implements the efficient allocation. A graphical explanation is provided in Figure 1.

In Figure 1, the flatter solid (blue) curve represents the indifference curve of the ex-ante utility at allocation $(c^*_1, c^*_2)$. The steeper solid (red) curve represents the indifference curve of the ex-post utility at allocation $(c^*_1, c^*_2)$. The perceived consumption
for delayed claiming have to be in the area bounded by the solid indifference curves in the north-west region to induce the doer to claim early rather than late. Furthermore, the incentive compatibility constraints for both types of agents provide upper and lower bounds to the difference in ex-ante utility of the two types of agents. In essence,

$$
\sum_{t=0}^{1} \left[ h_t \left( \frac{y^*_H}{\theta_L} \right) - h_t \left( \frac{y^*_L}{\theta_H} \right) \right] \geq \left[ u_1(c^*_H) + u_2(c^*_H) \right] - \left[ u_1(c^*_1) + u_2(c^*_2) \right] \geq \sum_{t=0}^{1} \left[ h_t \left( \frac{y^*_H}{\theta_L} \right) - h_t \left( \frac{y^*_L}{\theta_H} \right) \right].
$$

Therefore, given $y^*_m$, the perceived consumption from delayed claiming have to be within the dashed indifference curves to induce efficient output.

The benefits are less progressive if they are claimed at $t = 2$ than if they were claimed at $t = 1$. Since the ratio of the lifetime benefits to lifetime payroll taxes for $\theta_L$ agents who claimed at $t = 1$ is $R_{L,1} = \frac{2e^*}{2y^*_L - e}$, and the ratio for claiming at $t = 2$ is $R_{L,2} = \frac{e^*}{2y^*_L - e - 0.137}$. For $\theta_H$ agents, the ratio at $t = 1$ is $R_{H,1} = \frac{2e^*}{2y^*_H - e}$, and for claiming at $t = 2$, the ratio is $R_{H,2} = \frac{0.7415}{2y^*_L - e - 0.137}$. It can be shown that $R_{L,1} - R_{H,1} \approx 1.02$ and $R_{L,2} - R_{H,2} \approx 0.3359$, so the benefits are less progressive with withdrawal age.

This demonstrates how the optimal social security benefits for non-sophisticated
agents loads its information rents on less progressive benefits for late claiming ages. This induces the agents to work efficiently, incorrectly anticipating their future-selves collecting large benefits. However, their future-selves prefer immediate satisfaction and would claim benefits early, which are more progressive and help raise social insurance.

4.2 Liquidity of Defined Contribution Plans

The design of defined contribution (DC) plans is of growing interest. The literature has focused on how to influence DC plan enrollment behavior. Other aspects of the design of DC plans has also gained attention. In particular, Beshears et al. (2015b) showed the DC plans in the US to be very liquid. After separating from their employer, workers in the US can move their DC account balance to an IRA or Roth IRA and withdraw for any reason from the account before the eligibility age of 59.5 subject to a tax penalty of 10%. Such liquidation before eligibility is forbidden in many countries, except under special circumstances, such as a debilitating injury. Consequently, DC plans in the US are very flexible and meet the transitory needs of a worker. However, flexibility is undesirable if early withdrawal is due to present bias.

While making the DC plan more illiquid for time-inconsistent agents can be an effective commitment device, this paper provides an alternative view: such commitment should not be provided for free. The liquidity of DC plans can be redesigned to depend on income and act as a threat to sophisticated time-inconsistent agents, which incentivize agents to produce efficiently and allow the government to provide better insurance. The conditional commitment mechanism can be implemented with early withdrawal as the off-equilibrium threat.

To implement full efficiency with retirement accounts, consider the following timing: agents are endowed with $s_0 > 0$ in the accounts and work and deposit $s_{t+1}$ into their accounts in $t$, and are allowed to withdraw $\eta \in (0, s_0)$ early from their accounts and receive $\xi(y_0, s_1) - \eta$ as a subsidy in $t = 1$. Let $\tau(y_0, y_1, s_1)$ be the early withdrawal penalty, which is an off-equilibrium path threat contingent on income history.

\[\text{\footnotesize See Madrian and Shea (2001), Thaler and Benartzi (2004) and Carroll et al. (2009).}\]
\[\text{\footnotesize For example, countries such as Germany, Singapore and the UK.}\]
\[\text{\footnotesize For example, see Ashraf et al. (2006) and Beshears et al. (2015a).}\]
\[\text{\footnotesize For simplicity, I do not consider liquid savings accounts, like bank savings.}\]
A retirement account with contemporaneous income tax $T_t$ and savings subsidy $\rho_t$ is defined as $\mathcal{P}^{ra} = (s_0, \tau(y_0, y_1, s_1), \xi(y_0, s_1), \eta, \rho_1, \rho_2, T_0(y_0), T_1(y_1))$.

In $t=0$, the budget constraint is: $c_0 + \frac{s_1}{1 + \rho_1} \leq y_0 - T_0(y_0)$. In $t=1$, agents choose whether to withdraw early from the retirement account:

$$c_1 + \frac{s_2}{1 + \rho_2} \leq 1_{EW} \xi(y_0, s_1) + y_1 - T_1(y_1),$$

where $1_{EW}$ is equal to 1 if the agent withdrew early and zero otherwise. In $t=2$, agents face the following budget constraint:

$$c_2 \leq 1_{EW} (1 - \tau(y_0, y_1, s_1)) (s_0 + s_1 + s_2 - \eta) + (1 - 1_{EW}) (s_0 + s_1 + s_2).$$

Inequality (6) shows that, with early withdrawal, $c_2$ decreases proportionally to the penalty $\tau(y_0, y_1, s_1)$, where

$$\tau(y_0, y_1, s_1) = \begin{cases} \hat{\rho}(y_0, s_1) & \text{if } y_1 \geq \bar{y}(y_0, s_1) \\ 1 & \text{if } y_1 < \bar{y}(y_0, s_1) \end{cases}.$$ 

The penalty is structured so that if $y_1$ is commensurate with $y_0$ (i.e. $y_1 < \bar{y}(y_0, s_1)$), then the agent would not be tempted to withdraw from the retirement account (because, if he does withdraw early, then $c_2 = 0$). Therefore, the retirement account is illiquid and the agent is committed to having sufficient savings for retirement. If $y_1$ is much greater than $y_0$ (at least $\bar{y}(y_1, s_1)$), then the present-biased agent will withdraw early and be penalized by having lower retirement savings. Only agents who produced an inefficiently low output in $t=0$ with respect to their productivity would be tempted to withdraw early in $t=1$. Consequently, agents would produce efficiently in $t=0$ to avoid the temptation of withdrawing early in $t=1$. Here, the implementation of the conditional commitment mechanism uses the allocation from early withdrawal as the threat allocation.

**Proposition 4** If $\hat{\beta} \in [\beta, 1)$, then the efficient allocation $\{(c^*_m, y^*_m)\}_{\theta_m \in \Theta}$ can be implemented by $\mathcal{P}^{ra}$ for some $\tilde{y}(y_0, s_1) > y_0$.

Current retirement plans in the US are more liquid the lower the income, which is also what Proposition 4 proposes. However, there are two main differences between current plans and the plan proposed in Proposition 4. First, liquidity in the proposed
retirement plan is not indiscriminately available for all low income agents. To be able to withdraw early, an agent must earn a sufficiently larger income than the previous period to qualify. Another difference is that in Proposition 4, the early withdrawal penalty tax $\tau$ is applied to the residual amount left in the savings account. This penalty tax decreases the savings available in the retirement account to discipline the younger self. The current system has the early withdrawal penalty tax on the withdrawal amount. Though the current system discourages early withdrawals and helps smooth consumption, it does not have the disciplining effect on the younger-self to encourage efficient output.

4.2.1 An Example for Retirement Accounts

Consider the environment in Section 4.1.1. Assume the agents are fully sophisticated $\hat{\beta} = \beta$ and are not allowed to make deposits at $t = 0$, so $s_1 = 0$ or $\rho_1 = -1$. Let $\rho_2 = \frac{1}{\beta} - 1$ and $s_0 \in (0, c^*)$ and taxes be

$$ T_0 (y_0) = \begin{cases} y_0 - c^* & \text{if } y_0 \geq y_L^* \\ y_0 & \text{otherwise} \end{cases} , \quad T_1 (y_1) = \begin{cases} y_1 - c^* - \beta (c^* - s_0) & \text{if } y_1 \geq y_L^* \\ y_1 & \text{otherwise} \end{cases} . $$

With this setup, on-equilibrium path ($1_{EW} = 0$), the agents would consume the efficient consumption levels. Next, consider the off-equilibrium policy that supports efficient output on-equilibrium path. Suppose a $\theta_H$ agent produces $y_L^*$ (given $T_0$, there are no additional benefits to deviating to $y_0 \in (y_L^*, y_H^*)$), then he predicts his doer to solve the following if $y_1 < \bar{y} (y_0) : \max_{c_1, c_2, y_1} \ln (c_1) - \frac{1}{2} \left( \frac{y_1}{\theta_H} \right)^2 + \beta \ln (c_2)$ subject to $c_1 + \beta (c_2 - s_0) \leq y_1 - T_1 (y_1)$. Hence, the agent would produce $y_L^*$ and consume $c^*$ in both periods. If $y_1 = \bar{y} (y_0)$, the agent solves: $\max_{c_1, c_2} \ln (c_1) + \beta \ln (c_2)$ subject to $c_1 + \beta \left( \frac{c^*}{1 - \frac{c^*}{\theta_H(y_0)}} - s_0 + \eta \right) \leq \xi (y_0) + y_1 - T_1 (y_1)$, which yields the early withdrawal utility of

$$ \ln (\hat{c}_{L,1}) - \frac{1}{2} \left( \frac{\bar{y}}{\theta_H} \right)^2 + \beta \ln (\hat{c}_{L,2}) , $$

where $\hat{c}_{L,1} = \frac{2}{3} \left( \xi - \frac{1}{2} \eta \right) + c^*$ and $\hat{c}_{L,2} = (1 - \hat{\rho}) \left( \frac{2}{3} \left( \xi - \frac{1}{2} \eta \right) + c^* \right)$. Since there are only two types, for notational ease, I have denoted the policy variables without indicating their dependency on income. For $\theta_H$ agents who produced inefficiently in $t = 0$ to

\footnote{It is also possible to implement this as a repayment scheme where the agents are required to replenish their accounts by a large amount if they withdrew early.}
prefer withdrawing early, it must be that:

$$\ln \left( c_{L,1}^T \right) - \frac{1}{2} \left( \frac{\bar{y}}{\theta_H} \right)^2 + \beta \ln \left( c_{L,2}^T \right) \geq \ln (c^*) - \frac{1}{2} \left( \frac{y_L^*}{\theta_H} \right)^2 + \beta \ln (c^*).$$

To deter the $\theta_H$ agents from producing inefficiently in $t = 0$, it must be that:

$$\ln (c^*) - \frac{1}{2} \left( \frac{y_H^*}{\theta_H} \right)^2 + \ln (c^*) \geq \ln (c^*) - \frac{1}{2} \left( \frac{\bar{y}}{\theta_H} \right)^2 + \ln (c_{L,1}^T).$$  \hspace{1cm} (7)$$

Also, $\bar{y}$ is sufficiently high so the $\theta_L$ agents never withdraw early:

$$\ln (c^*) - \frac{1}{2} \left( \frac{y_L^*}{\theta_L} \right)^2 + \beta \ln (c^*) \geq \ln (c_{L,1}^T) - \frac{1}{2} \left( \frac{\bar{y}}{\theta_L} \right)^2 + \beta \ln (c_{L,2}^T).$$  \hspace{1cm} (8)$$

Consider the following early withdrawal policy: $\eta = 0.001, \xi = 4.1907, \hat{\rho} = 0.9945, \bar{y} = \sqrt{3}/3$. The policy for early withdrawal satisfies the inequalities above and thus implements the efficient allocation. Notice $\bar{y}$ is set slightly higher than $y_H^*$ so the $\theta_L$ agent would not find it optimal to withdraw early, but the $\theta_H$ agent would. Hence, $y_{L,1}^T = \bar{y}$ for the conditional commitment mechanism. To make early withdrawal appealing to the doer at $t = 1$, $\xi$ is high. To make early withdrawal unappealing to the planner at $t = 0$, the penalty $\hat{\rho}$ is high. The discrepancy between consumption at $t = 1$ and $t = 2$ for early withdrawals takes advantage of the disagreement between the planner and the doer. Early withdrawals tempt the doer against the will of the planner, and the only way for the planner to commit is to produce efficiently at $t = 0$. A graphical explanation is provided in Figure 2.

Figure 2 shows the consequences of early withdrawal so $\theta_H$ agents works efficiently. Let $\Phi_{j,k}^* = u_1(c_{j,1}^*) - h_1 \left( \frac{y_{j,1}}{\theta_k} \right)$ and $\Phi_{j,k}^T = u_1(c_{j,1}^T) - h_1 \left( \frac{\bar{y}}{\theta_k} \right)$, where $j, k \in \{L, H\}$. For notational ease, let $\Phi_{k,k}^* = \Phi_{k}^*$ and $\Phi_{k,k}^T = \Phi_{k}^T$. The steeper solid (red) curve represents the indifference curve of the ex-post utility at $t = 1$ for the $\theta_H$ agent who produced $y_H^*$ in $t = 0$. The flatter solid (blue) curve represents the indifference curve of the ex-ante utility at $t = 1$ for the $\theta_H$ agent who produced $y_H^*$ in $t = 0$. The dashed (red) curve represents the indifference curve of the ex-post utility for the efficient $\theta_L$ agent. Figure 2 shows that the government can choose off-equilibrium path policies such that inequality (7) is satisfied.
By single crossing, the government can choose $\bar{y}$ and $c_{L,1}^T$ in a way that the $\theta_L$ agents would never withdraw early. To see this, fix the choice of $c_{L,2}^T$ and $\Phi_{L,H}^T$ at the level shown in Figure 2. Inequalities (7) and (8) imply

$$\Delta u_2 \geq \Phi_{L,H}^T - \Phi_H^* + \Delta h_0 > \Phi_{L,H}^T - \Phi_{L,H}^* \geq \beta \Delta u_2 \geq \Phi_T^L - \Phi_L^*,$$

where $\Delta u_2 \equiv [u_2(c_2^*) - u_2(c_{L,2}^T)]$ and $\Delta h_0 = h_0 \left( \frac{\bar{y}_{H,0}}{s_{H}} \right) - h_0 \left( \frac{\bar{y}_{L,0}}{s_{H}} \right)$. The problem now is to find $c_{L,1}^T$ and $\bar{y}$ such that $u_1(c_{L,1}^T) - h_1 \left( \frac{\bar{y}}{s_{H}} \right) = \Phi_{L,H}^T$ and $\beta \Delta u_2 \geq \Phi_T^L - \Phi_L^*$. In essence, $(c_{L,1}^T, \bar{y})$ cannot be too tempting as to also lure the $\theta_L$ agents into withdrawing early.

In Figure 3, the flatter thick solid (blue) curve represents the indifference curve of $\Phi$ for the $\theta_H$ agents at allocation $(c_{1}^*, \bar{y}_{L,1}^*)$. The steeper solid (red) curve represents the indifference curve of $\Phi$ for the $\theta_L$ agents at allocation $(c_{1}^*, \bar{y}_{L,1}^*)$. The dashed (blue) curve represents the indifference curve of $\Phi$ for the $\theta_H$ agent at allocation $(c_{L,1}^T, \bar{y})$, chosen so that $u_1(c_{L,1}^T) - h_1 \left( \frac{\bar{y}}{s_{H}} \right) = \Phi_{L,H}^T$ and $\beta \Delta u_2 \geq \Phi_T^L - \Phi_L^*$.

### 4.3 Implementing the Hybrid Mechanism

Consider an environment where all agents are time-inconsistent, but sophistication and temptation are private information. By Section 3, the efficient allocation is
implementable by a hybrid mechanism. To implement the hybrid mechanism with social security and defined contribution plans, suppose the government endows the agents an initial savings of \(s_0 > 0\) in their DC accounts at the beginning of \(t = 0\). The agents can still deposit \(s_t\) in their DC accounts, which maintain the same penalty features introduced previously. The social security benefits still have the original structure in terms of progressivity, but the benefits are decreased by \(s_0\). The relatively sophisticated agents would be threatened by the liquid DC account if they worked inefficiently, while the relatively naïve agents would work efficiently due to the higher social security benefits from delay. The initial savings \(s_0\) is necessary, because the threat of a liquid DC account is potent only if there were funds in the account. If \(s_0 = 0\), a relatively sophisticated agent can always work inefficiently, choose a low \(s_1\) and claim social security benefits early to mimic relatively naïve agents.

An example of such an implementation is the central provident fund (CPF), which is a retirement account in Singapore. All conscripts are endowed with at least 5000 SGD in their CPF accounts after military service. Since national service is compulsory for all males in Singapore, this endowment is similar to the initial endowment of \(s_0\). However, the difference between the CPF and my proposal is that early withdrawal from the CPF account is not allowed.\footnote{Singaporeans are allowed to withdraw money from their CPF accounts for housing and education, but must repay it with interest. Medical expenses are exempt from repayment.}

---

28
Moser and de Souza e Silva (2015) analyzes a government screening both temptation and productivity. However, they do not consider off-equilibrium path policies and arrive at a different conclusion on the design of social security and defined contribution plans. Information rents are provided to more productive agents in the form of flexible savings plans like a defined contribution plan, while social security is less flexible and is for less productive agents. This runs counter to my implementation, where social security and defined contribution plans are used to separate sophistication and not productivity.

5 Model with Time-Consistent Agents

In this section, I will discuss the consequences of introducing time-consistent (TC) agents. The government does not observe whether agents are time-inconsistent (TI) or TC. I assume that \( \kappa = 1 \).

The environment with TC agents limits the ability of the government to bet with or threaten the agents. TC agents cause distortions, because they can follow through with their consumption plans and choose the imaginary allocation and avoid the threat allocation. This restricts the ability of the government to provide insurance, so the efficient optimum is no longer attainable. Here, I discuss the main results, and details are in the online appendix.\(^{30}\)

5.0.1 When \( \hat{\beta} < 1 \)

If all TI agents are not fully naïve (\( \hat{\beta} < 1 \)), it is possible to separate the TI agents from the TC agents. Suppose the government chooses to use a conditional commitment mechanism, then it can design the following menu for TC agents:

\[
C_{m}^{TC} = \left\{ (c_{P}^{m}, y_{P}^{m}); (c_{D}^{m}, y_{D}^{m}) \right\}
\]

and the following menu for TI agents: \( C_{m}^{TI} = \left\{ (c_{R}^{m}, y_{R}^{m}); (c_{T}^{m}, y_{T}^{m}) \right\} \). The allocation \( (c_{m}^{P}, y_{m}^{P}) \) is the persistent allocation, and it is the allocation the government implements for \( \theta_{m} \) type TC agents. The allocation \( (c_{D}^{m}, y_{D}^{m}) \) is referred to as the deterrent allocation, and it is meant to deter the TI agents from misreporting as \( \theta_{m} \) type TC agents.

agents. The deterrent allocation tempts the doer, which deters the TI agents from pretending to be TC. The threat and deterrent allocations leave no information rents for TI agents. The government offers $C = \{C^{TC}_m, C^{TI}_m\}_{\theta_m \in \Theta}$. The mechanism is meant to separate agents along two dimensions: productivity and consistency. Betting mechanisms can be defined analogously with imaginary allocations in $C^{TI}_m$.

In this environment, both the betting and conditional commitment mechanisms can achieve the same constrained efficient allocations and welfare. In particular, welfare increases with the population of TC agents, as shown in Figure 4. This is because the government can exploit TI agents by making them work more and consume less than their TC counterparts. Hence, as the proportion of TI agents increases, the government can provide more information rent per TC agent without causing more distortions.

5.0.2 When $\hat{\beta} = 1$

Separating consistency is not possible when $\hat{\beta} = 1$. If $\hat{\beta} = 1$, the government introduces the following menu at date $t = 0$, $C = \{C_m\}_{\theta_m \in \Theta}$, with $C_m = \{(c^R_m, y^R_m), (c^P_m, y^P_m)\}$. The government expects agents of the same productivity to pick the same menu and TI agents to choose the real allocations and TC agents to choose the persistent allocations. The persistent allocations serve a similar purpose as the imaginary allocations to the TI agents. TI agents evaluate their reporting strategies using the persistent allocations, but consume the real allocations.
The best the government can do is to require the TC agents to over-save and the TI agents to under-save. This is true even for the most productive agents, so there will be distortions at the top. More importantly, due to the government’s inability to separate TI from TC agents, the welfare when \( \hat{\beta} = 1 \) is always strictly less than the welfare when \( \hat{\beta} < 1 \). This has new implications on the education policy for TI agents. The government would prefer fully naïve agents to be at least partially aware of their temptation. Meanwhile, the government is indifferent between other levels of sophistication.

6 Discussion

6.1 Dynamic Stochastic Productivity Shocks

With stochastic productivity shocks, the betting and conditional commitment mechanisms can be implemented, as long as the planner is the one who learns the private information. However, betting mechanisms are more subtle. In each period, planners use a different imaginary promised utility than doers to make intertemporal decisions: the promised utility for planners is higher than the one for doers. Hence, betting mechanisms are dynamic, since the planners for the next period would use the imaginary allocations contained in the promised utility for the previous period’s doers. Conditional commitment mechanisms do not have this feature. With conditional commitment mechanisms, the dynamic taxation problem only needs to keep track of the history of reports. In fact, if productivity shocks are independent, the government can implement the same mechanism every period. In this case, the dynamic taxation problem is essentially static.

Another concern is learning. In betting mechanisms, the doers consume differently from what the planners expected to consume. This could trigger learning and the agents could become more sophisticated, see Ali (2011). Therefore, in dynamic betting mechanisms, the evolving sophistication of agents also needs to be considered. On the other hand, conditional commitment mechanisms do not have such concerns. This is because conditional commitment mechanisms confirm the planners’ beliefs about the bias, regardless of its accuracy.
6.2 Commitment vs. Flexibility

Recent works have highlighted the trade-off between commitment and flexibility. In these models, agents suffer from temptation but would also like to accommodate their intertemporal taste shock. The demand for commitment comes from temptation. The demand for flexibility comes from the future taste shock. In the context of this paper, this is a sequential screening problem. The government asks agents to report productivity first, and then the taste shock.

If temptation \( \beta \) is observable, then after learning the productivity, the government can implement a savings dependent transfer \( T_t(k_{t+1}) \). By Galperti (2015), an optimally chosen \( T_t(k_{t+1}) \), with \( T_t \) strictly increasing and \( T_t(k_{t+1}) > 0 \) above some threshold \( k_{t+1} \) and \( T_t(k_{t+1}) \leq 0 \) below it, can help attain the efficient allocation. This is because the transfer aligns the incentives of the planner and the doer. More importantly, since a betting or conditional commitment mechanism elicits all of the private information from the agent before the taste shock without any cost, the government can use the savings dependent transfer to ensure the doer saves according to the realized taste shock.

However, if temptation is unobservable by the government, then the optimal allocations may not only be distorted in output, as in Section 5, but also intertemporally, as in Galperti (2015).

6.3 Impediments to Implementation

I have shown how the presence of time-consistent agents cause distortions. Here, I discuss other situations where full efficiency may not be achievable.

6.3.1 Limited Promises and Punishments

The assumption that \( u_t \) is unbounded below and above provides a non-empty set of bets that could deceive non-sophisticated agents for any information rent. Consider the case where the utility function is bounded below. Figure 5 illustrates how betting can be limited for fully naïve agents in the two types example. The flatter solid (blue) curve represents the indifference curve of the ex-ante utility and the steeper solid (red) curve represents the indifference curve of the ex-post utility, both evaluated

---

at allocation \((c_1^*, c_2^*)\). The dotted (blue) curve indicates the minimum information rent necessary for the productive agents to be truthful. However, the best the government can do is to set the imaginary allocation at the boundary as indicated in Figure 5. Hence, asymmetric information would cause distortions. A similar argument can be made for conditional commitment mechanisms. Figure 6 illustrates how threats are limited in a sophisticated case with two productivity types.

Despite these limitations, the government is still able to improve welfare above the constrained efficient optimum by using the mechanisms introduced.

### 6.3.2 Outside Commitment Devices

In reality, self-control problems can be mitigated by a wide array of commitment devices available in the market. In the case of sophisticated agents, if commitment devices are available and its usage is unobservable, then threats become much less potent. This is because an agent can purchase illiquid assets and bind himself to an intertemporal allocation. This reduces the effectiveness of a threat, because the ex-post utility is maximized over a smaller consumption set. Therefore, private information could matter when commitment devices for sophisticated or partially naïve agents are available. However, for non-sophisticated agents, the government can always choose imaginary allocations that make buying an outside commitment device

---

32 There is a growing market for commitment devices. For example, StickK, Pact and Beeminder are some recent websites that offer commitment contracts.
7 Summary and Conclusion

This paper provided methods on utilizing the agents’ time inconsistency to increase welfare above the constrained efficient optimum, contrary to traditional policy proposals, where the primary goal was to mitigate the present bias. These methods are provide new insights on the progressivity of social security benefits and the liquidity of defined contribution plans.

The results of this paper could be applied to other settings, like the design of health or life insurance policies. The concept of betting and issuing threats could potentially be used in a wider array of mechanism design problems with agents suffering from other biases, such as overconfidence.

Though welfare increases with the proportion of time-inconsistent agents in the economy, this paper is not meant to advocate time-inconsistent behavior. The focus on savings has obscured other costs associated with being time-inconsistent, such as inadequate human capital development. Future work should explore this trade-off and its consequences on policy.

A Proofs

Proof of Theorem 1: First, set $\begin{pmatrix} R_m, y_m \end{pmatrix} = \begin{pmatrix} R^*_m, y_m \end{pmatrix}$ and $\begin{pmatrix} I_m, 0 \end{pmatrix} = \begin{pmatrix} I^*_m, 0 \end{pmatrix}$ for all $\theta_m \in \Theta$. Next, notice the following programming problem has no finite solution: for all $\theta_m \in \Theta$,

$$\max_{c_{m,1}, c_{m,2}} u_1 (c_{m,1}) + \beta u_2 (c_{m,2}),$$

subject to $u_1 (c_{m,1}) + \beta u_2 (c_{m,2}) \leq u_1 (c^*_m) + \beta u_2 (c^*_m)$, This follows from $\hat{\beta} > \beta$ and $u_2$ being unbounded above and $u_1$ unbounded below. Therefore, the executability and fooling constraints can be satisfied. Finally, notice that local downward incentive compatibility implies global incentive compatibility. Since the programming problem above has no finite solution, the imaginary allocations can always be chosen so the

---

33This has the additional benefit of preventing the non-sophisticated agents from using an inefficient amount of commitment (Heidhues and Koszegi (2009)).
incentive compatibility constraints are satisfied for any positive information rent. □

**Proof of Theorem 2** First, set \((c^R_m, y^R_m) = (c^*_m, y^*_m)\) and \((c^T_{m,0}, y^T_{m,0}) = (c^*_m, y^*_m)\) for all \(\theta_m \in \Theta\). Next, notice the following programming problem has no finite solution:

for all \(\theta_m \in \Theta\),

\[
\min_{c^{T}_{m,1}, c^{T}_{m,2}, y^{T}_{m,1}} u_1(c^{T}_{m,1}) - h_1 \left( \frac{y^{T}_{m,1}}{\theta_m} \right) + \beta u_2 \left( c^{T}_{m,2} \right)
\]

subject to

\[
u_1(c^{T}_{m,1}) - h_1 \left( \frac{y^{T}_{m,1}}{\theta_{m+1}} \right) + \beta u_2 \left( c^{T}_{m,2} \right) = u_1(c^*_m) - h_1 \left( \frac{y^*_m}{\theta_{m+1}} \right) + \beta u_2 \left( c^*_m \right).
\]

This is because \(h_1\) is strictly increasing and convex and \(\theta_{m+1} > \theta_m\) and \(u_1\) is unbounded above. Note that if \(y^{T}_{m,1} > y^*_m\) and \(9\) holds, then the threat allocation is more appealing for all agents more productive than \(\theta_{m+1}\). This implies the credible threat constraints are satisfied.

Suppose \(9\) holds and by Proposition the local downward incentive compatibility constraint for \(\theta_{m+1}\) can be written as

\[
(1 - \beta) \left[ u_2(c^*_2) - u_2 \left( c^{T}_{m,2} \right) \right] \geq h_0 \left( \frac{y^{*}_{m+1,0}}{\theta_{m+1}} \right) - h_0 \left( \frac{y^{*}_{m,0}}{\theta_{m+1}} \right) + h_1 \left( \frac{y^{*}_{m+1,1}}{\theta_{m+1}} \right) - h_1 \left( \frac{y^{*}_{m,1}}{\theta_{m+1}} \right).
\]

Since \(u_2\) is unbounded below and strictly increasing, there exists \(c^{T}_{m,2}\) such that all \(c_2 \leq c^{T}_{m,2}\) satisfies the local incentive compatibility constraint for \(\theta_{m+1}\). Trivially, \(c^{T}_{m,1}\) can be adjusted so the credible threat constraints still hold, because \(u_1\) is unbounded above. Finally, local downward incentive compatibility implies global incentive compatibility, which completes the proof. □

**Proof of Corollary** From the proof of Theorem 2 it is trivial to show the following programming problem has no finite solutions: for all \(\theta_m \in \Theta\),

\[
\min_{c^{T}_{m,1}, c^{T}_{m,2}, y^{T}_{m,1}} u_1(c^{T}_{m,1}) - h_1 \left( \frac{y^{T}_{m,1}}{\theta_m} \right) + \beta u_2 \left( c^{T}_{m,2} \right)
\]
subject to $u_1 (c^T_{m,1}) - h_1 \left( \frac{y^*_{m,1}}{\theta_{m+1}} \right) + \hat{\beta} u_2 (c^*_{m,2}) \geq u_1 (c^*_{m,1}) - h_1 \left( \frac{y^*_{m,1}}{\theta_{m+1}} \right) + \hat{\beta} u_2 (c^*_{m,2})$.

Since $\hat{\beta} > \beta$, this implies if the executability constraints hold with $c^*_{m,2} \geq c^T_{m,2}$, then the credible threat constraints hold. Finally, the same argument for incentive compatibility from the proof of Theorem 2 can be used.

**Proof of Lemma 1:** Suppose $\{(c^I_m, y^I_m)\}_{\theta_m \in \Theta}$ implements the efficient allocation for sophistication $\hat{\beta} > \beta$. This implies that $c^I_{m,2} > c^*_{m,2}$ for any $\theta_m > \theta_1$. Hence, the fooling constraints are satisfied for any $\hat{\beta}' > \hat{\beta}$, so it is incentive compatible. Notice for any $\hat{\beta}' > \hat{\beta}$, the executability constraints are satisfied. Therefore, $\{(c^I_m, y^I_m)\}_{\theta_m \in \Theta}$ also implements the efficient allocation for $\hat{\beta}' > \hat{\beta}$.

Suppose $\{(c^T_m, y^T_m)\}_{\theta_m \in \Theta}$ implements the efficient allocation for sophistication $\hat{\beta} > \beta$. This implies $c^*_2 > c^T_{m,2}$. To see this, note the incentive compatibility and threat constraints imply:

$$u_2 (c^*_2) - u_2 (c^T_{m,2}) > \left[ u_1 (c^T_{m,1}) - h_1 \left( \frac{y^T_{m,1}}{\theta_{m+1}} \right) \right] - \left[ u_1 (c^*_1) - h_1 \left( \frac{y^*_{m,1}}{\theta_{m+1}} \right) \right]$$

$$\geq \left[ u_1 (c^T_{m,1}) - h_1 \left( \frac{y^T_{m,1}}{\theta_{m+1}} \right) \right] - \left[ u_1 (c^*_1) - h_1 \left( \frac{y^*_{m,1}}{\theta_{m+1}} \right) \right]$$

$$\geq \hat{\beta} \left[ u_2 (c^*_2) - u_2 (c^T_{m,2}) \right].$$

Hence, the threat constraint is relaxed for any $\hat{\beta}' < \hat{\beta}$. Since the executability constraint also holds, this implies the credible threat constraints are satisfied for any $\hat{\beta}' < \hat{\beta}$. Therefore, for $\hat{\beta}' < \hat{\beta}$, incentive compatibility is satisfied and $\{(c^T_m, y^T_m)\}_{\theta_m \in \Theta}$ implements the efficient allocation too.

**Proof of Theorem 3:** First, set $(c^R_m, y^R_m) = (c^*_m, y^*_m)$, $y^I_m = y^*_m$, $c^I_{m,0} = c^T_{m,0} = c^*_m$ and $y^T_{m,0} = y^*_m$ for all $\theta_m \in \Theta$. Next, construct $\{(c^I_m, y^I_m)\}_{\theta_m \in \Theta}$ to implement the efficient allocation for some $\hat{\beta} \in (\beta, 1)$ and $\beta = \beta$. By Lemma 1, agents with $\hat{\beta}' > \hat{\beta}$ and $\beta = \beta$ would be fooled. Only the executability constraint depends on $\beta$. Since $c^I_{m,2} > c^*_2$, then for any $\beta < \beta$, the executability constraints also hold. Therefore, $\{(c^I_m, y^I_m)\}_{\theta_m \in \Theta}$ implements the efficient allocation for more naïve and tempted agents.

Next, construct $\{(c^T_m, y^T_m)\}_{\theta_m \in \Theta}$ to implement the efficient allocation for $\hat{\beta}$ and
Let $\beta = \beta$. For all $\theta_m$, fix $(c_{m,1}^l, c_{m,2}^l)$ and choose $(c_{m,1}^T, c_{m,2}^T)$ such that

$$u_1(c_{m,1}^l) - h_1\left(\frac{y_{m,1}^*}{\theta_{m+1}}\right) + \hat{\beta}u_2(c_{m,2}^l) \leq u_1(c_{m,1}^T) - h_1\left(\frac{y_{m,1}^T}{\theta_{m+1}}\right) + \hat{\beta}u_2(c_{m,2}^T). \quad (10)$$

Following the proof of Corollary 1, it is possible to construct $\{(c_{m}^{T}, y_{m}^{T})\}_{\theta_{m} \in \Theta}$ to implement the efficient allocation for $\hat{\beta}$ and $\beta = \beta$. Inequality (10) ensures agents who misreport would choose the threat allocation. Note that this does not change the reporting strategy for $\hat{\beta} > \hat{\beta}$ in a hybrid mechanism. By Lemma 1, $\{(c_{m}^{T}, y_{m}^{T})\}_{\theta_{m} \in \Theta}$ implements the efficient allocation for agents with $\hat{\beta} < \hat{\beta}$ and $\beta = \beta$. Finally, only the executability constraints depend on $\beta$. For any $\beta > \beta$, the executability constraint for threat allocations are relaxed. Hence, the hybrid mechanism implements the efficient allocation.

**Proof of Proposition 3:** Let $(c_{m}^{l}, y_{m}^{l})$ be imaginary allocations constructed in a direct mechanism that support the efficient allocation, where $y_{m}^{l} = y_{m}^{*}$ and $c_{m,0}^{l} = c_{m,0}^{*}$ for all $\theta_{m} \in \Theta$. Let $Y_{t}^{*} = \{y_{1,t}, \ldots, y_{m,t}, \ldots, y_{M,t}\}$. Furthermore, if $y_{t} \in Y_{t}^{*}$, let $y_{t}^{-1} \in \Theta$ denote the corresponding type. For example, $y_{m,t}^{-1} = \theta_{m}$. Let $EO = \{y = (y_{0}, y_{1}) | y_{0} \in Y_{0}^{*}, y_{1} \in Y_{1}^{*} \text{ and } y_{0}^{-1} = y_{1}^{-1}\}$ denote set of efficient output history. Consider the following taxes: $T_{0}(y_{0}) = y_{0} - c_{0}^{*}$, and $T_{1}(y_{0}, y_{1}, k_{1}) = y_{1} + k_{1} + 1_{1}b_{1}(y_{0}, y_{1}) + 1_{2}\alpha(y_{0}, y_{1}) - c_{1}^{*}$, and benefits:

$$b_{1}(y_{0}, y_{1}) = \begin{cases} c_{2}^{*} & \text{if } y \in EO \\ 0 & \text{otherwise} \end{cases}, \quad b_{2}(y_{0}, y_{1}) = \begin{cases} c_{m,2}^{l} & \text{if } y = y_{m}^{*} \\ 0 & \text{if } y \notin EO \end{cases},$$

with $\alpha(y_{0}, y_{1}) = c_{1}^{*} - c_{m,1}^{l}$, if $y \in EO$ and $y_{t}^{-1} = \theta_{m}$, otherwise $\alpha(y_{0}, y_{1}) = c_{1}^{*}$. This construction implements the efficient allocation.

Also, $b_{2}$ can be constructed to be less progressive than $b_{1}$. To see how, note that from the proof of Theorem 1, it is always possible to lower $c_{m,1}^{l}$ and increase $c_{m,2}^{l}$ such that incentive compatibility holds. Since $u_{t}$ is strictly concave, the increase in $c_{m,2}^{l}$ would need to be large compared to the decrease in $c_{m,1}^{l}$. This increases the ratio of lifetime benefits to taxes paid for higher productivity agents claiming at $t = 2$, which decreases progressivity.
Proof of Proposition 4: First, consider on-equilibrium path policies \((1_{EW} = 0)\).
Set \(s_0 \in (0, c_2^*)\), \(\rho_1 = -1\) and \(\rho_2 = \frac{1}{\beta} - 1\), so agents do not save at \(t = 0\). When \(1_{EW} = 0\), income taxes are

\[
T_0 (y_0) = \begin{cases} 
  y_0 - c_0^* & \text{if } y_0 \geq y_{1,0}^* \\
  y_0 & \text{otherwise}
\end{cases}, \quad T_1 (y_1) = \begin{cases} 
  y_1 - c_1^* - \beta (c_2 - s_0) & \text{if } y_1 \geq y_{1,1}^* \\
  y_1 & \text{otherwise}
\end{cases}
\]

With this setup, the doers in \(t = 0\) and \(t = 1\) would choose the efficient consumption on-equilibrium path.

Next, consider off-equilibrium path policies to support efficient output. A \(\theta_{m+1}\) agent with \(y_0 = y_{m,0}^*\) predicts his doer would solve the following in \(t = 1\) if \(y_1 < \bar{y} (y_{m,0}^*)\) : \(\max_{c_1, c_2, y_1} u_1 (c_1) - h_1 \left( \frac{y_1}{\eta_{m+1}} \right) + \beta u_2 (c_2)\) subject to \(c_1 + \beta (c_2 - s_0) = y_1 - T_1 (y_1)\). Let \((\tilde{c}_{m,1}, \tilde{c}_{m,2}, \tilde{y}_{m,1})\) denote the solution. Notice the solution to the problem does not depend on \(\theta_m\). If \(y_1 = \bar{y} (y_{m,0}^*)\), the doer in \(t = 1\) is predicted to solve the following: \(\max_{c_1, c_2} u_1 (c_1) + \beta u_2 (c_2)\) subject to \(c_1 + \beta \left( \frac{c_2}{1 - \rho (y_{m,0}^*)} - s_0 + \eta \right) = \xi (y_0) + y_1 - T_1 (y_1)\), and let \((\tilde{c}_{m,1}, \tilde{c}_{m,2})\) denote the solution, which does not depend on \(\theta_m\) nor \(\theta_{m+1}\).

Let \(\bar{y} (y_{m,0}^*) = y_{m,1}^* + \alpha_m\), with \(\alpha_m > 0\). Set \(\rho (y_{m,0}^*) > 1 - \beta\) to tempt the doer to increase \(c_1\) and decrease \(c_2\), and construct \(\xi (y_{m,0}^*) > 0\) to ensure the threat constraint binds. Let \(\xi_m (\rho (y_{m,0}^*), \alpha_m)\) denote this for \(y_0 = y_{m,0}^*\). Hence, types with \(\theta > \theta_{m+1}\) would strictly prefer withdrawing early if \(y_0 = y_{m,0}^*\).

Next, choose \(\rho (y_{m,0}^*)\) to satisfy the local downward incentive compatibility constraint of type \(\theta_{m+1}\). Notice that, for a given \(\alpha_m > 0\), \(\tilde{c}_{m,2} (\tilde{c}_{m,1})\) is a decreasing (increasing) function of \(\rho (y_{m,0}^*)\). The government can then increase \(\rho (y_{m,0}^*)\) till the incentive compatibility constraint holds, for any given \(\alpha_m > 0\). Let \(\hat{\rho} (\alpha_m)\) denote the level of early withdrawal penalty such that the incentive compatibility constraint holds. Hence, the government can set \(\xi (\hat{\rho} (y_{m,0}^*), \alpha_m)\) and \(\rho (y_{m,0}^*, \alpha_m)\) for the threat and local downward incentive compatibility constraints to hold for type \(\theta_m\).

Finally, to pin down \(\alpha_m\), the government can increase \(\alpha_m\) till the executability constraint for \(\theta_m\) holds. Also, since the threat constraint binds, the \(\theta_m\) agent would predict an output of \(y < \bar{y} (y_{m,0}^*)\) so the credible threat constraints for \(\theta_m\) hold. The process above can be repeated for all productivity types to ensure global incentive compatibility. ■
References


