Optimal Retirement Policies With Time-Inconsistent Agents *

Pei Cheng Yu †

December 28, 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 hidden present bias and sophistication. The paper shows how policies can utilize the time inconsistency of agents to improve welfare above the constrained efficient optimum. 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. The coexistence of both policies screens sophistication and present bias. These proposals outperform traditional policies, like linear savings subsidies or mandatory savings, by increasing redistribution and output efficiency. The resulting welfare improvement could be quantitatively significant depending on the size of the time-inconsistent population. (*JEL* Codes: D03, D62, D82, D84, D86, D91, H21)

Keywords: Retirement policy, Time inconsistency, Optimal taxation

*I thank Manuel Amador, Anmol Bhandari, V.V. Chari, Kim-Sau Chung, Emmanuel Farhi, Xavier Gabaix, Simone Galperti, Mike Golosov, Keiichi Kawai, Anton Kolotilin, Weiwen Leung, Hongyi Li, Benjamin Lockwood, Erzo Luttmer, Radek Paluszynski, 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] In response, strengthening social security and increasing participation in retirement accounts are mentioned as core issues by the previous 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) 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 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. Furthermore, empirical evidence shows that models with time-inconsistent preference can explain the consumption and savings patterns observed in the data (Angeletos et al., 2001; Laibson et al., 2017).

To study this question, this paper extends the Mirrlees taxation framework to incorporate quasi-hyperbolic discounting with hidden present bias and sophistication.[4] In essence, agents have private information about their productivity, the degree of present bias, and their beliefs regarding the severity of the bias. This paper provides a theoretical framework to designing retirement policies, which sheds light on features in social security and retirement accounts that could improve welfare.[5] The key is in utilizing the agent's time inconsistency, so policies go beyond mitigating the present bias. Traditional policies, such as linear savings subsidies or mandatory savings, increase savings by offsetting the bias independent of the asymmetric information (Krusell, Kuruscu, and Smith, 2010). Using traditional policies, the government is able to guarantee the constrained efficient optimum, but cannot do better. I consider policy instruments that could elicit private information at a lower cost while mitigating the present bias.

The contributions of this paper are twofold. First, this paper introduces off-equilibrium

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

[2] It also prescribes the use of lessons from behavioral economics to attain its goals (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] Chan (2017) finds evidence of large heterogeneity in present bias and sophistication.

[5] By retirement policies, I am referring to social security, policies regarding retirement accounts, and any other programs related to retirement welfare.

path policies to a Mirrlees taxation model, where agents have hidden present bias and sophistication. The gain in welfare from implementing off-equilibrium path policies in a Mirrlees setting is not straightforward. I show that if the utility from consumption is unbounded above and below, then the full information efficient allocation is implementable because the government is able to make arbitrarily large off-path threats and promises. I also argue how welfare can still be improved even when utility is bounded. Secondly, this paper also demonstrates how to implement off-equilibrium path policies in public finance and explores its implications on the design of retirement policies along with its quantitative impact.

To see how off-equilibrium path policies can improve welfare, consider a three period model with two productivity types, productive and unproductive, and all agents in the economy share the same degree of present bias. Agents work in the first two periods and retire in the last period.[6]

For fully naïve agents (unaware of present bias), the government designs policies that improve output efficiency by promising retirement benefits that cover a portion of the information rent. If an agent produces a high output in the first period, then an option of claiming higher retirement benefits would be available in the second period. However, in the second period, the present-biased agent prefers immediate gratification over delayed benefits, so the agent would forego the higher retirement benefits. Fully naïve agents do not foresee this and choose to produce efficiently due to over-estimating the value of retirement benefits to their future-selves. I will call this mechanism a betting mechanism.[7]

For sophisticated agents (fully aware of their bias and thus demand commitment), the government provides commitment in exchange for an increase in output efficiency. If an agent produces a low output in the first period, then the agent would face a menu of policies in the next period. One of the policies, the off-path policy, would exacerbate the present bias by increasing consumption in the second period at the expense of retirement welfare. A present-biased agent in the second period would prefer this policy from the menu. However, to qualify for it, the agent has to produce a sufficiently high output in the second period. Hence, a productive agent would choose the off-path policy if production was inefficiently low in the first period, while the unproductive agent would not find the off-path policy appealing due to the required increase in output. By backward induction, a productive agent is willing to exchange the information rent for commitment. I will refer to this mechanism as a conditional commitment mechanism.[8] Both betting and conditional committment mechanisms help

[6] In a Mirrlees taxation model, the relevant deviation is for agents to mimic less productive agents.

[7] Eliaz and Spiegler (2006) and Heidhues and Koszegi (2010) have analyzed exploitative contracts with naïvely time-inconsistent agents.

[8] Esteban and Miyagawa (2005), Galperti (2015) and Bond and Sigurdsson (2017) have also analyzed mechanisms with enlarged menus and time-inconsistent agents.

screen productivity for partially naïve agents (aware of present bias but underestimate its severity), because they are willing to exchange information rents for either commitment or speculative bets.

The first main result shows that in a Mirrlees economy with time-inconsistent agents and hidden present bias and sophistication, off-equilibrium path policies can improve welfare above the constrained efficient optimum. The government can choose a target level of sophistication, and design the betting and conditional mechanisms for the least and most present-biased agents at the targeted level. The paper shows that agents more naïve than the targeted level would self-select into the betting mechanism, while agents more sophisticated would self-select into the conditional commitment mechanism. For the betting mechanism, the rents are loaded in the future, so it needs to make sure the most present-biased agent prefers to tell the truth. This ensures all present-biased agents would prefer truth-telling. Similarly, for the conditional commitment mechanism, the off-path policies decrease retirement consumption, so if the most present-biased agents prefer the on-path policy, then less present-biased agents would prefer it too. As a result, screening sophistication and present bias is costless.[9] Notice that it is not necessary for the government to know the distribution of present-bias and sophistication beyond the range of present bias. With unbounded utility, screening productivity is also costless, so the efficient allocation is implementable. When all agents are present-biased, I show that there is an additional gain in welfare equivalent to a 0.46% increase in aggregate consumption above the constrained efficient optimum, which is close to $62 billion in current US dollars.

The second main result examines an economy with time-consistent agents. The presence of time-consistent agents limits the effectiveness of off-equilibrium path policies, because they do not require commitment and are not susceptible to speculative bets. Nevertheless, off-equilibrium path policies can still improve welfare by separating time-inconsistent agents from time-consistent agents and induce the time-inconsistent agents to increase output efficiency. This paper shows that welfare increases with the population of time-inconsistent agents. This is because total output increases with the proportion of present-biased agents, so the government can provide more information rent per time-consistent agent without additional distortions. This setting is important because off-equilibrium path policies may encourage agents to adopt outside commitment, and time-consistent agents could be interpreted as time-inconsistent agents with commitment. I also show how the presence of time-consistent agents places a natural restriction on speculative bets, so conditional com-

[9]This result differs from Eliaz and Spiegler (2006), where different sophistication level is provided with a different allocation. This is because their paper is exploitative, while the objective of this paper is paternalistic, so the government attempts to implement the same consumption for all agents regardless of sophistication or degree of present bias.

mitment mechanisms generally outperform betting mechanisms. As part of the numerical exercise, I quantify the optimal off-equilibrium path policies and show how the distortions decrease and welfare increases with the population of time-inconsistent agents.

This paper provides new and intuitive insights on the design of policies without recommending significant qualitative changes to 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.[10] National Commission on Fiscal Responsibility and Reform (2010) 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.[11] Non-sophisticated agents (fully naïve and partially naïve agents) would plan to claim at a later age and the decrease in progressivity would encourage them to work more 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 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,b). I show that increasing the liquidity of retirement accounts to allow for early withdrawals helps improve welfare. 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.[12] Thus, agents work efficiently for commitment. The paper shows how the redesigned retirement accounts and social security system can coexist to increase welfare through improvement in social insurance and production efficiency.

When time-consistent agents are present, I show how participation in retirement savings accounts should be voluntary. Time-inconsistent agents are encouraged to enroll in retirement accounts that provide them with commitment, but time-consistent agents should have full discretion on their savings.[13] The enrollment decision of agents helps the government separate time-inconsistent agents from time-consistent agents.

[10]See Coile et al. (2002), Shoven and Slavov (2014) and Knoll et al. (2015).

[11]Benefits 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 if the difference of this ratio between the low income individuals and higher income individuals also decreases.

[12]This implementation is similar to the idea of partially illiquid assets in Bond and Sigurdsson (2017). They analyze an endowment model so the implementation is restricted to savings.

[13]This is consistent with current proposals by many US states, like Oregon and Illinois, to provide all employees with retirement savings accounts which they can choose to opt out.

1.1 Related Literature

There have been several papers studying the design of policies with behavioral agents. Farhi and Gabaix (2015) study optimal taxation (Ramsey, Pigou and Mirrlees) with behavioral agents in a static environment 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 (2016) studies optimal income taxation when present-biased agents do not fully internalize the benefits of work to future earnings. In contrast, this paper emphasizes consumption and savings with present bias and focuses on optimal retirement policies.

Guo and Krause (2015) study a dynamic Mirrlees environment with sophisticated present-biased agents where the government does not have full commitment. Moser and de Souza e Silva (2017) consider a two-dimensional screening setup with hidden present bias and productivity and decentralizes the optimum using social security and retirement accounts. In Moser and de Souza e Silva (2017), present bias is stochastic, so providing a flexible retirement plan for high income agents while limiting the options for low income agents can relax incentive compatibility constraints and is optimal. This paper focuses on deterministic present bias, so flexible policies may induce under-savings, which is a distortion that could be avoided when agents have constant present bias.[14]

In other related work, Diamond and Spinnewijn (2011) discuss a model with heterogeneity in both productivity and time preference (agents are time-consistent). Krusell, Kuruscu, and Smith (2010) study the optimal taxation of consumers who suffer from temptation in a complete information environment. Amador, Werning, and Angeletos (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, Werning, and Angeletos (2006) with persistent shocks. Similar to this paper, Halac and Yared (2014) also point out how allowing the government to revise past reports in the future can help relax incentive compatibility and deter misreporting in the present.

This paper is also related to several behavioral contracting papers that have analyzed unused options to relax incentive compatibility constraints. 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 off-equilibrium path. Bond and Sigurdsson (2017) demonstrate how off-equilibrium path options in commitment contracts can help time-inconsistent agents follow through with

[14] Meier and Sprenger (2015) find evidence of present bias and obtain estimates for the quasi-hyperbolic model. They find the parameter estimates of β and δ to be relatively stable over time, which supports the theoretical environment of this paper.

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, Werning, and Angeletos (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 the agent's private information (productivity, time-inconsistency and sophistication) simultaneously.

The paper is organized as follows. Section 2 presents the model. Section 3 presents the general mechanism. Section 4 examines the setting with heterogeneous present bias and sophistication. Section 5 considers the effects of time-consistent agents. Section 6 discusses a reform of the social security and retirement accounts. It also provides a numerical exercise on the impact of the off-equilibrium path policies. Section 7 discusses some extensions and impediments to the mechanism and Section 8 concludes.

2 The Model

A continuum of agents of measure one live for three periods: $t \in \{0,1,2\}$. Agents are heterogeneous in productivity, which is denoted by $\Theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$, with $\theta_{m+1} > \theta_m$ and $|M| \geq 2$. Productivity is distributed according to $\Pr(\theta = \theta_m) = \pi_m > 0$, for all $\theta_m \in \Theta$ with $\sum_{m=1}^{M} \pi_m = 1$.[15]

The production technology is linear and depends 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 θ and l_t, but it observes y_t.

The period utilities $u_t : \mathbb{R}_+ \mapsto \mathbb{R}$ are twice differentiable and $u_t', -u_t'' > 0$. The dis-utility from labor $h_t : \mathbb{R}_+ \mapsto \mathbb{R}$ satisfies $h_t', h_t'' > 0$. Single crossing is automatically satisfied. For most of the paper, I will assume that u_t is unbounded below and above.

Assumption 1 *For any $t \in \{1,2\}$, u_t has full range ($u_t(\mathbb{R}_+) = \mathbb{R}$).*

In Section 4, I show that Assumption 1 is a sufficient condition for the efficient allocation to be incentive compatible. In Section 7, I discuss how the efficient allocation would not be implementable when Assumption 1 fails, but the main message of the paper still holds.

[15]For simplicity, the theoretical exposition will focus on the case where productivity does not change over the life-cycle. The main results still go through with a deterministic lifetime path of productivity: $\theta_m = (\theta_{m,1}, \theta_{m,2})$ and $\theta_{m,1} \neq \theta_{m,2}$. The quantitative analysis in Section 6.5 will analyse this case.

The utility of the agents at $t = 0$ is

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

while at $t = 1$ it is

$$V_1(c, y; \theta, \beta) = u_1(c_1) - h_1\left(\frac{y_1}{\theta}\right) + \beta\delta u_2(c_2),$$

à la quasi-hyperbolic discounting (Laibson, 1997). I will focus on the case where $\beta \in (0, 1)$, which measures the degree of present bias the agents suffer from.[16] I will denote the time-consistent utility ($\beta = 1$) as $U_t(c, y; \theta)$.

Following O'Donoghue and Rabin (2001), non-sophisticated agents at $t = 0$ perceive their present bias in $t = 1$ to be $\hat{\beta} \in (\beta, 1]$. Let $W_1(c, y; \theta, \beta, \hat{\beta})$ denote the non-sophisticated agents' perceived ex-post utility in $t = 1$:

$$W_1(c, y; \theta, \hat{\beta}) = u_1(c_1) - h_1\left(\frac{y_1}{\theta}\right) + \hat{\beta}\delta u_2(c_2).$$

If $\hat{\beta} = 1$, the agent is fully naïve and unaware of the future-self's present bias. If $\hat{\beta} = \beta$, the agent is sophisticated and fully aware of the bias. Partially naïve agents know they have present bias, $\hat{\beta} < 1$, but $\hat{\beta} > \beta$. In essence, they underestimate the severity of the future-self's bias. Non-sophistication refers to $\hat{\beta} \in (\beta, 1]$.[17] I will refer to $\hat{\beta}$ as *sophistication.*

Agents vary in present bias, β, and sophistication $\hat{\beta}$, so types are represented by $(\theta_m, \beta, \hat{\beta})$, where $\hat{\beta} \in [\beta, 1]$ and $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$ with $\overline{\beta} \geq \underline{\beta}$. In Section 4, I will analyze the case when $\overline{\beta} < 1$ and Section 5 will analyze the case when $\overline{\beta} = 1$. Let $\delta = 1$, which does not affect the results.

The timing is as follows: Before $t = 0$, the government designs the tax system, and has full commitment. The agents learn about their productivity at $t = 0$, and proceed to work, consume and save for each $t < 2$. The agents retire in $t = 2$.

The government tries to help the agents commit to the time-consistent counterpart, $U_0(c, y; \theta)$. I assume the non-sophisticated agents do not draw any inferences from the policies the government enacts, because they do not share the same prior as the government and are dogmatic in their beliefs. The government maximizes the following welfare criterion

$$\sum_{m=1}^{M} \pi_m U_0(c_m, y_m; \theta_m), \tag{1}$$

[16]Off-equilibrium threats and promises do not have bite if $\beta = 0$. The main idea of utilizing time inconsistency to raise welfare does not change if $\beta > 1$.

[17]If $\hat{\beta} < \beta$, it is still possible for the government to take advantage of the incorrect belief.

where (c_m, y_m) denotes the vector of allocations a θ_m agent consumes.[18] Since U is strictly concave in consumption, the government has a desire to insure agents against the realization of θ. I assume the government has no external revenue needs, so the feasibility constraint is

$$\sum_{t=0}^{2}\sum_{m=1}^{M}\pi_m\left(y_{m,t}-c_{m,t}\right)=0, \tag{2}$$

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

Finally, there are no private markets to insure against productivity shocks and no markets for illiquid assets or other commitment devices.

2.1 The Benchmarks

2.1.1 No Private Information

In the no private information case, the government maximizes social welfare (1) subject to the feasibility constraint (2).

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 $t>0$, $u_0'(c_{m,0}^*)=u_t'(c_{m,t}^*)$, and (iii.) efficient output: for any $t<2$, $u_t'(c_{m,t}^*)=\frac{1}{\theta_m}h_t'(\frac{y_{m,t}^*}{\theta_m})$.*

With complete information, the government achieves full insurance regardless of present bias or sophistication. This is because the agents work according to their productivity. The government then chooses an appropriate linear savings subsidy to correct the distortion caused by the present bias. To see this, consider an income tax $T_{m,t}^{\beta}$ for agents of productivity θ_m and present bias β at $t<2$. Let $y_t=y_{m,t}^*-T_{m,t}^{\beta}$ denote the after-tax income, which is the same for all productivity due to full insurance. Consider a savings subsidy of $\tau^s=(\tau_1^s,\tau_2^s)$. At $t=1$, for any savings s_1 made at $t=0$, the agents solve

$$\max_{c_1,s_2,c_2} u_1\left(c_1\right)-h_1\left(\frac{y_{m,1}^*}{\theta_m}\right)+\beta u_2\left(c_2\right)$$

subject to $c_1+s_2\le y_1+\left(1+\tau_1^s\right)s_1$ and $c_2\le\left(1+\tau_2^s\right)s_2$. If $\tau_2^s=\frac{1}{\beta}-1$, then $u_1'\left(c_1\right)=u_2'\left(c_2\right)$. Let $c_t\left(s_1\right)$ denote the optimal consumption given s_1. By backward induction, at $t=0$, the agents solve

$$\max_{c_0,s_1} u_0\left(c_0\right)-h_0\left(\frac{y_{m,0}^*}{\theta_m}\right)+\beta\left[u_1\left(c_1\left(s_1\right)\right)-h_1\left(\frac{y_{m,1}^*}{\theta_m}\right)+u_2\left(c_2\left(s_1\right)\right)\right]$$

[18] Much of the literature on dynamically inconsistent preferences have evaluated welfare with the time-consistent utility.

subject to $c_0 + s_1 = y_0$. Similarly, the government can choose τ_1^s such that $u_0'(c_0) = u_t'(c_t)$ for any $t > 0$. It can then adjust the income tax so the efficient consumption is implemented.

2.1.2 Private Information without Time Inconsistency

With private information, the implementable allocations must be incentive compatible. The government maximizes social welfare (1) subject to the feasibility constraint (2) and the incentive compatibility constraints, $\forall \theta_m, \theta_{m'} \in \Theta$,

$$U_0\left(c_m, y_m; \theta_m\right) \geq U_0\left(c_{m'}, y_{m'}; \theta_m\right), \tag{3}$$

which are evaluated at U_0 for time-consistent agents. 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 $\theta_m < \theta_M$, $u_t'(c_{m,t}^{**}) > \frac{1}{\theta_m} h_t'(\frac{y_{m,t}^{**}}{\theta_m})$.*

The constrained efficient allocation distorts the labor decisions of all agents except for the most productive agents θ_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.[19]

3 The General Mechanism

In this section, I will introduce the *betting mechanism* and the *conditional commitment mechanism* for known bias and sophistication. For a given present bias, β, and sophistication, $\hat{\beta}$, an enlarged menu C_m for type θ_m agents is defined as $C_m = \{(c_m, y_m), (c_m', y_m'), \ldots\}$. An agent is assigned a menu C_m after reporting θ_m. Let $\left(c_m^R, y_m^R\right) \in C_m$ be the *real allocation*, which is the optimal allocation the government implements. The government posts $C = \{C_1, \ldots, C_M\}$. The agents then choose a menu C_m from C after learning θ at $t = 0$. Without loss of generality, for all $\theta_m \in \Theta$, set all allocations in $t = 0$ equal to $\left(c_{m,0}^R, y_{m,0}^R\right)$, so agents start facing extraneous options at $t = 1$:

$$C_m = \left\{\left[\left(c_{m,0}^R, y_{m,0}^R\right), \left(c_{m,1}^R, y_{m,1}^R\right), c_{m,2}^R\right], \left[\left(c_{m,0}^R, y_{m,0}^R\right), \left(c_{m,1}', y_{m,1}'\right), c_{m,2}'\right], \ldots\right\},$$

[19]For more on the characterization of the constrained efficient allocation, see Hellwig (2007).

where the real allocations and extraneous options could be different.[20]

Incentive compatibility is characterized by what the agent perceives the future-self will choose under both honest and dishonest reporting. Let

$$C_m^{\hat{\beta}} = \left\{ (c_m, y_m) \in C_m \,\middle|\, (c_m, y_m) \in \arg \max_{(c'_m, y'_m) \in C_m} W_1(c'_m, y'_m; \theta_m, \hat{\beta}) \right\}.$$

$C_m^{\hat{\beta}}$ denotes the set of allocations a truthful θ_m agent with sophistication $\hat{\beta}$ predicts the future-self would choose at $t = 1$. Let

$$C_{m'|m}^{\hat{\beta}} = \left\{ (c_{m'}, y_{m'}) \in C_{m'} \,\middle|\, (c_{m'}, y_{m'}) \in \arg \max_{\left(c'_{m'}, y'_{m'}\right) \in C_{m'}} W_1(c'_{m'}, y'_{m'}; \theta_m, \hat{\beta}) \right\}.$$

$C_{m'|m}^{\hat{\beta}}$ denotes the set of allocations a θ_m agent with sophistication $\hat{\beta}$ predicts the future-self would choose after misreporting to be a $\theta_{m'}$ agent. Incentive compatibility is thus expressed as, $\forall \theta_m, \theta_{m'} \in \Theta$,

$$\max_{(c_m, y_m) \in C_m^{\hat{\beta}}} V_0 \left(c_m, y_m; \theta_m, \beta\right) \geq \max_{(c_{m'}, y_{m'}) \in C_{m'|m}^{\hat{\beta}}} V_0 \left(c_{m'}, y_{m'}; \theta_m, \beta\right). \tag{4}$$

The incentive compatibility constraints (4) ensure truthful reporting of productivity. Additional constraints are needed to ensure the real allocations are implemented at $t = 1$. The *executability constraints* are, $\forall \theta_m \in \Theta$,

$$\left(c_m^R, y_m^R\right) \in \arg \max_{(c_m, y_m) \in C_m} V_1 \left(c_m, y_m; \theta_m, \beta\right). \tag{5}$$

If the executability constraints (5) hold, the agent would choose the real allocations in $t = 1$. 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 The Betting Mechanism

Non-sophisticated agents mispredict their future behavior, so their reporting strategies reflect misguided expectations. The government implements a *betting mechanism* to exploit this incorrect belief.

[20]Non-sophisticated agents could potentially learn their present bias, as in Ali (2011). By delaying the appearance of the enlarged menu till $t = 1$, learning could be ignored.

Definition 1 *A direct betting mechanism for sophistication* $\hat{\beta} \in (\beta, 1]$ *has a menu* $C = \{C_m\}_{\theta_m \in \Theta}$ *with* $C_m = \left\{\left(c_m^R, y_m^R\right), \left(c_m^I, y_m^I\right)\right\}$ *and* $\left(c_m^R, y_m^R\right) \neq \left(c_m^I, y_m^I\right)$ *for some* θ_m *satisfying the fooling constraints:* $\forall \theta_m, \theta_{m'} \in \Theta$, $\left(c_m^I, y_m^I\right) \in C_m^{\hat{\beta}}$ *and* $\left(c_{m'}^I, y_{m'}^I\right) \in C_{m'|m}^{\hat{\beta}}$.

By Definition 1, the enlarged menu in a betting mechanism has the off-equilibrium allocation $\left(c_m^I, y_m^I\right) \in C_m$. Non-sophisticated agents of productivity θ_m predict they would choose (c_m^I, y_m^I) in $t = 1$, which, following Eliaz and Spiegler (2006), will be referred to as the *imaginary allocation*. However, the government intends the agents to choose allocation (c_m^R, y_m^R). Due to the *fooling constraints*, the benefits of truth-telling would be evaluated under the imaginary allocations.

Definition 2 *An allocation* $\left\{\left(c_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ *is truthfully implementable for present bias* β *by a direct betting mechanism for sophistication* $\hat{\beta} \in (\beta, 1]$, *if there exists* $\left\{\left(c_m^I, y_m^I\right)\right\}_{\theta_m \in \Theta}$ *such that the following are satisfied: (i.) incentive compatibility, and (ii.) executability.*

By Definition 2, to implement the real allocations, the executability constraints have to hold, which require the real allocations to be chosen at $t = 1$. Definition 2 also requires the imaginary allocations to satisfy incentive compatibility. This is because by the fooling constraints, the betting mechanism incentivizes truth-telling at $t = 0$ through off-equilibrium path imaginary allocations. This relaxes the incentive compatibility constraints because a portion of information rents is loaded on allocations that would not be chosen.

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 honors 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 the future-self would consume the imaginary allocation, the agent does not consider the belief and behavior of others.

3.2 Conditional Commitment Mechanism

For sophisticated agents, the government can design an off-equilibrium path option that exacerbates the present bias, which will be chosen only if an agent misreports productivity. This type of mechanism will be called a *conditional commitment mechanism*, since commitment is provided conditional on truth-telling 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. Therefore, conditional commitment mechanisms can be implemented for any agent who is not fully naïve, $\hat{\beta} \in [\beta, 1)$.

Definition 3 *A direct conditional commitment mechanism for sophistication* $\hat{\beta} \in [\beta, 1)$ *has a menu* $C = \{C_m\}_{\theta_m \in \Theta}$ *with* $C_m = \left\{ \left(c_m^R, y_m^R\right), \left(c_{m|n}^T, y_{m|n}^T\right)_{\theta_n > \theta_m} \right\}$ *and* $\left(c_m^R, y_m^R\right) \neq \left(c_{m|n}^T, y_{m|n}^T\right)$ *for some* θ_m *and all* $\theta_n > \theta_m$ *satisfying the threat constraints:* $\forall \theta_{m'} \in \Theta, \theta_m > \theta_{m'}$, $\left(c_{m'|m}^T, y_{m'|m}^T\right) \in C_{m'|m}^{\hat{\beta}}$, *and* $\left(c_m^R, y_m^R\right) \in C_m^{\hat{\beta}}$.

By Definition 3, the enlarged menu in a conditional commitment mechanism has the off-equilibrium allocations $\left(c_{m|n}^T, y_{m|n}^T\right)_{\theta_n > \theta_m} \in C_m$. I will refer to them as the *threat allocations*. The *threat constraints* ensure that agents perceive misreports would lead to their future-selves selecting the threat allocation. Notice that a conditional commitment mechanism for non-sophisticated agents require that $\left(c_m^R, y_m^R\right) \in C_m^{\hat{\beta}}$ for all productivity. This is because for $\hat{\beta} \in (\beta, 1)$, the threats are evaluated using $\hat{\beta}$ instead of the actual present bias β. Therefore, the threat constraints also need to ensure the agents perceive their future-selves choosing the real allocation when report is truthful. The constraint $\left(c_m^R, y_m^R\right) \in C_m^{\hat{\beta}}$ is redundant when agents are sophisticated, because it coincides with the executability constraint when $\hat{\beta} = \beta$.

Definition 3 also requires agents misreporting downward choose different threat allocations based on their inherent productivity, so the threat allocations are incentive compatible. In essence, agents decide whether to tell the truth in $t = 0$ or $t = 1$. If agents misreport in $t = 0$, then the incentive compatible off-path threats will uncover their lie.

Definition 4 *An allocation* $\left\{\left(c_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ *is truthfully implementable for present bias* β *by a direct conditional commitment mechanism when* $\hat{\beta} \in [\beta, 1)$ *if there exists* $\left\{ \left(c_{m|n}^T, y_{m|n}^T\right)_{\theta_n > \theta_m} \right\}_{\theta_m \in \Theta}$ *such that the following are satisfied: (i.) incentive compatibility, and (ii.) executability.*

By Definition 4, the threat, $\left(c_m^T, y_m^T\right)$, for type θ_m is designed such that, after preference reversal, a type θ_m agent who reports truthfully at $t = 0$ would never choose it in $t = 1$ (by the executability constraint). Definition 4 requires the threat allocations to satisfy incentive compatibility to deter the agents from misreporting. This helps relax the incentive compatibility constraints.

4 Hidden Present Bias and Sophistication

In this section, I will present the first main result of the paper. Consider an economy where all agents are time-inconsistent, but vary in present bias β and sophistication $\hat{\beta}$, so types are represented by $(\theta_m, \beta, \hat{\beta}) \in \Theta \times \left[\underline{\beta}, \overline{\beta}\right] \times [\beta, 1]$, where $\overline{\beta} < 1$. I will show that if Assumption 1 is satisfied, the efficient allocation is implementable by combining the betting and conditional commitment mechanisms. I will refer to it as the *hybrid mechanism.*

Definition 5 *A direct hybrid mechanism for sophistication* $\hat{\beta} \in (\beta, 1)$ *has a menu* $C = \{C_m\}_{\theta_m \in \Theta}$ *with* $C_m = \left\{ \left(c_m^R, y_m^R\right), \left(c_m^I, y_m^I\right), \left(c_{m|n}^T, y_{m|n}^T\right)_{\theta_n > \theta_m} \right\}$ *and* $\left(c_m^R, y_m^R\right) \neq \left(c_m^I, y_m^I\right)$ *for some* θ_m *satisfying the fooling constraints:* $\forall \theta_m, \theta_{m'} \in \Theta$, $\left(c_m^I, y_m^I\right) \in C_m^{\hat{\beta}}$ *and* $\left(c_{m'}^I, y_{m'}^I\right) \in C_{m'|m}^{\hat{\beta}}$, *and* $\left(c_m^R, y_m^R\right) \neq \left(c_{m|n}^T, y_{m|n}^T\right)$ *for some* θ_m *and all* $\theta_n > \theta_m$ *satisfying the threat constraints:* $\forall \theta_{m'} \in \Theta, \theta_m > \theta_{m'}$, $\left(c_{m'|m}^T, y_{m'|m}^T\right) \in C_{m'|m}^{\hat{\beta}}$, *and* $\left(c_m^R, y_m^R\right) \in C_m^{\hat{\beta}}$.

By Definition 5, the government implements a hybrid mechanism by choosing a fixed target sophistication $\hat{\beta}$ and designs both a betting and conditional commitment mechanisms for the targeted sophistication. The following definition defines a truthfully implementable hybrid mechanism for present bias β.

Definition 6 *An allocation* $\left\{\left(c_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ *is truthfully implementable for present bias* β *by a direct hybrid mechanism for* $\hat{\beta} \in (\beta, 1)$ *if there exists* $\left\{\left(c_m^I, y_m^I\right)\right\}_{\theta_m \in \Theta}$ *and* $\left\{ \left(c_{m|n}^T, y_{m|n}^T\right)_{\theta_n > \theta_m} \right\}_{\theta_m \in \Theta}$ *such that incentive compatibility (4) and executability constraints (5) are satisfied.*

The following Lemma analyzes the case for observable present bias. It shows that if both the betting and conditional commitment mechanisms for $\hat{\beta} \in (\beta, 1)$ can implement the efficient allocation, then the hybrid mechanism implements the efficient allocation with hidden productivity and sophistication.

Lemma 1 *For any* $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$, *betting mechanisms implementing the efficient allocation for* $\hat{\beta} \in (\beta, 1)$ *also implement it* $\forall \hat{\beta}' \geq \hat{\beta}$. *For any* $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$, *conditional commitment mechanisms implementing the efficient allocation for* $\hat{\beta} \in (\beta, 1)$ *also implement it* $\forall \hat{\beta}' \leq \hat{\beta}$.

Proof Suppose $\left\{\left(c_m^I, y_m^I\right)\right\}_{\theta_m \in \Theta}$ implements the efficient allocation for sophistication $\hat{\beta} > \beta$. The fooling and executability constraints imply $c_{m,2}^I > c_2^*$ for any $\theta_m > \theta_1$. Hence, the fooling constraints are satisfied for any $\hat{\beta}' > \hat{\beta}$, so it is incentive compatible for more naïve agents. Notice the executability constraints do not depened on sophistication, so they are satisfied for any $\hat{\beta}' > \hat{\beta}$. Therefore, $\left\{\left(c_m^I, y_m^I\right)\right\}_{\theta_m \in \Theta}$ also implements the efficient allocation for $\hat{\beta}' > \hat{\beta}$.

Suppose $\left\{ \left(c_{m|n}^T, y_{m|n}^T\right)_{\theta_n > \theta_m} \right\}_{\theta_m \in \Theta}$ implements the efficient allocation for sophistication $\hat{\beta} > \beta$. First, the executability constraint does not depend on sophistication, so it still holds for any $\hat{\beta}' \in \left[\beta, \hat{\beta}\right)$. Since $c^T_{m,0} = c^*_0$ and $y^T_{m,0} = y^*_{m,0}$, this implies $c^*_2 > c^T_{m|n,2}$ for any $\theta_n > \theta_m$, because by downward incentive compatibility and threat constraints: for any $\theta_m > \theta_{m'}$,

$$\begin{aligned}
u_2\left(c^*_2\right) - u_2\left(c^T_{m'|m,2}\right) &> \left[u_1\left(c^T_{m'|m,1}\right) - h_1\left(\frac{y^T_{m'|m,1}}{\theta_m}\right)\right] - \left[u_1\left(c^*_1\right) - h_1\left(\frac{y^*_{m,1}}{\theta_m}\right)\right] \\
&\geq \left[u_1\left(c^T_{m'|m,1}\right) - h_1\left(\frac{y^T_{m'|m,1}}{\theta_m}\right)\right] - \left[u_1\left(c^*_1\right) - h_1\left(\frac{y^*_{m',1}}{\theta_m}\right)\right] \\
&\geq \hat{\beta}\left[u_2\left(c^*_2\right) - u_2\left(c^T_{m',2}\right)\right].
\end{aligned}$$

The first inequality comes from incentive compatibility. The second inequality comes from the fact that the efficient allocation has $y^*_{m,1} > y^*_{m,1}$ when $\theta_{m'} > \theta_m$. The last inequality comes from the threat constraint. This implies that if $(c^*_m, y^*_m) \in C^{\beta}_m$, then $(c^*_m, y^*_m) \in C^{\hat{\beta}'}_m$ and if $\left(c^T_{m'}, y^T_{m'}\right) \in C^{\hat{\beta}}_{m'|m}$, then $\left(c^T_{m'}, y^T_{m'}\right) \in C^{\hat{\beta}'}_{m'|m}$, so the threat constraints are relaxed for $\hat{\beta}'$. Therefore, for $\hat{\beta}'$, incentive compatibility is satisfied and $\left\{\left(c^T_m, y^T_m\right)\right\}_{\theta_m \in \Theta}$ also implements the efficient allocation. ■

By Lemma 1, for a known β, a betting mechanism designed for sophistication $\hat{\beta}$ agents can also fool agents who are more naïve. This is because fooling the less naïve agents is more difficult, so incentives that could screen the productivity of less naïve agents will also work for more naïve agents. Similarly, a conditional commitment mechanism designed for sophistication $\hat{\beta}$ agents can also threaten agents who are more sophisticated. This is because threatening the less sophisticated agents is more difficult, so incentives that could separate the productivity of less sophisticated agents will also work for more sophisticated agents. The first main result of the paper shows that if Assumption 1 holds, then the efficient allocation is implementable for Mirrlees taxation with hidden present bias and sophistication by using a hybrid mechanism.

Theorem 1 *If all agents are time-inconsistent with hidden present bias and sophistication and Assumption 1 holds, the efficient allocation* $\{(c^*_m, y^*_m)\}_{\theta_m \in \Theta}$ *is truthfully implementable with a hybrid mechanism.*

Proof Set $\left(c^R_m, y^R_m\right) = \left(c^*_m, y^*_m\right), y^I_m = y^*_m, c^I_{m,0} = c^T_{m,0} = c^*_{m,0}$ and $y^T_{m,0} = y^*_{m,0}$ for all $\theta_m \in \Theta$. First, I will show that the betting mechanism can implement the efficient allocation for known β and $\hat{\beta} \in (\beta, 1)$ when Assumption 1 is satisfied. Choose $\left(c^I_{1,1}, c^I_{1,2}\right) = \left(c^*_1, c^*_2\right)$ and $u_1\left(c^I_{m,1}\right) + \beta u_2\left(c^I_{m,2}\right) = u_1\left(c^*_{m,1}\right) + \beta u_2\left(c^*_{m,2}\right)$, for all θ_m. The local downward incentive compatibility constraints can be rewritten as

$$\left(1-\hat{\beta}\right)\left[u_2\left(c^I_{m,2}\right)-u_2\left(c^I_{m-1,2}\right)\right] \geq \frac{1}{\beta}\left[h_0\left(\frac{y^*_{m,0}}{\theta_m}\right)-h_0\left(\frac{y^*_{m-1,0}}{\theta_m}\right)\right]+h_1\left(\frac{y^*_{m,1}}{\theta_m}\right)-h_1\left(\frac{y^*_{m-1,1}}{\theta_m}\right),$$

for all θ_m. Since $\hat{\beta} < 1$ and u_2 is unbounded above by Assumption 1, $c^I_{m,2}$ can be chosen to satisfy the local downward incentive compatibility constraints for all θ_m. Assumption 1 also implies u_1 is unbounded below, so $c^I_{m,1}$ is chosen such that the fooling constraints are binding. Since $c^I_{1,2} = c^*_2$, the local incentive compatibility constraints imply that $c^I_{m,2} > c^*_2$ for all $\theta_m > \theta_1$. Therefore, the executability constraints are satisfied. Finally, local downward incentive compatibility implies global incentive compatibility.

I will now show the conditional commitment mechanism can implement the efficient allocation for known β and $\hat{\beta} \in (\beta, 1)$ when Assumption 1 is satisfied. First, choose the threat constraints such that for all θ_m, $W_1\left(c^T_{m|m+1}, y^T_{m|m+1}; \theta_{m+1}, \hat{\beta}\right) = W_1\left(c^*_m, y^*_m; \theta_{m+1}, \hat{\beta}\right)$, and for all $\theta_n > \theta_{m+1}$, $W_1\left(c^T_{m|n}, y^T_{m|n}; \theta_n, \hat{\beta}\right) = W_1\left(c^T_{m|n-1}, y^T_{m|n-1}; \theta_n, \hat{\beta}\right)$. This essentially requires the off-path threats to be locally downward incentive compatible. If $y^T_{m|n,1}$ is monotonically increasing in θ_n, then it is sufficient for global incentive compatibility due to the single crossing condition. It implies

$$\begin{aligned} u_1\left(c^T_{m|n,1}\right) =& u_1\left(c^*_1\right) + \hat{\beta}\left[u_2\left(c^*_2\right) - u_2\left(c^T_{m|n,2}\right)\right] + \left[h_1\left(\frac{y^T_{m|m+1,1}}{\theta_{m+1}}\right) - h_1\left(\frac{y^*_{m,1}}{\theta_{m+1}}\right)\right] \\ &+ \sum_{s=m+2}^{n}\left[h_1\left(\frac{y^T_{m|s,1}}{\theta_s}\right) - h_1\left(\frac{y^T_{m|s-1,1}}{\theta_s}\right)\right], \end{aligned}$$

and the incentive compatibility constraint can be rewritten as

$$\begin{aligned} \left(1-\hat{\beta}\right)\left[u_2\left(c^*_2\right)-u_2\left(c^T_{m|n,2}\right)\right] \geq & \frac{1}{\beta}\left[h_0\left(\frac{y^*_{n,0}}{\theta_n}\right)-h_0\left(\frac{y^*_{m,0}}{\theta_n}\right)\right]+h_1\left(\frac{y^*_{n,1}}{\theta_n}\right)-h_1\left(\frac{y^*_{m,1}}{\theta_{m+1}}\right) \\ &+ \sum_{s=m+2}^{n}\left[h_1\left(\frac{y^T_{m|s-1,1}}{\theta_{s-1}}\right) - h_1\left(\frac{y^T_{m|s-1,1}}{\theta_s}\right)\right]. \end{aligned}$$

By Assumption 1, $c^T_{m|n,2}$ is strictly decreasing in θ_n, and if $y^T_{m|n,1}$ is strictly increasing in θ_n, then $c^T_{m|n,1}$ is strictly increasing in θ_n.

Next, I will show that $y^T_{m|m+1,1} > y^*_{m,1}$ and $y^T_{m|n,1}$ is strictly increasing in θ_n, which would imply $(c^*_m, y^*_m) \in C^{\hat{\beta}}_m$, so the threat constraints are satisfied. To see this, let the executability constraints bind: $\forall \theta_m, \theta_n > \theta_m$, $V_1\left(c^T_{m|n}, y^T_{m|n}; \theta_m, \beta\right) = V_1\left(c^*_m, y^*_m; \theta_m, \beta\right)$. When $\theta_n = \theta_{m+1}$, the executability constraint can be written as

$$\left(\hat{\beta}-\beta\right)\left[u_2\left(c_2^*\right)-u_2\left(c_{m|m+1,2}^T\right)\right]=\left[h_1\left(\frac{y_{m|m+1,1}^T}{\theta_m}\right)-h_1\left(\frac{y_{m|m+1,1}^T}{\theta_{m+1}}\right)\right]-\left[h_1\left(\frac{y_{m,1}^*}{\theta_m}\right)-h_1\left(\frac{y_{m,1}^*}{\theta_{m+1}}\right)\right],$$

Since $c_2^* > c_{m|m+1,2}^T$ and $\hat{\beta} > \beta$, the strict convexity of h_1 would imply $y_{m|m+1,1}^T > y_{m,1}^*$. By the construction above, for any $\theta_n > \theta_m$, the executability constraint is

$$\left(\hat{\beta}-\beta\right)\left[u_2\left(c_{m|n-1,2}^T\right)-u_2\left(c_{m|n,2}^T\right)\right]=\left[h_1\left(\frac{y_{m|n,1}^T}{\theta_m}\right)-h_1\left(\frac{y_{m|n,1}^T}{\theta_n}\right)\right]-\left[h_1\left(\frac{y_{m|n-1,1}^T}{\theta_m}\right)-h_1\left(\frac{y_{m|n-11}^T}{\theta_n}\right)\right].$$

Since $c_{m|n,2}^T$ is strictly decreasing in θ_n, $y_{m|n,1}^T$ is strictly increasing in θ_n. Hence, the threat constraints are satisfied, and incentive compatibility and executability are satisfied by construction. More specifically, the threat allocations can be recovered by first observing that $c_{m|m+1,2}^T$ is pinned down by the efficient allocation. With $c_{m|m+1,2}^T$, $y_{m|m+1,1}^T$ is backed out from the executability constraint. Repeating this process would yield us the threat allocations.

Finally, I will show that with hidden present bias and sophistication, the hybrid mechanism can implement the efficient allocation. Construct $\left\{\left(c_m^I, y_m^I\right)\right\}_{\theta_m\in\Theta}$ to implement the efficient allocation for some $\hat{\beta} \in \left(\overline{\beta}, 1\right)$ satisfying the following inequalities:

$$u_1\left(c_{m,1}^*\right)+\hat{\beta}u_2\left(c_{m,2}^*\right)=u_1\left(c_{m,1}^I\right)+\hat{\beta}u_2\left(c_{m,2}^I\right), \tag{6}$$

$$V_0\left(c_m^I, y_m^I; \theta_m, \underline{\beta}\right) \geq V_0\left(c_{m-1}^I, y_{m-1}^I; \theta_m, \underline{\beta}\right). \tag{7}$$

By (6), agents with the targeted sophistication would weakly prefer the imaginary allocation over the efficient allocations in $t = 1$. Since $c_{m,2}^I > c_{m,2}^*$ and $\hat{\beta} > \overline{\beta}$, then for any $\beta \leq \overline{\beta}$, the executability constraints also hold and by Lemma 1, all agents with $\hat{\beta}' \geq \hat{\beta}$ would be fooled. By (7), the most present-biased agents would prefer truth-telling. It is clear that it would still be incentive compatible for any $\beta > \underline{\beta}$.

Construct $\left\{\left(c_m^T, y_m^T\right)\right\}_{\theta_m\in\Theta}$ to implement the efficient allocation for $\hat{\beta}$ and $\beta = \underline{\beta}$. For all θ_m, fix $\left(c_{m,1}^I, c_{m,2}^I\right)$ and choose $\left(c_{m|m+1,1}^T, c_{m|m+1,2}^T\right)$ such that

$$u_1\left(c_{m,1}^I\right)-h_1\left(\frac{y_{m,1}^*}{\theta_{m+1}}\right)+\hat{\beta}u_2\left(c_{m,2}^I\right)=u_1\left(c_{m|m+1,1}^T\right)-h_1\left(\frac{y_{m|m+1,1}^T}{\theta_{m+1}}\right)+\hat{\beta}u_2\left(c_{m|m+1,2}^T\right). \tag{8}$$

Inequality (8) ensures agents who misreport would be indifferent between the threat and imaginary allocations. Following the construction of threat allocations that I have shown, it is possible to construct $\left\{\left(c_m^T, y_m^T\right)\right\}_{\theta_m\in\Theta}$ to implement the efficient allocation for $\hat{\beta}$ and $\beta = \underline{\beta}$. Note that this does not change the reporting strategy for $\hat{\beta}' > \hat{\beta}$ in a hybrid mechanism. By Lemma 1, $\left\{\left(c_m^T, y_m^T\right)\right\}_{\theta_m \in \Theta}$ implements the efficient allocation for agents with $\hat{\beta}' < \hat{\beta}$ and $\beta = \underline{\beta}$. Finally, only the executability and incentive compatibility constraints depend on β. For any $\beta > \underline{\beta}$, both the executability and incentive compatibility constraints for threat allocations are relaxed. Hence, the hybrid mechanism implements the efficient allocation. ■

Theorem 1 depends on three key observations. First, the efficient allocation is implementable with betting and conditional commitment mechanisms for a known present bias and sophistication when Assumption 1 holds. Therefore, the betting and conditional mechanisms facilitate the government in screening productivity. Secondly, the government can design betting and conditional commitment mechanisms for a target sophistication to screen sophistication. This is because by Lemma 1, agents more naïve than the target would self-select into the betting mechanism, while agents less naïve than the target would self-select into the conditional commitment mechanism. Finally, to screen present bias, both mechanisms only need to be designed for the most present-biased agents. If the most present-biased agents prefer to tell the truth in a betting mechanism, then less present-biased agents would prefer truth-telling as well. Similarly, if the real allocation is executable for the most present-biased agent in a conditional commitment mechanism, then it is also executable for less present-biased agents. The key is to choose the targeted sophistication $\hat{\beta}$ to be greater than $\overline{\beta}$. If the target is less than $\overline{\beta}$, then sophisticated agents with $\beta \in \left(\hat{\beta}, \overline{\beta}\right)$ would choose the imaginary allocations over the real allocations.

It is also important to point out that Theorem 1 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 $\underline{\beta}$ and $\overline{\beta}$.

To see how the betting and conditional commitment mechanisms screen productivity and why Assumption 1 is a sufficient condition for implementing the efficient allocation, consider an economy with two productivity types $\Theta = \{\theta_L, \theta_H\}$, where $\theta_H > \theta_L$, and fixed present bias β. Let $\left\{\left(c_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ be the efficient allocation, where $c_{H,t}^R = c_{L,t}^R = c_t^*$ and $y_{m,t}^R = y_{m,t}^*$ for all $\theta_m \in \Theta$. The efficient allocation satisfies Proposition 1, so $y_H^* > y_L^*$. For simplicity, I will demonstrate the betting mechanism for the fully naïve case: $\hat{\beta} = 1$, and the conditional commitment mechanism for the fully sophisticated case $\hat{\beta} = \beta$.

I will first examine the betting mechanism. Set $c_{m,0}^I = c_0^*$ and $y_{m,t}^I = y_{m,t}^*$ for all θ_m. In Figure 1, the flatter solid (blue) curve represents the indifference curve from the perspective of $t = 0$ at allocation (c_1^*, c_2^*). The present-biased agents value c_2 less at $t = 1$ than at $t = 0$, so the steeper solid (red) curve represents the indifference curve from the perspective of $t = 1$

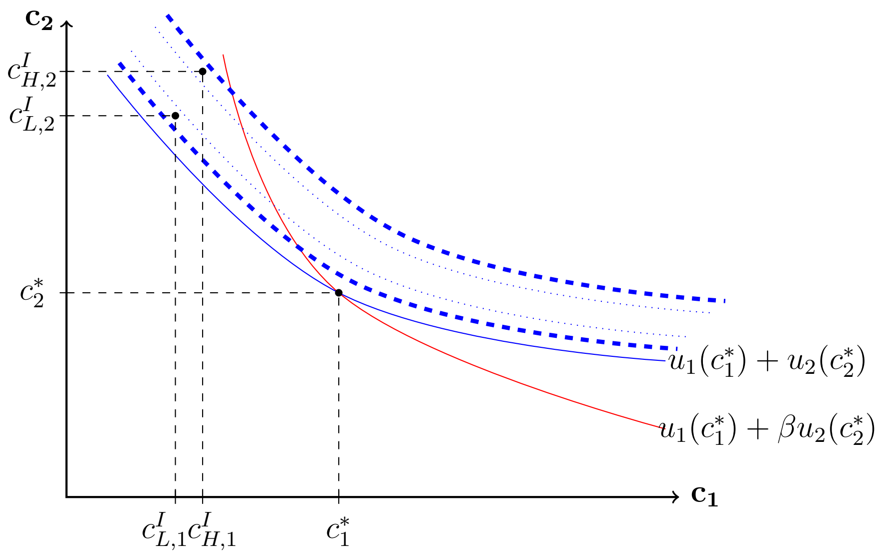

Figure 1: Finding the Imaginary Allocations

at allocation (c_1^*, c_2^*). The imaginary allocations have to be in the area bounded by the solid indifference curves in the north-west region: below the red curve and above the blue curve. This is because for the efficient allocations to be implemented, the agents have to prefer it over the imaginary allocations. This would ensure the executability constraints are satisfied.

Furthermore, the incentive compatibility constraints provide upper and lower bounds to the difference in utility between the two types of agents evaluated at the imaginary allocations. In essence,

$$\begin{aligned}\sum_{t=0}^{1} \beta^t \left[h_t\left(\frac{y^*_{H,t}}{\theta_L}\right) - h_t\left(\frac{y^*_{L,t}}{\theta_L}\right)\right] &\geq \beta\left[u_1(c^I_{H,1}) + u_2(c^I_{H,2})\right] - \beta\left[u_1(c^I_{L,1}) + u_2(c^I_{L,2})\right] \\ &\geq \sum_{t=0}^{1} \beta^t \left[h_t\left(\frac{y^*_{H,t}}{\theta_H}\right) - h_t\left(\frac{y^*_{L,t}}{\theta_H}\right)\right],\end{aligned}$$

where the upper bound is derived from the incentive compatibility constraint for θ_L and the lower bound is derived from the incentive compatibility constraint for θ_H. Figure 1 shows that the difference in utility between the θ_H and θ_L imaginary allocations have to be greater than the thin dashed lines (the lower bound) and less than the thick dashed lines (the upper bound) for incentive compatibility to be satisfied.

When Assumption 1 is satisfied, the indifference curves are bounded away from the axis. Hence, it would always be possible to find imaginary allocations that satisfy the incentive compatibility, fooling and executability constraints by increasing $c^I_{H,2}$ and decreasing $c^I_{H,1}$. In other words, if Assumption 1 is satisfied, the government can always decrease consumption in $t=1$ and load the information rent on retirement consumption to simultaneously satisfy both incentive compatibility and executability.

Next, to see how the conditional commitment mechanism implements the efficient allocation for sophisticated agents when Assumption 1 holds, set $c_{L,0}^{T}=c_{0}^{*}$ and $y_{L,0}^{T}=y_{L,0}^{*}$, so the threat occurs at $t=1$. Let $\Phi_{j,k}^{i}=u_{1}(c_{j,1}^{i})-h_{1}\left(y_{j,1}^{i}/\theta_{k}\right)$ and $\Phi_{k,k}^{i}=\Phi_{k}^{i}$, where $i\in\{R,T\}$ and $j,k\in\{L,H\}$. This demonstration will proceed in two steps. For the first step, I will first show how to construct threat allocations to deter misreporting. I will then show how it can be adjusted so that truthful agents would never choose it. From incentive compatibility, threat and executability constraints and by $\theta_{H}>\theta_{L}$ and $\beta<1$, the efficient and threat allocations have to satisfy: $\Phi_{L,H}^{T}>\Phi_{L,H}^{R}>\Phi_{L}^{R}>\Phi_{H}^{R}$, and $c_{2}^{*}>c_{L,2}^{T}$.

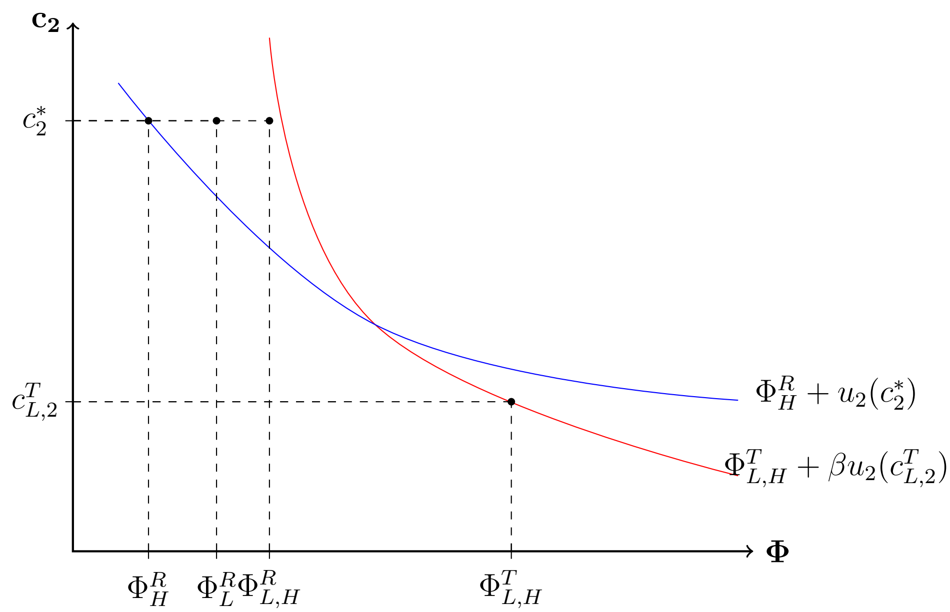

Figure 2: Finding the Threat Allocation: Part I

Figure 2 shows how the incentive compatibility constraint restricts the set of threat allocations. The steeper solid (red) curve represents the indifference curve from the perspective of $t=1$ for the θ_{H} agent who pretended to be θ_{L} in $t=0$. The flatter solid (blue) curve represents the indifference curve from the perspective of $t=0$ for the θ_{H} agent who reported truthfully in $t=0$. Figure 2 shows when Assumption 1 holds, the government can choose $\left(c_{L}^{T},y_{L}^{T}\right)$ such that the incentive compatibility constraint is satisfied by decreasing $c_{L,2}^{T}$ and increasing $\Phi_{L,H}^{T}$. Furthermore, when Assumption 1 holds, it is possible to increase $c_{L,1}^{T}$ so that the threat constraint holds for any arbitrarily small $c_{L,2}^{T}$.

Finally, I will show that $\left(c_{L,1}^{T},y_{L,1}^{T}\right)$ can be chosen so that the executability constraint is satisfied. To see this, fix the choice of $c_{L,2}^{T}$ and $\Phi_{L,H}^{T}$ at the level shown in Figure 2. If the threat satisfies incentive compatibility, threat and executability constraints, it implies

$$\Delta u_2 \geq \Phi_{L,H}^T - \Phi_H^R + \frac{1}{\beta}\Delta h_0 > \Phi_{L,H}^T - \Phi_{L,H}^R \geq \beta \Delta u_2 \geq \Phi_L^T - \Phi_L^R,$$

where $\Delta u_2 \equiv \left[u_2(c_2^*) - u_2(c_{L,2}^T)\right]$ and $\Delta h_0 = h_0\left({}^{y_{H,0}^*}/_{\theta_H}\right) - h_0\left({}^{y_{L,0}^*}/_{\theta_H}\right)$. The problem now is to find $c_{L,1}^T$ and $y_{L,1}^T$ such that $u_1(c_{L,1}^T) - h_1\left({}^{y_{L,1}^T}/_{\theta_H}\right) = \Phi_{L,H}^T$ and satisfies the executability constraint, $\beta\Delta u_2 \geq \Phi_L^T - \Phi_L^R$.

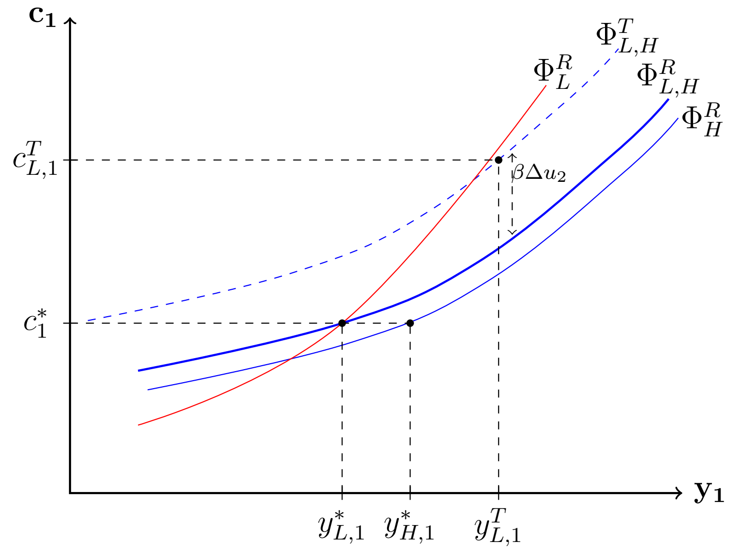

Figure 3: Finding the Threat Allocation: Part II

In Figure 3, the flatter thick solid (blue) curve represents the indifference curve of Φ for the θ_H agents at allocation $\left(c_1^*, y_{L,1}^*\right)$. The steeper solid (red) curve represents the indifference curve of Φ for the θ_L agents at allocation $\left(c_1^*, y_{L,1}^*\right)$. The dashed (blue) curve represents the indifference curve of Φ for the θ_H agent at allocation $\left(c_{L,1}^T, y_{L,1}^T\right)$, chosen so that $u_1(c_{L,1}^T) - h_1\left({}^{y_{L,1}^T}/_{\theta_H}\right) = \Phi_{L,H}^T$ and the executability constraint holds. In essence, the government can increase $y_{L,1}^T$ to discourage θ_L agents from choosing the threat allocation. While by Assumption 1, it can increase $c_{L,1}^T$ the threat remains potent for θ_H agents. Hence, Assumption 1 provides the government sufficient leverage to discipline the agents through off-equilibrium path policies.

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. TC agents cause distortions, because they follow through with their consumption plans, so the effectiveness of off-path policies are limited and the efficient optimum is no longer attainable. I also present the second main result of the paper, which shows that welfare increases monotonically with the proportion of time-inconsistent agents in the economy. The proofs for this section are in the online appendix.

I will focus on TI agents with the same present bias β and sophistication $\hat{\beta} < 1$. With Assumption 1, this environment is the same as a setting with heterogeneous present bias and sophistication where $\underline{\beta} = \beta$ and the least sophisticated agent has sophistication $\hat{\beta}$. I have excluded fully naïve agents and I will focus on conditional commitment mechanisms. The government is uncertain whether the agents are TC ($\beta = 1$) or TI ($\beta < 1$), with probability $\Pr(TI) = \phi$. The TC agents know their consistency, while TI agents could be non-sophisticated. I assume the distribution of productivity is independent of the agents' consistency. I will first present the result for conditional commitment mechanisms and then discuss betting mechanisms and how they are less effective in separating TC from TI agents.

5.1 Conditional Commitment Mechanisms

If the government implements a conditional commitment mechanism, then it can design the following menu for TC agents: $C_m^{TC} = \left\{ \left(c_m^P, y_m^P\right); \left(c_m^D, y_m^D\right) \right\}$, and the following menu for TI agents: $C_m^{TI} = \left\{ \left(c_m^R, y_m^R\right); \left(c_{m|n}^T, y_{m|n}^T\right)_{\theta_n > \theta_m} \right\}$. The allocation $\left(c_m^P, y_m^P\right)$ is the *persistent allocation*, and it is the allocation the government implements for θ_m TC agents. The allocation $\left(c_m^D, y_m^D\right)$ is referred to as the *deterrent allocation*, and it is meant to deter the TI agents from misreporting as θ_m TC agents. The idea is similar to Galperti (2015), where unused options were introduced to deter time-inconsistent agents from mimicking time-consistent agents. I will show that the optimal allocations in this environment will always make the presence of deterrent allocations necessary. The deterrent allocations work in a similar fashion as threat allocations. The government offers $C = \left\{ C_m^{TC}, C_m^{TI} \right\}_{\theta_m \in \Theta}$. The mechanism is meant to separate agents along two dimensions: productivity and consistency.

Let

$$C_{m'|m}^{1|\hat{\beta}} = \left\{ (c_{m'}, y_{m'}) \in C_{m'}^{TC} \,\middle|\, (c_{m'}, y_{m'}) \in \arg \max_{\left(c'_{m'}, y'_{m'}\right) \in C_{m'}^{TC}} W_1(c'_{m'}, y'_{m'}; \theta_m, \hat{\beta}) \right\}.$$

$C_{m'|m}^{1|\hat{\beta}}$ denotes the set of allocations a TI agent with productivity θ_m and sophistication $\hat{\beta}$ predicts the future-self will choose in $t = 1$ after misreporting to be TC with productivity

$\theta_{m'}$. Let

$$C_m^1 = \left\{ (c_m, y_m) \in C_m^{TC} \middle| (c_m, y_m) \in \arg \max_{(c'_m, y'_m) \in C_m^{TC}} U_0(c'_m, y'_m; \theta_m) \right\}.$$

C_m^1 denotes the set of allocations a truth-telling TC agent would choose. Let $C_m^{\hat{\beta}}$ and $C_{m'|m}^{\hat{\beta}}$ be defined as before. The following definition defines a direct conditional commitment mechanism with TC agents.

Definition 7 *A direct conditional commitment mechanism for sophistication $\hat{\beta} \in (\beta, 1)$ with TC agents has $C = \left\{ C_m^{TC}, C_m^{TI} \right\}_{\theta_m \in \Theta}$ satisfying: (i.) threat constraints: $\forall \theta_m, \theta_{m'} \in \Theta$, $\left(c_m^R, y_m^R \right) \in C_m^{\hat{\beta}}$ and $\left(c_{m'|m}^T, y_{m'|m}^T \right) \in C_{m'|m}^{\hat{\beta}}$, (ii.) deterrent constraints: $\forall \theta_m, \theta_{m'} \in \Theta$, $\left(c_m^P, y_m^P \right) \in C_m^1$, and $\left(c_{m'}^D, y_{m'}^D \right) \in C_{m'|m}^{1|\hat{\beta}}$.*

The incentive compatibility constraints for the TI agents are, $\forall \theta_m, \theta_{m'} \in \Theta$,

$$\max_{(c_m, y_m) \in C_m^{\hat{\beta}}} V_0 \left(c_m, y_m; \theta_m, \beta \right) \geq \max \left\{ \begin{array}{l} \max\limits_{(c_{m'}, y_{m'}) \in C_{m'|m}^{\hat{\beta}}} V_0 \left(c_{m'}, y_{m'}; \theta_m, \beta \right), \\ \max\limits_{(c_{m'}, y_{m'}) \in C_{m'|m}^{1|\hat{\beta}}} V_0 \left(c_{m'}, y_{m'}; \theta_m, \beta \right) \end{array} \right\}. \tag{9}$$

By (9), it is optimal for the TI agents to report truthfully about their productivity and consistency. Let

$$C_{m'|m}^{\hat{\beta}|1} = \left\{ (c_{m'}, y_{m'}) \in C_{m'}^{TI} \middle| (c_{m'}, y_{m'}) \in \arg \max_{\left(c'_{m'}, y'_{m'}\right) \in C_{m'}^{TI}} U_0(c'_{m'}, y'_{m'}; \theta_m) \right\},$$

$$C_{m'|m}^{1} = \left\{ (c_{m'}, y_{m'}) \in C_{m'}^{TC} \middle| (c_{m'}, y_{m'}) \in \arg \max_{\left(c'_{m'}, y'_{m'}\right) \in C_{m'}^{TC}} U_0(c'_{m'}, y'_{m'}; \theta_m) \right\}.$$

Hence, $C_{m'|m}^{\hat{\beta}|1}$ denotes the set of allocations a TC agent with productivity θ_m would select from a menu for TI agents with productivity $\theta_{m'}$, while $C_{m'|m}^1$ denotes the set of allocations a TC agent would select if productivity was misreported as $\theta_{m'}$ and was truthful about consistency. The incentive compatibility constraints for the TC agents are, $\forall \theta_m, \theta_{m'} \in \Theta$,

$$\max_{(c_m,y_m)\in C_m^1} U_0\left(c_m, y_m; \theta_m\right) \geq \max \left\{ \max_{(c_{m'},y_{m'})\in C^1_{m'|m}} U_0\left(c_{m'}, y_{m'}; \theta_m\right), \max_{(c_{m'},y_{m'})\in C^{\hat{\beta}|1}_{m'|m}} U_0\left(c_{m'}, y_{m'}; \theta_m\right) \right\}. \tag{10}$$

Incentive compatibility constraints (10) discourage the TC agents from misreporting productivity or consistency. The executability constraints are defined by (5).

Definition 8 *The allocation* $\left\{\left(c_m^P, y_m^P\right), \left(c_m^R, y_m^R\right)\right\}_{\theta_m\in\Theta}$ *is truthfully implementable for present bias* β *by a direct conditional commitment mechanism for sophistication* $\hat{\beta}$ *with TC agents if there exists* $\left\{\left(c_m^D, y_m^D\right), \left(c_{m|n}^T, y_{m|n}^T\right)_{\theta_n>\theta_m}\right\}_{\theta_m\in\Theta}$ *such that (i.) incentive compatibility, and (ii.) executability are satisfied.*

The government maximizes welfare

$$\sum_{\theta_m\in\Theta} \pi_m \left[\phi U_0\left(c_m^R, y_m^R; \theta_m\right) + (1-\phi) U_0\left(c_m^P, y_m^P; \theta_m\right)\right], \tag{11}$$

subject to the incentive compatibility constraints, executability constraints, credible threat constraints, deterrent constraints and the feasibility constraint

$$\sum_{\theta_m\in\Theta} \left\{ \phi\pi_m \left[\sum_{t=0}^{2} \left(y_{m,t}^R - c_{m,t}^R\right)\right] + (1-\phi)\pi_m \left[\sum_{t=0}^{2} \left(y_{m,t}^P - c_{m,t}^P\right)\right] \right\} = 0, \tag{12}$$

where $y_2 = 0$. The threat and deterrent allocations can help relax (9). If Assumption 1 holds, the government only needs to deter misreporting from the TC agents. The following theorem shows how the government takes advantage of the TI agents.

Theorem 2 *For a conditional commitment mechanism, if Assumption 1 holds and* $\phi \in (0,1)$ *and* $\hat{\beta} \in [\beta, 1)$*, then there exists* $\bar{\theta} > \theta_1$ *such that (i.) for* $\theta_m \geq \bar{\theta}$*,* $c_m^P > c_m^R$ *with* $y_m^R > y_m^P$*, (ii.) for* $\theta_1 < \theta_m < \bar{\theta}$*,* $\left(c_m^P, y_m^P\right) \leq \left(c_m^R, y_m^R\right)$*, (iii.)* $\left(c_1^P, y_1^P\right) = \left(c_1^R, y_1^R\right)$.

By Theorem 2, full insurance is no longer incentive compatible when TC agents are in the economy. The high productivity ($\theta_m \geq \bar{\theta}$) TC agents require information rents, so they have lower marginal utilities from consumption and lower disutility from effective labor y.

When Assumption 1 holds, the incentive compatibility constraints of TI agents are nonbinding. The government exploits the higher productivity TI agents by requiring them to work more and consume less, which increases the resources available for redistribution. As a result, TC agents with $\theta_m \geq \bar{\theta}$ would never misreport to be TI agents of the same productivity. For agents with lower productivity ($\theta_m < \bar{\theta}$), the government would exploit the TI agents by requiring them to work more, but also compensate them with more consumption for insurance. The consumption is limited by the incentive compatibility constraint for the higher productivity agents. This increase in production more than offsets the increase in consumption, so it also increases the resources available for redistribution. The government refrains from exploiting the TI agents with the lowest productivity by bunching them with the TC agents, because any exploitation would only lead to less insurance. The only binding incentive compatibility constraints are the downward adjacent incentive compatibility constraints for the TC agents. For $\theta_m > \theta_1$, TI agents have strictly lower lifetime utility than TC agents of the same productivity. As a result, deterrent allocations must always be present in the menu for TC agents.

To see how deterrent allocations can be constructed, for any given $\left\{\left(c_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ and $\left\{\left(c_m^P, y_m^P\right)\right\}_{\theta_m \in \Theta}$, let $\left(c_m^D, y_m^D\right)$ with $y_m^D = y_m^P$, $c_{m,0}^D = c_{m,0}^P$ and $c_{m,t}^D = c_t^D$ for all $t > 0$ be the deterrent allocation satisfying:

$$\min_{\theta_m, \theta_{m'} \in \Theta} W_1\left(c_m^D, y_{m'}^D; \theta_m, \hat{\beta}\right) \geq \max_{\theta_m, \theta_{m'} \in \Theta} W_1\left(c_{m'}^P, y_{m'}^P, \theta_m, \hat{\beta}\right), \tag{13}$$

and

$$\min_{\theta_m \in \Theta} V_0\left(c_m^R, y_m^R; \theta_m, \beta\right) \geq \max_{\theta_m, \theta_{m'} \in \Theta} V_0\left(c_{m'}^D, y_{m'}^D; \theta_m, \beta\right). \tag{14}$$

By inequality (13), any TI agent who misreports as TC would select the deterrent allocation over the persistent allocation. Inequality (14) guarantees the TI agents would prefer to report their consistency truthfully. If (14) is satisfied, the TC agents would never choose the deterrent allocations over the persistent allocations. If (13) and (14) hold, then TI agents of any productivity would never misreport to be TC agents. Finally, the deterrent allocations can always be constructed such that (13) and (14) are satisfied.[21]

Define the intertemporal wedge as $\tau_t^C = 1 - \frac{u_t'(c_t)}{u_{t+1}'(c_{t+1})}$, and the intratemporal wedge as $\tau_t^L = 1 - \frac{\frac{1}{\theta} h_t'\left(\frac{y}{\theta}\right)}{u_t'(c_t)}$, which are also the implicit marginal tax rate on savings and labor respectively. The following theorem characterizes the optimal allocation and wedges in an environment with TC agents.

[21]To see how, first choose c_m^D so (14) holds. Next, increase c_1^D and decrease c_2^D such that $V_0\left(c_{m'}^D, y_{m'}^D; \theta_m, \beta\right)$ remains unchanged, and since $1 > \hat{\beta}$ and Assumption 1 holds, then it is possible to find c_1^D and c_2^D such that (13) holds.

Theorem 3 *For a conditional commitment mechanism, if Assumption 1 holds and $\phi \in (0,1)$ and $\hat{\beta} \in [\beta, 1)$, then the optimal allocation has the following properties: for any $t < 2$, (i.) $\tau_t^C = 0$ for all agents, (ii.) $\tau_t^L = 0$ for all TI agents with $\theta_m \geq \bar{\theta}$ and θ_M TC agents, (iii.) $\tau_t^L \geq 0$ for all TI agents with $\theta_m < \bar{\theta}$ and $\tau_t^L > 0$ for all TC agents with $\theta_m < \theta_M$.*

The usual trade-off between insurance and output efficiency is present in this economy. Theorem 3 demonstrates how the the output of the less productive agents is distorted downwards. This is standard in Mirrlees taxation. Also, the government is able to provide consumption smoothing for all agents, so the intertemporal wedge is undistorted.

The next corollary shows that as long as $\hat{\beta} \in [\beta, 1)$, then the conditional commitment mechanism can implement the same optimal allocation for different levels of sophistication. This follows from the fact that the optimal allocation in a conditional commitment mechanism does not depend on the sophistication of the TI agents.

Corollary 1 *In a conditional commitment mechanism, the optimal allocation is the same for any sophistication $\hat{\beta} \in [\beta, 1)$.*

The constrained efficient optimum is achieved when no TI agents are present ($\phi = 0$). Section 4 has shown that the efficient optimum is attainable when the economy is populated solely by TI agents ($\phi = 1$). Let $\mathcal{W}^T(\phi)$ denote the welfare under a conditional commitment mechanism with measure ϕ of TI agents. The second main result of the paper shows that social welfare increases as the proportion of TI agents increases, as shown in Figure 4.

Theorem 4 *$\mathcal{W}^T(\phi)$ increases with ϕ from the constrained efficient optimum to the full information efficient optimum.*

As the mass of TI agents increases, the first order effect is an increase in the available resources for redistribution, which comes from the allocation patterns in Theorem 2. Theorem 2 shows that the optimal policy induces TI agents to produce more output and consume less. This relaxes the resource constraint. The second order effect is that with less TC agents, the government can provide each TC agent more information rent using fewer resources. This relaxes the incentive compatibility constraints for TC agents (10), which enables the government to provide better insurance with more TI agents.

5.2 Betting Mechanisms

The government can also implement a betting mechanism when $\hat{\beta} \in (\beta, 1)$.[22] The government introduces the following menu for agents of productivity $\theta_m : C_m = \left\{C_m^{TC}, C_m^{TI}\right\}$,

[22] If $\hat{\beta} = 1$, TI agents believe their future-selves to be time-consistent. As a result, it is not possible to screen consistency using a betting mechanism without introducing additional distortions.

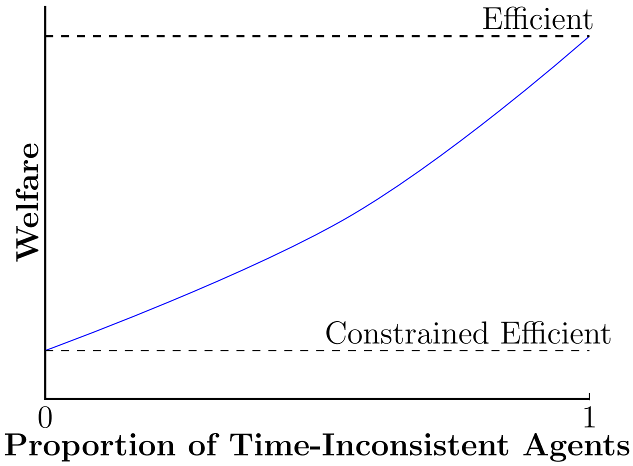

Figure 4: Welfare and Proportion of Time-Inconsistent Agents

where C_m^{TC} consists of the persistent and deterrent allocations and C_m^{TI} consists of the real and imaginary allocations.

Definition 9 *A direct betting mechanism for sophistication $\hat{\beta} \in (\beta, 1)$ with TC agents has $C = \left\{C_m^{TC}, C_m^{TI}\right\}_{\theta_m \in \Theta}$, with $C_m^{TC} = \left\{\left(c_m^P, y_m^P\right); \left(c_m^D, y_m^D\right)\right\}$ and $C_m^{TI} = \left\{\left(c_m^R, y_m^R\right); \left(c_m^I, y_m^I\right)\right\}$ satisfying: (i.) fooling constraints: $\forall \theta_m, \theta_{m'} \in \Theta$, $\left(c_m^I, y_m^I\right) \in C_m^{\hat{\beta}}$ and $\left(c_{m'}^I, y_{m'}^I\right) \in C_{m'|m}^{\hat{\beta}}$, (ii.) deterrent constraints: $\forall \theta_m, \theta_{m'} \in \Theta$, $\left(c_m^P, y_m^P\right) \in C_m^1$, and $\left(c_{m'}^D, y_{m'}^D\right) \in C_{m'|m}^{1|\hat{\beta}}$.*

The allocations that are truthfully implementable by a direct betting mechanism with TC agents is bounded by the incentive compatibility constraints (9) and (10) and the executability constraints (5). The definition of truthfully implementable allocations in this mechanism is similar to Definition 8.

Definition 10 *The allocation $\left\{\left(c_m^P, y_m^P\right), \left(c_m^R, y_m^R\right)\right\}_{\theta_m \in \Theta}$ is truthfully implementable for present bias β by a direct betting mechanism for sophistication $\hat{\beta}$ with TC agents if there exists $\left\{\left(c_m^D, y_m^D\right), \left(c_m^I, y_m^I\right)\right\}_{\theta_m \in \Theta}$ such that (i.) incentive compatibility, and (ii.) executability are satisfied.*

The betting mechanism with TC agents can lead to lower welfare than conditional commitment mechanisms. Notice that the threat allocations in the conditional commitment mechanism can be designed such that the TC agents would never choose it. However, it is difficult to deter TC agents from selecting the imaginary allocations in a betting mechanism.

To see this, let the persistent and real allocations be the ones implemented in an optimal conditional commitment mechanism. For them to be implemented in a betting mechanism, the incentive compatibility constraint for TC agents imply $U_0\left(c_m^P, y_m^P;\theta_m\right) \geq U_0\left(c_m^I, y_m^I;\theta_m\right)$, which can be expressed as

$$\left[u_0\left(c_{m,0}^P\right) - h_0\left(\frac{y_{m,0}^P}{\theta_m}\right)\right] - \left[u_0\left(c_{m,0}^R\right) - h_0\left(\frac{y_{m,0}^R}{\theta_m}\right)\right] \geq U_1\left(c_m^I, y_m^I;\theta_m\right) - U_1\left(c_m^P, y_m^P;\theta_m\right).$$

Similarly, the downward incentive compatibility constraint of the TI agents imply

$$U_1\left(c_m^I, y_m^I;\theta_m\right) - U_1\left(c_{m-1}^I, y_{m-1}^I;\theta_m\right) \geq \frac{1}{\beta}\left(\left[u_0\left(c_{m-1,0}^R\right) - h_0\left(\frac{y_{m-1,0}^R}{\theta_m}\right)\right] - \left[u_0\left(c_{m,0}^R\right) - h_0\left(\frac{y_{m,0}^R}{\theta_m}\right)\right]\right).$$

Notice that the optimal allocations implemented in a conditional commitment mechanism might not be implementable in a betting mechanism, because there may not exist imaginary allocations that satisfy both TC and TI incentive compatibility constraints for low values of β.[23] This is because imaginary allocations load the information rent on imaginary retirement consumption $c_{m,2}^I$. Imaginary retirement consumption would have to increase as β decreases for the incentive compatibility constraints of TI agents to hold. This would simultaneously make it more appealing for TC agents to misreport their consistency, because TC agents can mimic TI agents and select the imaginary allocations, which places an additional constraint on the imaginary allocations. As a result, when TC agents are present, conditional commitment mechanisms are more effective than betting mechanisms.

This has two critical implications. Firstly, this implies that conditional commitment mechanisms are more appealing than betting mechanisms. Secondly, since conditional commitment mechanisms have no effect on fully naïve agents, there is incentive for the government to help agents learn self-control so they would be at least partially naïve.

6 Social Security and Retirement Savings Accounts

In this section, decentralization of the optimal allocations in Sections 4 and 5 will be presented. I will first discuss the design of policies in an environment without TC agents. The betting mechanism can be decentralized with social security, and the conditional commitment mechanism can be decentralized with retirement savings accounts. A combination of both social security and retirement savings accounts can decentralize the hybrid mechanism. I will then discuss a decentralization with TC agents. This section also emphasizes the significant differences between the policy recommendations here and the ones that are being discussed

[23] By Theorem 3, if $\theta_{m-1} > \overline{\theta}$, then $u_0\left(c_{m-1,0}^R\right) - h_0\left(\frac{y_{m-1,0}^R}{\theta_m}\right) > u_0\left(c_{m,0}^R\right) - h_0\left(\frac{y_{m,0}^R}{\theta_m}\right)$.

in the literature and by policymakers. At the end of the section, I will provide an estimate of the quantitative impact of the policies.

6.1 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.[24] Also, while the US population is living longer, the average retirement age has remained steady for the past decade.[25] 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 for the rest of his/her life.[26] 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 10). 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.[27] As a result, early claimants are stuck with lower monthly benefits for the rest of their lives.

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. Knoll et al. (2015) devise effective choice architectures to delay claiming.[28] I propose a new approach to this issue.

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

[24]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.

[25]See Munnell (2015).

[26]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 $1932.

[27]Only 2% of the population choose to delay benefits till 70.

[28]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.

designing the benefits. The government can vary the progressivity of social security benefits with the age of initial claiming to decentralize the betting mechanism.[29]

With less progressive benefits for agents who claim late, non-sophisticated agents can be encouraged to work more efficiently at a younger age. Unknowingly, they 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.[30] The government and the agents bet on when the agents would claim their benefits and the progressivity of the benefits 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 $t = 2$.[31] 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 $t = 2$:

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

where $\mathbf{1}_t$ is an indicator function that is equal to 1 if benefits are claimed at t and zero otherwise.[32] In $t = 2$, the agents face the following budget constraint:

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

so consumption in retirement depends on savings and benefits. Notice that if the agent started claiming at $t = 1$, then the benefit in $t = 2$ is b_1. This models the current social security system, where the benefits depend on the time it was initially collected and the amount would perpetuate till death. The following proposition demonstrates the decentralization in an environment with fixed present bias and sophistication.

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

[29]See Footnote 11 for the definition of progressivity of benefits.

[30]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.

[31]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.

[32]Though most people choose to claim benefits during retirement, it is possible to claim benefits while working. For implementation, a penalty can be included for retiring early. One of the key recommendations of National Commission on Fiscal Responsibility and Reform (2010) is to raise the full retirement age.

Proposition 3 shows how social security can implement the efficient allocation with non-sophisticated agents. 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 stated in the National Commission on Fiscal Responsibility and Reform (2010).

6.2 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.[33] 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 relatively 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 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.[34] Comparatively, DC plans in the US are flexible and meet the transitory needs of a worker.[35] However, flexibility is undesirable if early withdrawal is due to present bias.[36]

Beshears et al. (2015a) showed that making the DC plan more illiquid for time-inconsistent agents can be an effective commitment device and increase savings. This paper provides an alternative view: such commitment can be provided in exchange for more efficient labor supply. The liquidity of DC plans can be redesigned to depend on income and act as a threat to sophisticated time-inconsistent agents, which incentivizes agents to produce efficiently and allow the government to provide better insurance. The conditional

[33]See Madrian and Shea (2001), Thaler and Benartzi (2004) and Carroll et al. (2009).

[34]For example, countries such as Germany, Singapore and the UK.

[35]Argento, Bryant, and Sabelhaus (2015) find 45% of contributions to retirement accounts among participants under the age of 55 in 2010 were offset by early withdrawals, which is higher than years prior to the Great Recession. Munnell and Webb (2015) estimates that if early withdrawals were not possible total 401(k) wealth would be 25% higher and total IRA wealth would be 23% higher.

[36]For example, see Ashraf, Karlan, and Yin (2006) and Beshears et al. (2015a).

commitment mechanism can be decentralized with early withdrawal as the off-path threat.[37]

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 a fixed amount $\eta \in (0, s_0]$ early from their accounts and receive $\xi(y_0, y_1, s_1) - \eta$ as a subsidy in $t = 1$.[38] Let $\tau(y_0, y_1, s_1)$ be the early withdrawal penalty, which is an income contingent off-path threat. A retirement account with contemporaneous income tax T_t and savings subsidy ρ_t is defined as $P^{ra} = (s_0, \tau(y_0, y_1, s_1), \xi(y_0, s_1), \eta, \rho_0, \rho_1, T_0(y_0), T_1(y_1))$.

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:

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

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

$$c_2 \leq \mathbf{1}_{EW}(1 - \tau(y_0, y_1, s_1))(s_0 + s_1 + s_2 - \eta) + (1 - \mathbf{1}_{EW})(s_0 + s_1 + s_2). \tag{15}$$

Inequality (15) shows that, with early withdrawal, c_2 decreases proportionally to the penalty $\tau(y_0, y_1, s_1)$, where $\tau(y_0, y_1, s_1) = \tau_m(y_1, s_1)$ if $y_0 \in \left[y^*_{m,0}, y^*_{m+1,0}\right)$ and

$$\tau_m(y_1, s_1) = \begin{cases} \hat{\rho}_M(y_0, s_1) & \text{if } y_1 \geq \bar{y}_M(y_0, s_1) \\ \hat{\rho}_{M-1}(y_0, s_1) & \text{if } y_1 \in [\bar{y}_{M-1}(y_0, s_1), \bar{y}_M(y_0, s_1)) \\ \vdots & \vdots \\ 1 & \text{if } y_1 < \bar{y}_{m+1}(y_0, s_1) \end{cases}.$$

The penalty is structured so that if y_1 is commensurate with y_0, then the agent would not be tempted to withdraw from the retirement account. Therefore, the account is illiquid and the agent is committed to having sufficient savings for retirement. If the agent produces more than his/her income history indicates, then the present-biased agent will withdraw early and be penalized by having lower retirement savings. The severity of the penalty increases with y_1 to deter productive agents from producing inefficiently. 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 $t = 1$. The following proposition demonstrates the

[37]This idea is similar to the implementation of partial illiquidity in Bond and Sigurdsson (2017), where the agent in $t = 1$ can increase c_1 and decrease c_2 to punish misbehavior in $t = 0$.

[38]For simplicity, I do not consider liquid savings accounts, like bank savings.

decentralization in an environment with fixed present bias and sophistication.

Proposition 4 *If $\hat{\beta} \in [\beta, 1)$, then the efficient allocation can be decentralized by P^{ra}.*

There are proposals to make the account more liquid the lower the income, which is consistent with Proposition 4.[39] However, there are two main differences between current plans and the plan proposed in Proposition 4. First, unlike the proposals where reliefs from adverse income shocks are the main concern, 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.[40] Another difference is that in Proposition 4, the early withdrawal penalty tax τ 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 increase labor efficiency.

6.3 Decentralizing the Hybrid Mechanism

The hybrid mechanism can be decentralized with social security and retirement savings accounts. The government endows the agents an initial savings of $s_0 > 0$ in their retirement accounts at the beginning of $t = 0$. The agents deposit s_t in their DC accounts, which maintain the same penalty features introduced previously. The social security benefits have the original structure in terms of progressivity, but the benefits are decreased by s_0. The government designs both P^{ss} and P^{ra} for some sophistication $\hat{\beta} \in (\overline{\beta}, 1)$ and the most present-biased agents $\beta = \underline{\beta}$. The initial savings s_0 is necessary, because the threat of a liquid retirement account is credible only if there were funds in the account. If $s_0 = 0$, sophisticated agents can always work inefficiently, choose a low s_1 and claim social security benefits early to mimic non-sophisticated agents.

Proposition 5 *If all agents are time-inconsistent with hidden present bias and sophistication, the efficient allocation $\{(c_m^*, y_m^*)\}_{\theta_m \in \Theta}$ can be decentralized by P^{ss} and P^{ra}.*

Proposition 5 follows from Propositions 3 and 4 and the proof of Theorem 1. An example of such an implementation is the central provident fund (CPF), a retirement account

[39] Obama proposed a hardship exception to early withdrawal penalties for those who have received unemployment for more than 26 weeks, see Fiscal Year 2017 Budget (2016).

[40] 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.

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 $s_0 > 0$. However, early withdrawal from the CPF account is not allowed.[41]

It should be noted that Moser and de Souza e Silva (2017) also provide a similar implementation albeit in a different environment. They consider an environment with dynamically stochastic present bias, so off-equilibrium path policies are ineffective. As a result, they arrive at a different conclusion on the design of social security and retirement accounts. 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 is different from the implementation in this paper, where social security and retirement accounts are used to separate sophistication and present bias.

6.4 Retirement Savings Policies with Time-Consistent Agents

From Section 5, the key insight in the mechanism with TC agents is the necessity of deterrent allocations. I will demonstrate an implementation where the decision to enroll in retirement savings accounts is voluntary. More specifically, TI agents are encouraged to enroll in the program, which provides them with commitment. TC agents are encouraged to opt out of the program and save on their own. This is because TC agents do not need the commitment provided by the retirement savings accounts, while full discretion in savings is undesirable for TI agents.

More specifically, to implement the optimal allocation presented in Section 5, the government makes participation in retirement savings accounts a voluntary decision. The retirement savings accounts, P^{ra}, takes the form described in Proposition 4. Agents have a choice of enrolling in P^{ra} or save on their own at interest rate $r = \frac{1}{\delta} - 1 = 0$. Income taxes would depend on the enrollment decision of the agents. Let T_t^{out} denote the income tax for agents who did not enroll in P^{ra}. Furthermore, to deter less present-biased agents from opting out, it may be necessary to introduce consumption subsidies conditional on savings, $\tau_1^{out,ns}$, and after-tax income $I_1^{out,ns}$ in $t = 1$ for those who chose to save on their own, which is triggered when intertemporal wedge in $t = 1$ is not zero: $u_1'(c_1) \neq u_2'(c_2)$.[42]

Proposition 6 *For the environment with TC agents, the optimal allocation can be decentralized by* $\{\tau_1^{out}, T_t^{out}, P^{ra}\}$.

[41] 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.

[42] For agents with sufficiently small β, the lack of commitment from saving on their own is a strong enough incentive to participate in retirement savings accounts.

Proposition 6 leads us to two important insights. Firstly, not only should the retirement savings accounts be partially liquid, its participation should also be voluntary. Secondly, when agents are not fully naïve, with a properly designed partially liquid retirement savings account in place, the government does not need to provide additional paternalistic measures to help agents save more for retirement. In fact, TI agents should be punished with low retirement savings and high present consumption when they do not participate in retirement savings accounts.

6.5 Quantitative Analysis

In this section, I will provide an estimate of the quantitative impact of the mechanism for a parametrized version of the economy. I will compute the wedges, the welfare gains and characterize the off-path distortions.

6.5.1 Data and Parameters

To obtain deterministic lifetime wage paths, I follow the procedure delineated in Weinzierl (2011) and use education level for productivity types.[43] The sample is divided into two education groups: high school or less (θ_L), and beyond high school (θ_H), with $\pi_H = 0.494$. I assume individuals work for 40 years, and spend 20 years in retirement. The data is divided into two working age periods: the young ($25 - 40$ years old), and the old ($41 - 64$ years old).[44] Individuals are assumed to start their retirement at the age of 65.[45] The wage paths are shown in Table 1.

	θ_L	θ_H
Young	13.67	19.38
Old	14.85	25.99

Table 1: Hourly Wage Paths by Productivity in 1999 US Dollars

I assume the period utility function is

$$\log(c) - \frac{1}{1+\sigma} l_t^{1+\sigma}.$$

[43]Wages are in 1999 U.S. dollars and the data is comprised of household heads from US PSID core sample for 1969-1999.

[44]Weinzierl (2011) runs a regression of wages on dummy variables for each age and finds a significant break between 40 and 41 years old. Best and Kleven (2013) also divides the sample into two age groups and uses the age of 41 as the cutoff.

[45]For individuals born in the United States before 1938, the full retirement age as defined by the Social Security Administration is 65 years old.

where $\sigma = 5$ so the Frisch elasticity is 0.2.[46] Notice that a logarithmic consumption utility satisfies Assumption 1. I use the estimates in Laibson et al. (2017), which estimate the annual discount factor $\delta = 0.986$ and $\beta = 0.519$ when relative risk aversion is 1. Let R denote the annual gross rate, I also assume $\delta R = 1$, which implies an annual gross rate of return of 1.3%. I will conduct the quantitative analysis for different values of ϕ.[47]

6.5.2 Intratemporal Distortions and Welfare Gain

By Theorem 3, when TC agents are present ($\phi < 1$) and none of the agents are fully naïve, the only distortions in the economy are the intratemporal distortions for the θ_L agents. Table 2 presents the intratemporal wedges (τ_t^L) for θ_L individuals at different values of ϕ.

	$\phi = 0$	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$
θ_L-young	0.08	0.079	0.078	0.077	0
θ_L-old	0.087	0.086	0.085	0.084	0

Table 2: Intratemporal Distortions for θ_L

The intratemporal wedge decreases as the population of present-biased agents grows. This is because θ_H-TI inividuals would consume less and work more than their time-consistent counterparts. As a result, when the population of present-biased agents increases, the government is able provide θ_H-TC individuals more consumption and leisure which relaxes the incentive compatibility constraint. This allows the government to decrease the intratemporal distortion on the θ_L individuals. Notice the drastic change in distortion when $\phi = 1$. This is because the efficient optimum is implemented and there are no distortions.

	$\phi = 0$	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$
θ_H-TC welfare	6.073	6.104	6.135	6.165	NA
θ_H-TI welfare	NA	5.849	5.883	5.917	5.839
θ_L welfare	5.801	5.836	5.871	5.904	6.05
Total welfare	5.935	5.937	5.939	5.941	5.946

Table 3: Welfare

Table 3 shows the overall welfare and lifetime utility by type for different proportions of TC individuals. Though the values are only meaningful ordinally, Table 3 shows the welfare ranking of the individuals in the economy. An increase in the population of TI individuals

[46]This is consistent with many studies on labor supply elasticity, see Keane (2011) for a survey.

[47]Mahajan and Tarozzi (2011) estimates that 30% to 40% of individuals in their sample are time-consistent. Meier and Sprenger (2010) finds 36% and 9% of the individuals in their sample to be present-biased and future-biased respectively. On the other hand, Chan (2017) estimates that 95% of individuals in his sample have $\beta < 0.9$, so ϕ is close to 1.

would increase total output in the economy., With the higher total output, even the θ_H-TC agents can consume more and work less at the optimum. Table 4 shows the output of each agent and Table 5 shows the consumption.

	$\phi = 0$	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$
$(y^P_{H,0}, y^P_{H,1})$	$(20.2, 28.72)$	$(20.15, 28.66)$	$(20.11, 28.6)$	$(20.07, 28.54)$	NA
$(y^R_{H,0}, y^R_{H,1})$	NA	$(20.51, 29.16)$	$(20.46, 29.09)$	$(20.41, 29.03)$	$(20.52, 29.18)$
$(y_{L,0}, y_{L,1})$	$(13.55, 14.94)$	$(13.52, 14.91)$	$(13.49, 14.87)$	$(13.45, 14.84)$	$(13.5, 14.91)$
Total Output	$(16.83, 21.74)$	$(16.84, 21.75)$	$(16.84, 21.77)$	$(16.85, 21.78)$	$(16.96, 21.95)$

Table 4: Output

	$\phi = 0$	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$
c^P_H	15.77	15.94	16.11	16.28	NA
c^R_H	NA	14.61	14.78	14.95	14.57
c_L	13.14	13.32	13.49	13.66	14.57

Table 5: Consumption

Secondly, the welfare for $\phi = 0$ can be interpreted as the social welfare from implementing traditional policies. In essence, if the government observes present bias and uses a linear savings subsidy to off-set it, then the bias is mitigated independent of the asymmetric information, and the optimum for $\phi = 0$ is achieved. Table 6 presents the welfare gains in terms of increase in consumption. For example, if $\phi = 1$, then there is an additional gain in welfare equivalent to a 0.46% increase in aggregate consumption or roughly $62 billion in current U.S. dollars annually. In other words, under traditional policies, if the government can increase the consumption of individuals by 0.46% without changing labor supply, then the welfare would be equivalent to the setting with $\phi = 1$.

	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$	$\phi = 1$
% Increase in Consumption	0.068%	0.148%	0.241%	0.46%

Table 6: Welfare Gain Over Traditional Policies in Consumption Equivalents

6.5.3 Off-Path Policies

I will examine the off-path distortions used to sustain the welfare improvements. The set of off-path policies is large. I will focus on the minimal or least distortionary off-path allocations necessary to implement the optimum.[48] There are few studies that estimate

[48]For the conditional commitment mechanism, the threat allocations are chosen such that the following constraints are binding: the local downward incentive compatibility, executability at β and threat constraint

sophistication, and a wide range of estimates have been produced.[49] I will set $\hat{\beta} = 0.68$, which is greater than β and is consistent with the mean estimate of $\hat{\beta}$ in Chan (2017).[50]

First, I will analyze the off-path distortions in an environment with TC agents. Section 5 showed how a conditional commitment mechanism would be more effective than a betting mechanism when TC agents are present. Hence, I will focus on the off-path distortions in a conditional commitment mechanism. Table 7 presents the minimal intertemporal and intratemporal wedges, τ_1^C and τ_1^L respectively, necessary to sustain the optimum.

	$\phi = 0.25$	$\phi = 0.5$	$\phi = 0.75$
Intertemporal Distortion	0.951	0.949	0.947
θ_H Intratemporal Distortion	0.822	0.824	0.826
θ_L Intratemporal Distortion	−4.124	−4.056	−3.991

Table 7: Off-Path Distortions in a Credible Commitment Mechanism

Notice how the off-path policies exploit the present bias and distort the intertemporal wedge to sustain the optimum. Firstly, the off-path policy induces extremely low retirement consumption and loads almost all consumption at $t = 1$. Secondly, the off-path intertemporal distortion decreases with the proportion of present-biased individuals. This is because the optimal utility for θ_H-TI individuals increases with ϕ, so it becomes easier to deter them from misreporting and threats can be less extreme. Finally, notice that the off-path policy induces too much labor supply, so θ_L individuals are deterred from choosing it.

Finally, I will analyze the off-path distortions in an environment with only TI agents. Table 8 presents the minimal intertemporal and intratemporal wedges necessary to sustain the efficient optimum.

	Conditional Commitment	Betting
Intertemporal Distortion	0.996	−249.9
θ_H Intratemporal Distortion	0.377	0.835
θ_L Intratemporal Distortion	−16.912	−3.73

Table 8: Off-Path Distortions in When All Agents are Time-Inconsistent

Notice that to implement the full information efficient optimum, the off-path policies have to be quite extreme. The distortions for the conditional commitment mechanism are qualitatively similar to the ones presented in Table 7, but are significantly larger quantitatively. The off-path consumption for the betting mechanism has individuals believing that

at $\hat{\beta}$. For the betting mechanism, the imaginary allocations are chosen such that the local downward incentive compatibility at β and fooling constraint at $\hat{\beta}$ are binding.

[49] See Fang and Wang (2015) and Chan (2017).

[50] In Fang and Wang (2015), the annual discount factor ranged from 0.681 to 0.947 depending on the specification, but $\hat{\beta}$ was close to one in all specifications with a short-run discount factor of 0.68.

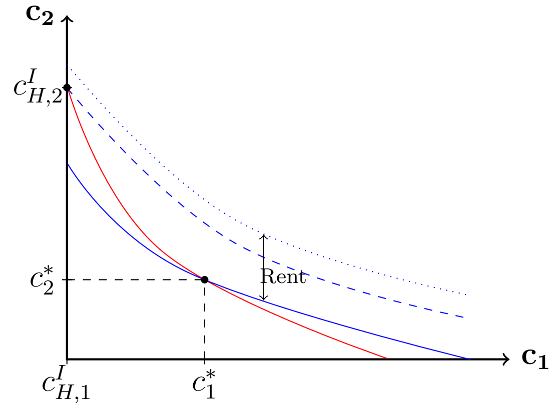

Figure 5: Limited Fooling

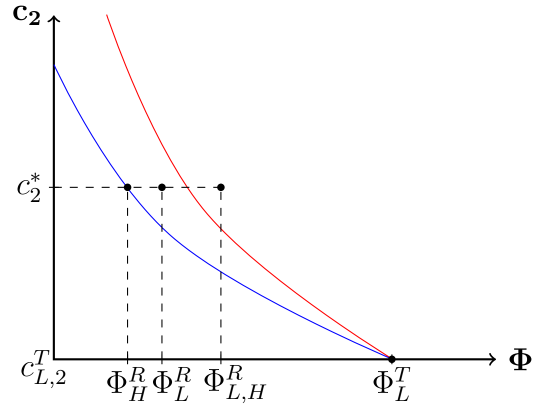

Figure 6: Limited Threats

they will consume very little before retirement, but consumption is extremely backloaded. The reason for this is because there is significantly more redistribution from θ_H to θ_L when $\phi = 1$. As a result, the off-path policies are more extreme. However, it should be noted that some estimates suggest a significant population of time-consistent individuals (see footnote 47), so such policies may not be advisable.

7 Discussion

7.1 Limited Promises and Punishments

Assumption 1 provides a non-empty set of bets that could deceive non-sophisticated agents for any information rent. If Assumption 1 fails, the efficient optimum might not be implementable. To see this, consider the case where the utility function is bounded below. Figure 5 illustrates how betting can be limited for fully naïve agents with $\theta \in \{\theta_L, \theta_H\}$. 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.

As a result, asymmetric information causes distortions if the imaginary allocations cannot fully cover the minimal information rent necessary for truth-telling. The government would have to distort consumption and output downwards for lower typesA similar argument can be made for conditional commitment mechanisms. Figure 6 illustrates how threats are limited in a sophisticated case with two productivity types.

Nevertheless, the government is still able to improve welfare above the constrained efficient optimum by using betting or conditional commitment mechanisms when Assumption 1 fails. This is because a portion of information rents is loaded on the off-equilibrium path allocations, which helps relax incentive compatibility and decrease distortions.

7.2 Trembling Hand

A legitimate concern when decentralizing the betting and conditional commitment mechanisms is the potential for agents to unintentionally select the off-path allocations. This could lead to significant welfare loss for the agents.

For illustrative purposes, I will focus the discussion on a fixed present bias β and sophistication $\hat{\beta}$. Also assume that $u_t = u$ and $h_t = h$ for all t. Consider a betting mechanism where agents might unintentionally choose the imaginary allocation with probability $\kappa \in (0, 1)$. As a result, the imaginary allocations would enter the feasibility constraint. Assume that the agents are oblivious to the possibility of making a mistake. The following theorem characterizes the intertemporal wedge in this environment.

Proposition 7 *If $\kappa \in (0, 1)$, the intertemporal wedge has the following properties for $\theta_m > \theta_1$: (i.) $\tau_1^C \leq 0$ for the imaginary allocations with strict inequality for θ_M and (ii.) $\tau_1^C \geq 0$ for the real allocations with strict inequality for θ_M.*

Proposition 7 shows how consumption smoothing is not implementable for agents with $\theta_m > \theta_1$, and the best the government can do is to provide consumption smoothing for them in expectation. Since the imaginary allocation attempts to load the information rent on retirement consumption, those who choose it would over-save. Therefore, the agents choosing the real allocations would under-save. This is true even for the most productive agents, so there will be distortions at the top.

On the other hand, for conditional commitment mechanisms, if agents might choose the threat allocation, consumption smoothing is not implementable for lower productivity types. This is because the threat is introduced into the menu of low productivity types. Also, since the threat allocation works by exacerbating present bias, the threat allocation would have agents under-saving so the real allocation would have agents over-saving.

7.3 Outside Commitment Devices

In reality, self-control problems can be mitigated by a wide array of commitment devices available in the market.[51] In the case of sophisticated agents, if commitment devices are

[51] There is a growing market for commitment devices. For example, StickK, Pact and Beeminder are some recent websites that offer commitment contracts.

available and its usage is unobservable, then threats are less potent. This is because agents can purchase commitment and bind themselves to an intertemporal allocation. Therefore, screening of productivity would be more costly 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 undesirable. This has the additional benefit of preventing the non-sophisticated agents from using an inefficient amount of commitment (Heidhues and Koszegi, 2009).

7.4 Paranoid Time-Consistent Agents

Previous analysis assumed the TC agents were sophisticated ($\hat{\beta} = \beta = 1$). A paranoid agent is a non-sophisticated TC agent who believes the future-self is present biased. Paranoia affects the behavior of TC agents, and consequently government policy can be adjusted to exploit it. Paranoid agents respond to threats, and can also be fooled. It is easy to see how a conditional commitment mechanism can extract information rents from paranoid agents, since the government can construct the threat allocation using similar methods for the TI agents. The betting mechanism is more subtle.

To see how a betting mechanism achieves the efficient optimum in an economy with only paranoid agents, $\hat{\beta} < \beta = 1$, consider the example with $\Theta = \{\theta_L, \theta_H\}$. Let $C_m = \left\{(c^*, y_m^*), \left(c_m^I, y_m^*\right)\right\}$ with $c_{m,0}^I = c_0^*$ and $c_H^I = c^*$. The allocations satisfy the fooling constraints $u_1\left(c_{L,1}^I\right) + \hat{\beta} u_2\left(c_{L,2}^I\right) \geq u_1\left(c_1^*\right) + \hat{\beta} u_2\left(c_2^*\right)$, the executability constraints $u_1\left(c_1^*\right) + u_2\left(c_2^*\right) \geq u_1\left(c_{L,1}^I\right) + u_2\left(c_{L,2}^I\right)$, and the incentive compatibility constraints, which implies the following

$$\sum_{t=0}^{1} h_t\left(\frac{y_{H,t}^*}{\theta_L}\right) - h_t\left(\frac{y_{L,t}^*}{\theta_L}\right) \geq \left[u_1(c_1^*) + u_2(c_2^*)\right] - \left[u_1(c_{L,1}^I) + u_2(c_{L,2}^I)\right] \geq \sum_{t=0}^{1} h_t\left(\frac{y_{H,t}^*}{\theta_H}\right) - h_t\left(\frac{y_{L,t}^*}{\theta_H}\right).$$

The fooling and executability constraints imply $c_{m,1}^I > c_1^*$ and $c_{m,2}^I < c_2^*$. Combined with the incentive compatibility constraints, it must be that $c_{L,1}^I > c_{H,1}^I$ with $c_{L,2}^I < c_{H,2}^I$. In essence, the government fools the paranoid agents by choosing the imaginary allocations to exacerbate their fears. A paranoid agent would predict choosing the imaginary allocations even though the agent is strictly worse off by choosing it, because he/she does not think the real allocation is attainable. The government takes advantage of this by making the imaginary allocation for the θ_L agent even worse. Hence, the paranoid θ_H agent produces efficiently because there is a fear that by misreporting, he/she would have even less savings.

The way the betting mechanism works for paranoid agents is in stark contrast to the logic presented in the previous sections. Non-sophisticated TI agents are fooled by empty promises, but paranoid agents are fooled by empty threats.

If the economy has both paranoid and TI agents, then using a betting mechanism could be problematic. For example, imagine an economy with paranoid TC agents with incorrect belief $\hat{\beta} < 1$, which corresponds to the belief of the non-sophisticated TI agents with $\beta < \hat{\beta}$. If the government tries to fool the agents, then it is not possible for it to separate agents along consistency level. In this particular case, depending on who the government chooses to fool, either the paranoid TC agents would end up selecting the imaginary allocation used to fool the TI agents or the TI agent would choose the imaginary allocation used to fool the TC agents. The resulting welfare would be lower compared to when TC agents are sophisticated. However, such a problem does not arise when the government uses a conditional commitment mechanism, because the same threats for TI agents can also deter paranoid agents from misreporting.[52]

7.5 Dynamic Stochastic Shocks

The recent literature on optimal dynamic Mirrlees taxation have focused on stochsti-cally evolving productivity.[53] This is in contrast to the current paper, which assumes a constant productivity. This assumption seems innocuous since Keane and Wolpin (1997) found that the bulk of labor-market uncertainty can be explained by skill endowments in adolescence. Also, Guvenen et al. (2016) showed that most individuals experience very little income change in a given year, which suggests a highly persistent income process. However, Guvenen et al. (2016) also showed that the distribution of earnings changes exhibit high kurtosis, so a non-negligible number of people experience large income changes. Therefore, a characterization of optimal betting and conditional commitment mechanisms with dynamic stochastic productivity shocks would be useful for the design of policy.

Conditional commitment mechanisms could help raise welfare in settings with dynamic shocks and time-inconsistent agents. Halac and Yared (2014) focus on public strategies, so players can only report their current taste shock. However, they point out that by allowing the players to report on past shocks, it can potentially relax the incentive compatibility constraint by punishing inconsistent reports. However, they show that this would not work in a setting with independent shocks as in Amador, Werning, and Angeletos (2006), so conditional commitment mechanisms are useful as long as shocks are persistent.

To see how, consider two productivity types, $\Theta = \{\theta_L, \theta_H\}$, and suppose Assumption 1

[52] This is true because the credible threat constraints and the incentive compatibility constraints are the same for both the TI and TC agents who share the same beliefs. Finally, the executability constraint is more relaxed for the TC agents than the TI agents.

[53] See surveys Golosov, Tsyvinski, and Werning (2006) and Kocherlakota (2010) for a comprehensive list of papers.

holds. If productivity shocks are independent across time, then θ_0 provides no information on θ_1, so off-equilibrium threats do not deter misreports. However, if shocks are persistent, then θ_0 is informative of θ_1. In particular, if $\Pr(\theta_{L,1}|\theta_{L,0}) = 1$ and $\Pr(\theta_{H,1}|\theta_{H,0}) > 0$, then the efficient allocation is implementable. The low productivity shocks need to be fully persistent, because low productivity agents in $t = 0$ need to be protected from the off-equilibrium punishment in $t = 1$, which is similar to Proposition 3 in Bond and Sigurdsson (2017).

Meier and Sprenger (2015) has documented the temporal stability of time preferences, which supports the environment in this paper. Moser and de Souza e Silva (2017) analyzed an environment where present bias is independently stochastic over time. They show that in their environment off-path policies do not relax the incentive compatibility constraints. Indeed, off-path threats require the agents knowing what their future-selves might do and independently stochastic time inconsistency hinders the ability of agents to predict.

8 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 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 provision of commitment 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 does not 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 Proposition 3: Let $\left(c^I_m, y^I_m\right)$ be imaginary allocations constructed in a direct mechanism that support the efficient allocation, where $y^I_m = y^*_m$ and $c^I_{m,0} = c^*_{m,0}$ for all $\theta_m \in \Theta$. Let $Y^*_t = \left\{y^*_{1,t}, \dots, y^*_{m,t}, \dots, y^*_{M,t}\right\}$. Furthermore, if $y_t \in Y^*_t$, let $y^{-1}_t \in \Theta$ denote the corresponding type. For example, $y^{-1}_{m,t} = \theta_m$. Let $EO = \left\{y = (y_0, y_1) \,|y_0 \in Y^*_0, y_1 \in Y^*_1 \text{ and } y^{-1}_0 = y^{-1}_1\right\}$ denote set of efficient output history. Consider the following taxes: $T_0(y_0) = y_0 - c^*_0$, and $T_1(y_0, y_1, k_1) = y_1 + k_1 + \mathbf{1}_1 b_1(y_0, y_1) +$

$\mathbf{1}_2 \alpha(y_0, y_1) - c_1^*$, and benefits:

$$b_1(y_0, y_1) = \begin{cases} c_2^* & \text{if } y \in EO \\ 0 & \text{otherwise} \end{cases}, \quad b_2(y_0, y_1) = \begin{cases} c_{m,2}^I & \text{if } y = y_m^* \\ 0 & y \notin EO \end{cases},$$

with $\alpha(y_0, y_1) = c_1^* - c_{m,1}^I$, 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, by Assumption 1, it is always possible to lower $c_{m,1}^I$ and increase $c_{m,2}^I$ such that incentive compatibility holds. Since u_t is strictly concave, the increase in $c_{m,2}^I$ would need to be large compared to the decrease in $c_{m,1}^I$. 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-path policies ($\mathbf{1}_{EW} = 0$). 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 $\mathbf{1}_{EW} = 0$, income taxes are

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

With this setup, agents in $t = 0$ and $t = 1$ would choose the efficient consumption on-path.

Let $\left\{ \left(c_{m|n}^T, y_{m|n}^T \right)_{\theta_n > \theta_m} \right\}_{\theta_m \in \Theta}$ be threat allocations constructed in a direct conditional commitment mechanism that support the efficient allocation. Set $\bar{y}_n(y_0, s_1) = y_{m|n,1}^T$ if $y_0 \in \left[y_{m,0}^*, y_{m+1,0}^* \right)$. Given the contemporary income tax $T_t(y_t)$, off-path output will be $y_{m|n,1}^T$ for agents with productivity θ_n who produced $y_{m,0}^*$ in $t = 0$. Next, set $\eta = s_0$ and if $y_0 \in \left[y_{m,0}^*, y_{m+1,0}^* \right)$, $s_1 \geq 0, y_1 \in [\bar{y}_n(y_0, s_1), \bar{y}_{n+1}(y_0, s_1))$,

$$\hat{\rho}_n(y_0, s_1) = 1 - \frac{\beta u_1' \left(c_{m|n,1}^T \right)}{\hat{\beta} u_2' \left(c_{m|n,2}^T \right)}; \xi(y_0, y_1, s_1) = c_{m|n,1}^T - c_1^T + \beta \left[\frac{c_{m|n,2}^T}{1 - \hat{\rho}(y_0, s_1)} - (c_2^* - s_0) \right].$$

Given this off-equilibrium path policy, agents would consume the threat allocations if output was not efficient in $t = 0$. ■

Proof of Proposition 6: Following the same process as in the proof of Proposition 4, let P^{ra} be chosen to implement the real allocations $\left(c_m^R, y_m^R \right)$ for all $\theta_m \in \Theta$.

Next, let $Y_t^P = \left\{ y_{1,t}^P, \ldots, y_{m,t}^P, \ldots, y_{M,t}^P \right\}$. Furthermore, if $y_t \in Y_t^P$, let $y_t^{-1} \in \Theta$ denote the corresponding type. Let $EO^{out} = \left\{y = (y_0, y_1) \,|y_0 \in Y_0^P, y_1 \in Y_1^P \text{ and } y_0^{-1} = y_1^{-1}\right\}$ denote the set of equilibrium output history. For all $t < 2$, let $\hat{c}_t : \mathbb{R}_+ \mapsto \mathbb{R}_+$ be such that

$$\hat{c}_t\left(y_t\right) \equiv \max_{y_t' \leq y_t} \hat{c}\left(y_t'\right) \text{ subject to } y_t' \in Y_t^P \text{ and } \hat{c}_t\left(y_{m,t}^P\right) = c_{m,t}^P.$$

Similarly, for all $t < 2$, and $\hat{c}_2 : \mathbb{R}_+ \mapsto \mathbb{R}_+$ define $\hat{c}_2\left(y_1\right) \equiv \max_{y_1' \leq y_1} \hat{c}\left(y_1'\right)$ subject to $y_1' \in Y_1^P$ and $\hat{c}_2\left(y_{m,1}^P\right) = c_{m,2}^P$. Also, choose T_t^{out} as a function of output history:

$$T_0^{out}\left(y_0\right) = \begin{cases} y_0 - \hat{c}_0\left(y_0\right) & \text{if } y_0 \geq y_{1,0}^P \\ y_0 & \text{otherwise} \end{cases}, \; T_1^{out}\left(y_0, y_1\right) = \begin{cases} y_1 - \hat{c}_1\left(y_1\right) - \hat{c}_2\left(y_1\right) & \text{if } y \in EO^{out} \\ y_1 & \text{otherwise} \end{cases}.$$

By the construction of $\hat{c}_t$ and T_t^{out}, agents do not have an incentive to produce outside of the set Y_t^P for all t. Furthermore, by the construction of T_1^{out}, agents of productivity θ_m would produce y_m^P, since after-tax income would be zero if $y \notin EO^{out}$. Since the persistent allocations are incentive compatible, TC agents would produce according to their productivity.

Finally, let c_1^D and c_2^D satisfy (13) and (14). To separate TI agents from TC agents, choose $1 + \tau_1^{out,ns} = \frac{u_1'\left(c_1^D\right)}{\hat{\beta} u_2'\left(c_2^D\right)}$ and $I^{out,ns} = \left(1 + \tau_1^{out,ns}\right) c_1^D + c_2^D$. As a result, a TI agent mimicking a TC agent would predict consuming $\left(c_1^D, c_2^D\right)$ when solving: $\max_{c_1, c_2} u_1\left(c_1\right) + \hat{\beta} u_2\left(c_2\right)$ subject to $\left(1 + \tau_1^{out,ns}\right) c_1 + c_2 \leq I^{out,ns}$, which is triggered whenever $u_1'\left(c_1\right) \neq u_2'\left(c_2\right)$. ■

References

Ali, Nageeb. 2011. "Learning Self-Control." *Quarterly Journal of Economics* 126:857–893.

Amador, Manuel, Ivan Werning, and George-Marios Angeletos. 2006. "Commitment vs. Flexibility." *Econometrica* 74 (2):365–396.

Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman, and Stephen Weinberg. 2001. "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation." *Journal of Economic Perspectives* 15 (3):47–68.

Argento, Robert, Victoria Bryant, and John Sabelhaus. 2015. "Early Withdrawals from Retirement Accounts During the Great Recession." *Contemporary Economic Policy* 33 (1):1–16.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Phillipines." *Quarterly Journal of Economics* 121 (2):635–672.

Auerbach, Alan J., Laurence J. Kotlikoff, Darryl Koehler, and Manny Yu. 2016. "Is Uncle Sam Inducing the Elderly to Retire?" *NBER Working Paper* (22770).

Beshears, John, James Choi, Christopher Harris, David Laibson, Brigitte Madrian, and Jung Sakong. 2015a. "Self Control and Commitment: Can Decreasing the Liquidity of a Savings Account Increase Deposits?" *NBER Working Paper* (21474).

Beshears, John, James Choi, Joshua Hurwitz, David Laibson, and Brigitte Madrian. 2015b. "Retirement Savings and Household Decisions." *American Economic Review, Papers and Proceedings* 105 (5):420–425.

Best, Michael and Henrik Kleven. 2013. "Optimal Income Taxation with Career Effects of Work Effort." *Working Paper* .

Bond, Philip and Gustav Sigurdsson. 2017. "Commitment Contracts." *Review of Economic Studies* Forthcoming.

Carroll, Gabriel, James Choi, David Laibson, Brigitte Madrian, and Andrew Metrick. 2009. "Optimal Defaults and Active Decisions." *Quarterly Journal of Economics* 124 (4):1639–1674.

Chan, Marc. 2017. "Welfare Dependence and Self-Control: An Empirical Analysis." *Review of Economic Studies* 84:1379–1423.

Coile, Courtney, Peter Diamond, Jonathan Gruber, and Alan Jousten. 2002. "Delays in Claiming Social Security Benefits." *Journal of Public Economics* 84.

Diamond, Peter and Johannes Spinnewijn. 2011. "Capital Income Taxes with Heterogeneous Discount Rates." *American Economic Journal: Economic Policy* 3.

Eliaz, Kfir and Ran Spiegler. 2006. "Contracting with Diversely Naive Agents." *Review of Economic Studies* 73.

Esteban, Susanna and Eiichi Miyagawa. 2005. "Optimal Menu of Menus with Self-Control Preferences." *NAJ Economics, Peer Reviews of Economics Publications* 12.

Executive Order No. 13707. 2015. *3 CFR* 80 (181).

Fang, Hanming and Yang Wang. 2015. "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions." *International Economic Review* 56 (2).

Farhi, Emmanuel and Xavier Gabaix. 2015. "Optimal Taxation with Behavioral Agents." *Working Paper* .

Fiscal Year 2017 Budget. 2016. "Great Explanations of the Administration's Fiscal Year 2017 Revenue Proposals." .

Gabaix, Xavier. 2014. "A Sparsity-Based Model of Bounded Rationality." *Quarterly Journal of Economics* 129 (4).

Galperti, Simone. 2015. "Commitment, Flexibility, and Optimal Screening of Time Inconsistency." *Econometrica* 83 (4).

Golosov, Mikhail, Aleh Tsyvinski, and Ivan Werning. 2006. "New Dynamic Public Finance: A User's Guide." *NBER Macroeconomics Annual 2006* .

Guo, Jang-Ting and Alan Krause. 2015. "Dynamic Nonlinear Income Taxation with Quasi-Hyperbolic Discounting and No Commitment." *Journal of Economic Behavior and Organization* 109.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2016. "What Do Data on Millions of US Workers Reveal about Life-Cycle Earnings Dynamics." *Working Paper* .

Halac, Marina and Pierre Yared. 2014. "Fiscal Rules and Discretion Under Persistent Shocks." *Econometrica* 82 (5).

Heidhues, Paul and Botond Koszegi. 2009. "Futile Attempts at Self-Control." *Journal of the European Economic Association* 7 (2-3).

———. 2010. "Exploiting Naivete about Self-Control in the Credit Market." *American Economic Review* 100 (5).

Hellwig, Martin F. 2007. "A Contribution to the Theory of Optimal Utilitarian Income Taxation." *Journal of Public Economics* 91.

Keane, Michael. 2011. "Labor Supply and Taxes: A Survey." *Journal of Economic Literature* 49 (4).

Keane, Michael and Kenneth Wolpin. 1997. "The Career Decisions of Young Men." *Journal of Political Economics* 105.

Knoll, Melissa, Kirstin Appelt, Eric Johnson, and Jonathan Westfall. 2015. "Time to Retire: Why Americans Claim Benefits Early and How to Encourage Delay." *Behavioral Science & Policy* 1 (1).

Knoll, Melissa and Anya Olsen. 2014. "Incentivizing Delayed Claiming of Social Security Retirement Benefits Before Reaching the Full Retirement Age." *Social Security Bulletin* 74 (4).

Kocherlakota, Narayana. 2010. "The New Dynamic Public Finance." *Princeton University Press* .

Krusell, Per, Burhanettin Kuruscu, and Anthony Smith. 2010. "Temptation and Taxation." *Econometrica* 78 (6).

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112 (2).

Laibson, David, Peter Maxted, Andrea Repetto, and Jeremy Tobacman. 2017. "Estimating Discount Functions with Consumption Choices over the Lifecycle." *Working Paper* .

Lockwood, Benjamin. 2016. "Optimal Income Taxation with Present Bias." *Working Paper* .

Madrian, Brigitte and Dennis Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics* 116 (4).

Mahajan, Aprajit and Alessandro Tarozzi. 2011. "Time Inconsistency, Expectations and Technology Adoption: The case of Insecticide Treated Nets." *Working Paper* .

Meier, Stephan and Charles Sprenger. 2010. "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics* 2 (1).

———. 2015. "Temporal Stability of Time Preferences." *Review of Economics and Statistics* 97 (2).

Moser, Christian and Pedro Olea de Souza e Silva. 2017. "Paternalism vs Redistribution: Designing Retirement Savings Policies with Behavioral Agents." *Working Paper* .

Munnell, Alicia. 2015. "The Average Retirement Age - An Update." *Briefs from Center for Retirement Research at Boston College* 15-4.

Munnell, Alicia and Anthony Webb. 2015. "The Impact of Leakages from 401(k)s and IRAs." *Working Paper from Center for Retirement Research at Boston College* (2015-2).

National Commission on Fiscal Responsibility and Reform. 2010. "The Moment of Truth: Report of the National Commission on Fiscal Responsibility and Reform." .

National Research Council. 2012. *Aging and the Macroeconomy: Long-Term Implications of an Older Population*.

O'Donoghue, Ted and Matthew Rabin. 2001. "Choice and Procrastination." *Quarterly Journal of Economics* 116.

Scholz, John Karl, Ananth Seshadri, and Surachai Khitatrakun. 2006. "Are Americans Saving 'Optimally' for Retirement?" *Journal of Political Economy* 114 (4).

Shoven, John B. and Sita Nataraj Slavov. 2014. "Does it Pay to Delay Social Security?" *Journal of Pension Economics and Finance* 13 (2).

Thaler, Richard and Shlomo Benartzi. 2004. "Save More Tomorrow: Using Behavioral Economics to Increase Employee Savings." *Journal of Political Economy* 112.

Weinzierl, Matthew. 2011. "The Surprising Power of Age-Dependent Taxes." *Review of Economic Studies* 78.